An Investigation Into Fault Diagnosis of Planetary Gearboxes Using A Bispectrum Convolutional Neural Network

Xinyu Pang, Xuanyi Xue, Wangwang Jiang, and Kaibo Lu

***Abstract*—To improve the efficiency and accuracy of fault diagnostics of planetary gearboxes, an intelligent diagnosis approach is proposed based on deep convolutional neural networks (CNNs) and vibration bispectrum (BSP). Rather than using raw vibration signals, BSP is appreciated as the input for the CNN models (denoted as BSP-CNN) because the BSP allows nonlinear feature enhancement and noise reduction. In addition, transfer learning (TL) is accompanied to address the challenges of CNN difficulties. The proposed BSP-CNN is verified firstly to diagnose a number of common faults including gear states: normal, tooth wear, tooth root crack, tooth breakage and missing tooth, achieving an accuracy of 97.36% in identifying different faults. Then, its TL capability is evaluated based on the sun gear faults datasets. The classification accuracy of the planet gear faults is over 95.1%. After the transfer learning, the classification accuracy of the sun gear fault is still higher than 97.9%, and the computational time consumed by proposed method is also less compared to other diagnosis methods. This article has twofold contributions: first, the development of a BSP-based CNN model for fault diagnosis; andsecond, the extensive evaluation of CNN-TL methods for monitoring and diagnosing planetary gearboxes.**

***Index Terms*—Bispectrum (BSP) analysis, deep convolutional neural network (CNN), fault diagnosis, planetary gearbox, transfer learning (TL).**

I. Introduction

A PLANETARY gearbox has a very complex structure, a large transmission ratio, strong bearing capacity, and high transmission efficiency. It has been widely used in large industrial equipment, machine tools, and heavy machinery (e.g., aviation, agricultural, and transportation) [1]. Different types of fault, such as wear, crack or even broken, will occur in planetary gearbox after long-term working under the conditions with low speed, heavy loading and strong impact. If these faults cannot be detected and diagnosed in time, it may lead to the failure of entire transmission system, even cause safety accidents [2]. Therefore, planetary gearbox fault diagnosis has become great significance to the safe operation and health maintenance of mechanical equipment.

The vibration signals generated when the planet gears mesh simultaneously with both the sun gear and the inner ring gear will produce significant nonlinear coupling phenomenon which will result in a large number of modulation components when the failure occurs. With the interference of the vibration signals of the shaft and bearing, the signals excited by the fault of the planetary gearbox are often submerged in the noise, with weak fault characteristics, low signal-to-noise ratio, and complex nonlinear characteristics [3], [4]. Therefore, it is necessary to preprocess the raw signal for enhancing the feature information, so as to get a better diagnosis. The bispectrum (BSP) technique is a 2-D discrete Fourier transform having a third-order cumulant, which can well-detect phase coupling, remove uncoupled frequency components, deal with nonlinear signals well and suppress Gaussian noise effectively. Therefore, BSP is particularly effective in extracting feature of modulation signals [5]. In addition to image processing in medicine [6], [8], BSP analysis is also gradually used in vibration signal processing of mechanical equipment. Li *et al*. [9] proposed a fault diagnosis method based on BSP entropy and deep belief network which accurately predicted the trends and random fluctuations during the performance degradation of the hydraulic pump. Howard [10] used the BSP and the trispectrum to monitor the machine vibration condition by detecting phase coherence between various frequency components. Jiang *et al*. [11] took the BSP distribution as features for roller bearing and gear fault diagnosis.

Although BSP analysis has been successfully applied in fault feature extraction, it still needs considerable domain knowledge to accurately identify fault types and locations. In order to realize automatic fault identification, researchers have investigated many intelligent diagnosis algorithms, and combined the extracted signal features with these algorithm [12]–[15].

Manuscript received April 5, 2020; revised July 1, 2020 and August 25, 2020; accepted September 21, 2020. Date of publication October 6, 2020; date of current version August 13, 2021. Recommended by Technical Editor H.-T. Yau and Senior Editor H. Qiao. This work was supported in part by the National Natural Science Foundation of China under Grant 51805352 and in part by the Natural Science Foundation of Shanxi Province under Grant 201901D111062 *(Corresponding author: Xinyu Pang.)*

Xinyu Pang is with the College of Mechanical and Vehicle Engineering, Taiyuan University of Technology, Taiyuan 030024, China (e-mail: typangxy@163.com).

Xuanyi Xue, Wangwang Jiang, and Kaibo Lu are with the College of Mechanical and Vehicle Engineering, Taiyuan University of Technology, Taiyuan 030024, China (e-mail: 1760186560@qq.com; jiangwangwang2017@163.com; lvkaibo4@163.com).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMECH.2020.3029058

Intelligent algorithms, such as support vector machines (SVMs) and artificial neural networks are shallow machine learning models, which need to take features manually extracted as input. Convolutional neural network (CNN) is a deep learning model and can resolve the inherent insufficient expression ability in shallow network. CNN presents strong feature extraction and classification ability, but is prone to overfitting [16]. It has nonlinear transformation capability and can mine deeper information from data, including images. Simultaneously, the network has characteristics of shared weight and translation invariance [17]. In terms of mechanical fault diagnosis, many fault diagnosis methods based on CNN are proposed. For example, a method based on empirical mode decomposition (EMD) and deep CNNs was applied to the fault diagnosis of planetary gearbox [18], deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings [19], fault diagnosis system for induction motors by CNN using empirical wavelet transform [20], fault diagnosis of bearings using envelope spectrum and CNNs [21], fault diagnosis for rotating machinery using multiple sensors and CNNs [22], motor fault diagnosis based on short-time Fourier transform (STFT) and CNN [23]. The earlier methods are applied to convolution neural network through different signal feature extraction ways. Nevertheless, the time-frequency analysis methods employed, such as EMD, wavelet transform, envelope spectrum, STFT, cannot retain the phase information between frequency components and are not suitable for the fault feature extraction of planetary gearbox, thus they will affect the correctness of training samples and satisfactory results in a CNN [24]. Based on the above analysis, BSP analysis is highly appreciated for preprocessing the raw vibration signals of planetary gearbox to obtain high-quality CNN input samples because BSP has the properties of noise reduction and information enhancement.

In the fault diagnosis of planetary gearbox, many issues will also affect the performance of CNN intelligent diagnosis. Limited collected industrial data is a common issue since there are many possible fault profiles. Naturally, it remains difficult to train network models because that the lower quality sample, the deeper layers and the more parameters are required. When target task for identification is repurposed, the CNNs should be trained from beginning, which will result in more time consuming in training. Aiming at tackling these issues, recent studies have attempted to put forward deep CNN-based transfer learning for fault diagnosis. Transfer learning is a new machine-learning method that uses existing knowledge to solve new problems in various fields [25], such as image recognition and text processing. Hasan and Kim [26] introduced transfer learning for bearing fault diagnosis and achieved higher accuracies. To achieve the fault diagnosis of fixed shaft gearbox under variable working conditions with small training samples, an improved deep transfer autoencoder is proposed in [27]. A deep CNN-based transfer learning approach for fixed shaft gearbox fault diagnosis utilizing preprocessing-free raw vibration signals as CNN input is proposed in [24]. These efforts have shown good degrees of success in diagnosing mechanical fault types and severities. However, these progresses in fault diagnosing were made based on 1-D data of vibration from fixed shaft gearboxes. Less attention was paid to monitoring planetary gearboxes which are becoming more widely used in wind turbine generators but fault-related vibration signals are more complicated due to their compact constructions and multiple dynamic interactions.

Aiming at advancing the diagnostic performance of planetary gearbox, this paper proposes an intelligent diagnosis approach. It proposes developing deep CNNs, such as CNN with certain data pre-pressing to enhance the data quality and reduce train difficulties. In particular, it suggests to use BSP to enhance the complicated vibration response from planetary gearboxes and thereby puts forward a BSP CNN model to implement fault diagnostics.

The rest of this article is organized as follows. Section II introduces the theoretical basis of using CNN with BSP image as input and transfer learning for gear fault diagnosis. The model architecture and application steps of the proposed method are explained in Section III. Then, Section IV shows the diagnostic results and discussion by analyzing the experimental data using the proposed approach, and comparisons with respect to different methods are conducted as well. Finally, Section V concludes this article.

II. Theoretical Basis

A. BSP Analysis

In this article, BSP analysis is used to transform 1-D raw vibration signals into 2-D images, which are taken as the inputs of the fault diagnosis model.

The BSP is a third-order statistic, which is defined as a 2-D Fourier transform with third-order autocorrelation. It expresses the correlation between the three spectral elements comprising the spectral values and two frequency components. Moreover, it can reveal the nonlinearity and non-Gaussian properties of the signals [28] that are resulted from the nonlinear interactions of gear engagements.

Let the third-order cumulant be absolutely summable, that is

$$\sum_{\tau_1=-\infty}^{\infty}\sum_{\tau_2=-\infty}^{\infty}|c_{3x}(\tau_1,\tau_2)|<\infty. \tag{1}$$

Then, the BSP expression is

$$B_x(\omega_1,\omega_2)=\sum_{\tau_1=-\infty}^{\infty}\sum_{\tau_2=-\infty}^{\infty}c_{3x}(\tau_1,\tau_2)e^{-j(\omega_1\tau_1+\omega_2\tau_2)}. \tag{2}$$

For a discrete-time signal with limited energy, $x(t)$, its BSP expression is

$$B_x(\omega_1,\omega_2)=X(\omega_1)X(\omega_2)X^*(\omega_1+\omega_2) \tag{3}$$

where $X(\omega)$ is the Fourier transform of signal $x(t)$, and $X^*(\omega)$ is the complex conjugate of $X(\omega)$.

BSP keeps the information of the signal phase and amplitude. Thus, it is a complex function having period and symmetry. BSP analysis only suppresses the linear-phase information and keeps the nonlinear-phase information, whereas the power spectrum suppresses all phase information. In the Gaussian process, $B_x(\omega_1,\omega_2)=0$ [29]. This critical property allows the multiple

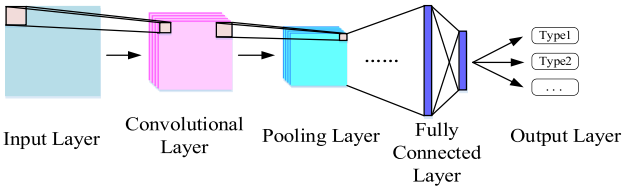


Fig. 1. Typical structure of a CNN model.

modulations of planetary gearbox to be highlighted. Simultaneously the random noises are suppressed. These important properties will be critical to develop CNN models efficiently in terms of less samples required and shorter time taken for training.

B. Pre-Trained Model-CNN

By learning 2-D signals layer-by-layer, CNN continuously transforms and extracts the underlying features to extract different signal features at different levels. It finally forms the highly abstract distributed features to reveal the essential characteristics hidden in the original signal [30]. A typical CNN structure is shown in Fig. 1, which includes an input layer, a convolutional layer, a pooling layer, fully connected layer, and output layer [17]. Compared with single or double hidden layer BP neural networks, CNN has a much deeper layer [31].

An input layer is used to accept input images with predefined labels as valid training samples. The quality of input images will not only affect the amount of training data, the total amount of layers used in CNN, but also the performance, such as accuracy and efficiency, of specific CNN architecture.

The convolutional layer is the feature extraction layer. The neurons on the convolutional layer are connected to the local receptive field of the previous hidden layer through the convolution kernel. The output of the convolutional layer is the extracted feature. The convolutional layer usually has several convolution kernels, which are weight matrices. Different convolution kernels get different weights through network training. Thus, different features can be extracted from a single input. The operation expression of convolutional layer is

$$x_j^l = f\left(\sum_{i=1}^{M} x_i^{l-1} * k_{ij}^l + b_j^l\right), j = 1, \cdots, N \tag{4}$$

where l is the current number of layers, M is the number of input feature graphs, x_j^l is the jth feature graph of layer l, x_i^{l-1} is the i-th feature graph of layer l-1, N is the number of convolution filters, k_{ij}^l is the weight matrix of convolution kernel, b_j^l represents the bias term corresponding to the jth feature graph at level l, and f is the activation function, and $*$ denotes the convolutional operation [32]. The amount of parameters of the convolutional layer can be obtained by

$$P = N \times (s \times s \times M + 1) \tag{5}$$

where $s \times s$ is the size of the convolution kernel.

Batch normalization (BN) layer results in normalized batches of data. When training the network, a change of the input value distribution in each layer will make the network training more complex and slower. BN can solve the instability of gradient diffusion or explosion in a deep network to a certain extent and accelerate the convergence speed of the model. The expression of BN is given as

$$x_i^l = \gamma \frac{x_i^{l-1} - \mu}{\sigma} + \beta \tag{6}$$

where μ is the mean value ofx_i^{l-1},σ is the standard variance of x_i^{l-1},γ is the scale factor, and $\beta$is the shift [33].

The rectified linear unit (ReLU) layer is used as the activation function to increase the sparsity of the network, to improve the generalization ability of the network, and to accelerate the convergence speed of the network [34]. Its expression is shown as

$$x_i^l = \max(0, x_i^{l-1}). \tag{7}$$

The pooling layer is a feature-mapping layer that reduces the complexity of network computing by down-sampling the extracted features. Sampling methods include max pooling and mean pooling. Max pooling calculates the maximum value of each sampling area as an output of the pooling layer, and mean pooling takes the mean value of each sampling area as an output of the pooling layer. Generally, mean pooling retains more background information of images, whereas max pooling retains more texture information. For the purposes of this article, max pooling is adopted. The calculation formula of the output feature map of the lth layer is as follows:

$$x_j^l = f\left(w_j^l down(x_j^{l-1}) + b_j^l\right), j = 1, \cdots, M \tag{8}$$

where w_j^l represents the weight of the jth feature image output at the lth layer, and $down()$ represents the down-sampling function [35].

The purpose of the fully connected layer is to map the distributed feature representation to sample space labels. Each neuron in the fully connected layer is connected from all neurons in the previous layer. The mapped feature can be classified using a SoftMax classifier. The output vector of the lth layer is [36]

$$x_j^l = f\left(\sum_{i=1}^{M} x_i^{l-1} \times w_{ij}^l + b_j^l\right), j = 1, \cdots, N \tag{9}$$

where M is the dimension of the input vector, N is the dimension of the output vector, and w_{ij}^l is the weight of the jth output connected to the ith input. The number of parameters of the fully connected layer can be calculated with (10). SoftMax is the most commonly used CNN classifier. The output expression of SoftMax regression is shown in (11) [37]

$$P = N \times M + 1 \tag{10}$$

$$q(x_j^l) = \frac{e^{x_j^{l-1}}}{\sum_{k=1}^{N} e^{x_k^{l-1}}}. \tag{11}$$

CNN training requires an optimization algorithm to accelerate its learning rate. The stochastic gradient descent (SGD) is the most common optimization algorithm, and it is easy to converge to the local optimal solution. One disadvantage of SGD is that

its update direction completely depends on the current batch. Thus, its update is very unstable. In this article, the SGD-with-momentum (SGDM) algorithm is adopted. The momentum accelerates the SGD in the relevant direction, increases stability and accelerates convergence [38].

C. Transfer Learning

Transfer learning is a machine-learning method that transfers knowledge from one task to another [39]. Transfer learning represents the capability of knowledge transformation between different fields or tasks [40]. The tasks of traditional machine learning are independent of one another, whereas the tasks of various source domain in transfer learning are no longer independent. The knowledge related to the target task is mined from different data concerning different source tasks to help target-task learning. In CNN terms, that is the layers at the convolutional stages of the CNN trained on source dataset indeed extract general features of input, while the layers of fully connected stages are more specific to target task which should be trained by fine-tuning pretrained network using target domain dataset [24].

There are two basic concepts of transfer learning: field and task. The existing knowledge is called the *source field*, and the new knowledge to be learned is called the *target field*. The task refers to the model needed to solve the problem.

1) "Field D" comprises two parts: the feature space χ and the marginal probability distribution $P(x)$, where $X = \{x_1, x_2, \cdots, x_n\} \in \chi$. Thus,$X = \{\chi, P(X)\}$.
2) "Task T" comprises the label category space γ and the target prediction function $f(x)$, which can be learned from the training dataset and used for a new sample x for category prediction. The prediction result is $f(x)$, where $T = \{\gamma, f(x)\}$.

The purpose of transfer learning is to improve the ability of the target prediction function, $f_t(x)$ in target-task T_t by using the knowledge of source field $D_s$and source-taskT_s.For example, the prediction classifier function of sun gear fault diagnosis can be improved by using planet gear health conditions as source domain and planet gear fault diagnosis as source target, and vice versa.

D. Fine-Tuning Pretrained Network

In this article, deep transfer learning is used to initialize and update the weights of the fully connected layer. The steps of fine-tuning the pretrained network algorithm are as follows:

1) Load the pretrained network and target field dataset.
2) Build a transfer network consisting of a transfer layer and a new layer. The layers in front of the fully connected layer of the pretrained network are extracted as the transfer layer. The parameters used in the fully connected layer, the SoftMax layer, and the classification output layer of the pretrained network are trained using the new training data, and the layers number of the fully connected layer equals to the class number of the new data.
3) Fine-tuning the hyperparameter: A, reduce the initial learning rate, slow down the learning rate of the transfer layer; B, increase the learning-rate factor of the fully connected layer, speed up the learning rate of the new layer; C, reduce the number of training rounds.
4) Train and verify the network in the target domain. The target domain training dataset is used to train the deep network model. The SGDM algorithm is used to optimize the model and update the weight, record the model error and classification accuracy, and complete the training. The training process is repeated until the final classification accuracy is no longer significantly improved. After the training, the network parameters are fixed, and the trained fault diagnosis model is obtained.

TABLE I
SPECIFICATIONS OF THE PROPOSED ARCHITECTURE

Stage	Layer	Type	Parameter
1	1	Image Input	344×344×3
2(transferred)	2	Convolution	F=3; N=8; S=1; C=1
	3	Batch Normalization	N=8
	4	ReLU	N=8
	5	Max Pooling	F=2; N=8; S=2
3(transferred)	6	Convolution	F=3; N=16; S=1; C=8
	7	Batch Normalization	N=16
	8	ReLU	N=16
	9	Max Pooling	F=2; N=16; S=2
4(transferred)	10	Convolution	F=3; N=32; S=1; C=16
	11	Batch Normalization	N=32
	12	ReLU	N=32
5(to be trained)	13	Fully Connected	D=5
	14	Softmax	D=5
	15	Classification Output	D=5

III. CNN-TL ARCHITECTURE FOR PLANETARY GEARBOXES FAULT DIAGNOSIS

Based on a study of the vibration BSP under baseline conditions, CNN-TL model architectures are configured for implementing monitoring tasks of the planetary gearbox detailed in Section IV. The specific architecture along with corresponding specifications is given in Table I. A total of 5 stages (15 layers) are used in the proposed architecture. Stage 1 uses BSP images of the vibration signals as inputs. The stage 2 to stage 4 serve as a general well-trained tool for feature extraction, the parameters and specifications used in these stages (the layers 2 to 12) can be transferred to facilitate gear fault diagnosis. The final stage is left to be trained as a classifier using the experimental data specific to the fault diagnosis task.

F is the size of the filters, *N* is the number of filters, *S* is the stride step size, *C* is the number of channels, and *D* is the number of nodes

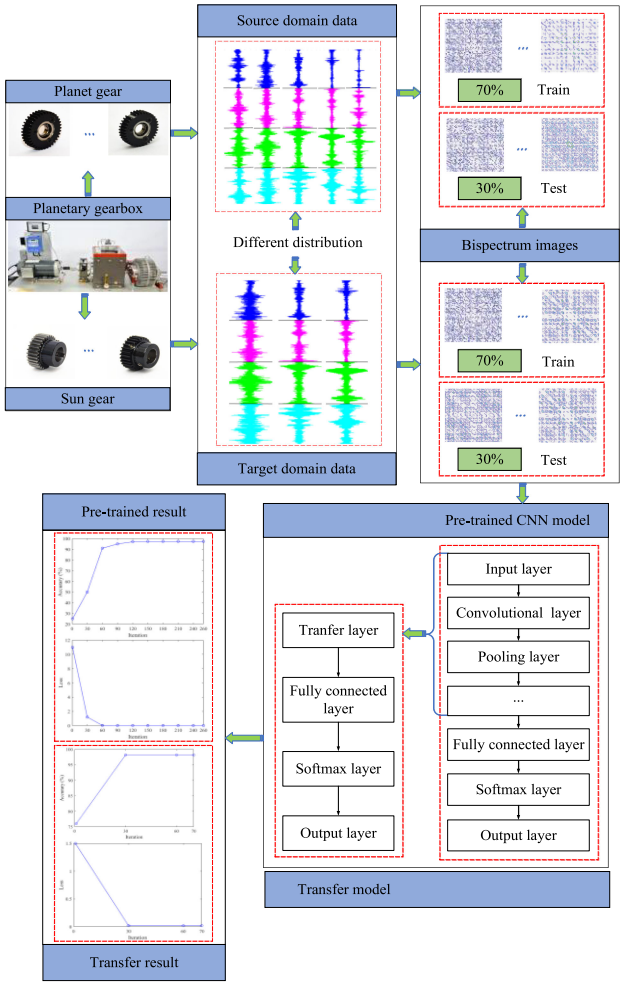


Fig. 2. Schematic diagram illustrating the process of the proposed approach.

The specific process of the proposed approach is shown in Fig. 2.

1) *Acquire the vibration signals in the source and target fields*: Vibration signals based on different single type of gear fault under different loads and speeds are collected using one accelerometer mounted on the gearbox casing. The source domain data is from planet gear, and the target domain data is from sun gear.
2) *Obtain training and testing BSP images*: Each sample is obtained by dividing the data collected in 1) into fixed-length samples, and there is not overlap among these samples. Then, the sample data is processed using BSP analysis so as to achieve BSP image.
3) *Pretrained CNNs and testing*: The pretrained CNN used in this article comprises three convolutional layers, two pooling layers, one fully connected layer, one SoftMax layer, and one classification output layer. The output of three convolutional layers are normalized in batches and activated by applying ReLU operation to accelerate the training of the network and to improve its generalization capability. The parameters of the model are given in Table I.

TABLE II
TRAINING PARAMETERS OF CNN

Parameters	Values
Minibatch size	128
Learning rate	0.01
L2Regularization	0.0001
Max epochs	20
Validation frequency	30
Momentum	0.9
Weight learn rate factor	10
Bias learn rate factor	10

The size of the input BSP image is 344×344×3, herein 3 represents number of color channels. After a large number of comparisons for convolution kernels selection, 3×3 convolution kernels were selected for these three convolutional layers with the kernels number of 8, 16, and 32, respectively. Stride is 1. All convolutional layers are edge-processed using "same." The size of the max pooling filter is 2 × 2, and the numbers of the max pooling filter are 8 and 16, respectively, as given in Table1. Padding is set to 0 to the boundaries of the input matrix to avoid size changes.

The setting of training parameters is very important to the recognition accuracy of CNN. Minibatch size is the number of samples contained in one batch. The larger the minibatch, the faster the data is processed. Additionally, more epochs are needed to achieve the same precision. Minibatch size is usually set to 2^n, which is 128 in this article. The learning rate is set to 0.01. The regular term of the fully connected layer is set to L2 to avoid overfitting, and the regularization coefficient is 0.0001. The SGDM momentum value is set to 0.9. During the training process, the network is verified every 30 iterations. The learning-rate factors, weight learn rate factor, and bias learn rate factor are all 10. The number of epoch represents the number of times required to train all datasets. The epoch is set to 20. The detailed training parameter of the pretrained network are given in Table II [17], which is configured with MATLAB 2019a.

4) *The BSP of the target field is used to fine-tuning the pretrained network, and the transfer network is obtained and tested*: Transfer learning is achieved by extracting all the transferred layers in pre-trained CNN, and replacing the last three layers of the pretrained network using fine-tuning process. The parameters *D* of these last three layers are set to the number of sun gear faults types, and weight learn rate factor and bias learn rate factor are set to 20. To avoid overfitting, a random dropout layer is added after the transfer layer, and the dropout rate is 0.6. Initial learning rate is 0.001, and the epoch is set to 10. Other hyperparameter settings are the same as those in the pre-trained network. The layer and parameters of the transfer network are given in Table III, and the settings of training hyperparameter are given in Table IV.

TABLE III
LAYER AND PARAMETERS OF THE TRANSFER NETWORK

Layer	Type	Parameters
1	Transfer	---
2	Dropout	R=0.6
3	Fully Connected	D=3
4	SoftMax	D=3
5	Classification Output	D=3

R is the dropout rate and *D* is the number of nodes.

TABLE IV
TRAINING HYPERPARAMETERS OF THE TRANSFER NETWORK

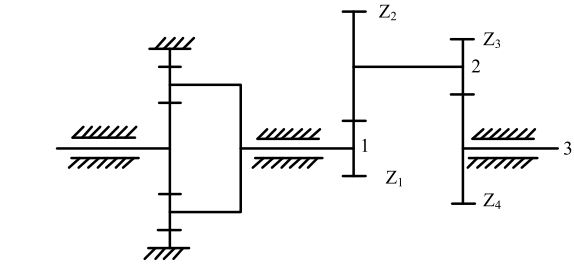
Hyperparameters	Values
Minibatch size	128
Learning rate	0.001
L2Regularization	0.0001
Max epochs	10
Validation Frequency	30
Momentum	0.9
Weight learn rate factor	20
Bias learn rate factor	20

IV. EXPERIMENTAL EVALUATION

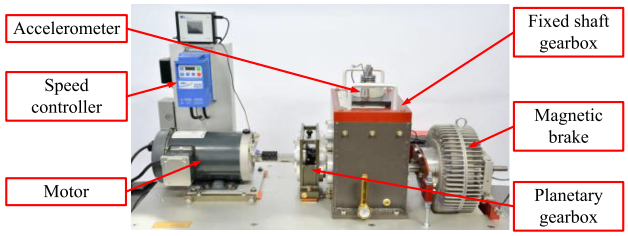
This section focuses on evaluating the proposed method using datasets from a drivetrain dynamics simulator (DDS). As mentioned in Section II-C, the case in which the deep CNN-TL can transfer planet gear fault knowledge to a sun gear fault is examined. First, the BSP of the measured vibration signal is obtained. Then, the CNN model is trained using planet gear image data to obtain the pretrained network. Next, the pretrained network is fine-tuned using the sun gear data to obtain the transfer network. Finally, the performance of the method is discussed.

A. Experimental Setup

A DDS manufactured by SpectraQuest Company is applied. The transmission schematic and the physical model are shown in Fig. 3. The planetary gearbox used in the test is a single planetary gearing system that includes four planet gears. The ring gear is stationary, the sun gear is the input, and the planetary carrier is the output. The reduction ratio of the whole planet gear system is 4.571. The relevant parameters of the planetary gearbox is given in Table V. As shown in Fig. 4, five different gear states (i.e., normal, tooth root crack, tooth wear, tooth breakage, and missing tooth) are preset for the planet gear, and three different gear conditions (i.e., tooth root crack, tooth wear, and tooth breakage) are introduced to the sun gear. The vibration signals of the planet gears are collected using a vibration acceleration sensor. The sampling frequency is 10 kHz. Two different torques (0 and 40 N.m) and two input speeds (1800 and 2400 r/min) are set to form four operational conditions. Vibration and acceleration signals of eight types of faults under four operating states are



(a)



(b)

Fig. 3. DDS test bench. (a) Transmission schematic of the DDS gearbox. (b) Physical model of the DDS gearbox.

TABLE V
PARAMETERS OF PLANETARY GEARBOX

Gears	Sun gears	Planet gears	Gear ring	Z_1	Z_2	Z_3	Z_4
Number of teeth	Z_S=28	Z_P=36	Z_r=100	29	100	36	90

collected, respectively. A total of 120 000 points are collected for each operating state, then equally divided into 120 data samples. The number of points in a single sample is 1000. Therefore, the number of samples of each fault is 480. The details about these datasets are given in Table VI. Source domain data from these five kinds of planet gear states are used to train CNN to obtain a pre-trained network, whereas the data of these three kinds of sun gear faults are used as target domain data to train the transfer network.

B. BSP Analysis

The vibration signals corresponding to the planet gears of five health conditions and the sun gears of three faults under four typical working conditions: 1800 r/min with 0 N.m, 1800 r/min with 40 N.m, 2400 r/min with 0 N.m, 2400 r/min with 40 N.m were collected. Figs. 5 and 6 show the BSP images of planet gear and that of sun gear with different health condition under the same operation condition of 1800 r/min with 0 N.m, respectively. It can be observed that there are distinctive differences among these BSP distribution regions and distribution patterns. They reflect the dynamic behaviors and modulation components corresponding to the different fault type. This is in agreement with the model prediction and of that of previous studies [41],

Fig. 4. Photos of the gears. (a) Photos of 5 planet gears. (b) Photos of 3 sun gears.

TABLE VI
STATES OF PLANETARY GEARBOX

Planetary gearbox health states	Categories	The number of samples
Normal	Class 1	480
Tooth root crack fault of the planet gear	Class 2	480
Teeth wear fault of the planet gear	Class 3	480
Tooth breakage fault of the planet gear	Class 4	480
Missing tooth fault of the planet gear	Class 5	480
Tooth root crack fault of the sun gear	Class 6	480
Tooth wear fault of the sun gear	Class 7	480
Tooth breakage fault of the sun gear	Class 8	480

[42]. These enhanced differences will help to develop a more reliable CNN model.

In particular, for the planet gear cases there is not any apparent peaks in normal state, whereas a number of distinctive peak appear near image center for tooth root crack, these peaks are distributed along diagonal; some BSP peaks present with the

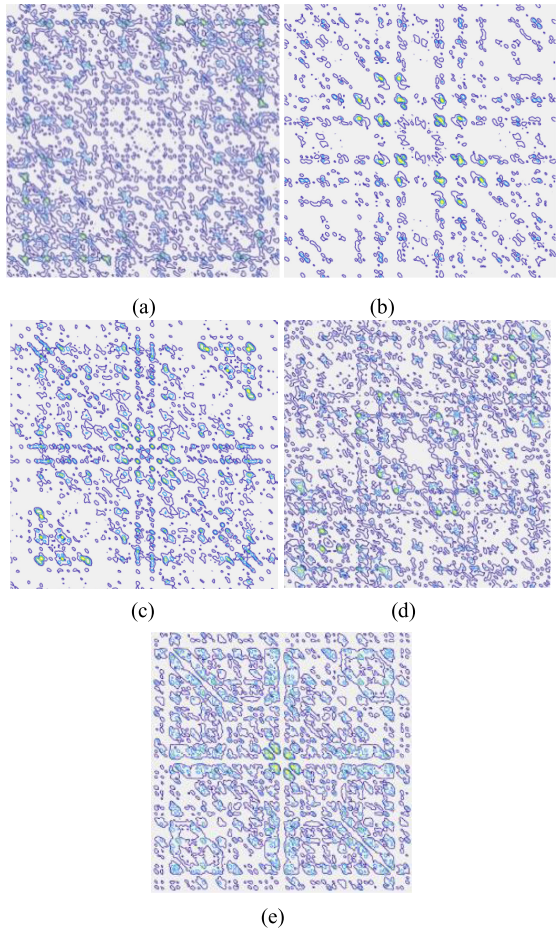


Fig. 5. BSP images of vibration for different faults on planet gears. (a) Normal. (b) Tooth root crack. (c) Tooth wear. (d) Tooth breakage. (e) Missing tooth.

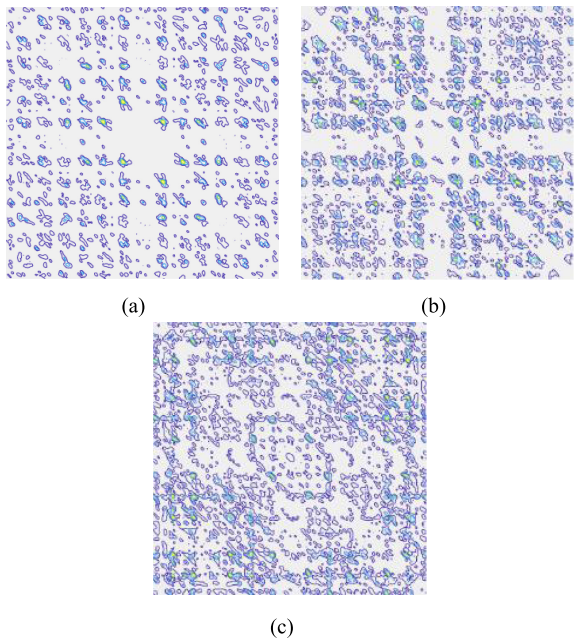


Fig. 6. BSP images of vibration for different faults on sun gears. (a) Tooth root crack. (b) Tooth wear. (c) Tooth breakage.

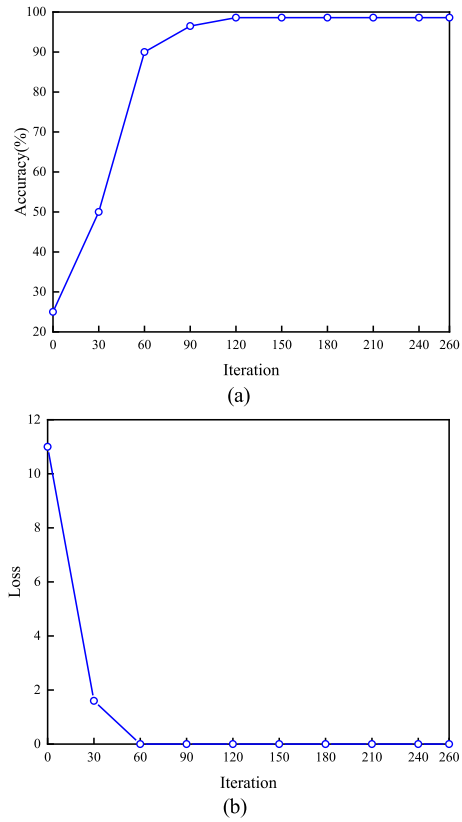


Fig. 7. Verification process of pre-trained network. (a) Accuracy. (b) Loss.

distribution of both center and ends of diagonal due to tooth wear; the peaks distribution regions for tooth breakage and missing tooth are similar to that of tooth wear, but they have more large distribution range. The distribution range of missing tooth is wider than that of tooth breakage, and the higher peaks are distributed at the center of image. Furthermore, although it is hard to describe in words, it also can be observed clearly from these graphs that BSP distribution patterns are different corresponding to different fault type. Therefore, the proposed method is used to characterize high-level abstractions for better achievements in accuracy and efficiency. For sun gear, BSP images present similar appearance to those of planet gear.

C. Fault Diagnosis Results

Of the planet gear test datasets, 70% are randomly selected for training and the rest (30%) are for testing. The verification processes of the pretrained model are shown in Fig. 7. The figure shows that verification process gradually converges in terms of both accuracy and loss as the increase of iterations. The accuracy increases from 50% to 95% with the iteration number increasing from 30 to 90, and when the number of iterations is greater than 120, the classification accuracy reaches the plateau and can achieve its maximum of 97.36%. The loss exhibits a decreasing trend and finally remain stable at very lower level near 0. The confusion matrix is shown in Fig. 8, where the horizontal axis represents the predicted fault type of the planet gear, and the vertical axis represents the real fault type

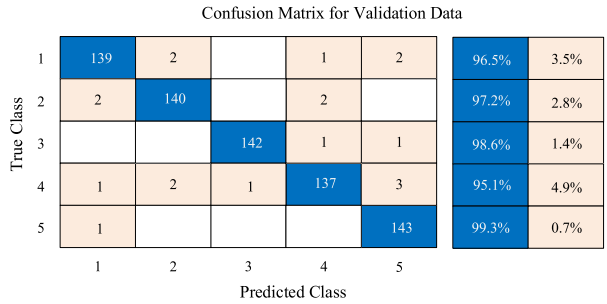


Fig. 8. Validation data confusion matrix of pretrained network.

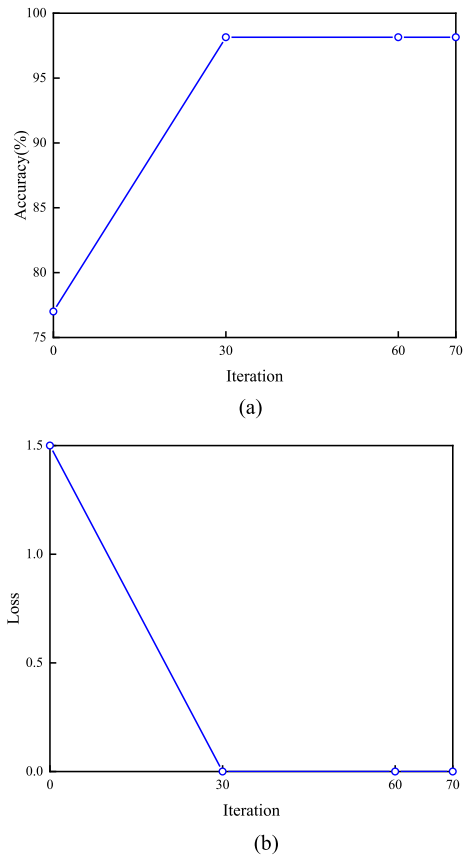


Fig. 9. Training processing of transfer network. (a) Accuracy. (b) Loss.

of the planet gear. As can be seen from Fig. 8, the classification accuracy of missing tooth is 99.3%, and that of tooth wear is 98.6%. The identification rate of normal condition and tooth root crack are slightly lower, 96.5% and 97.2%, respectively. And even the identification rate of tooth breakage is the lowest, it still reaches 95.1%.

Of the sun gear fault data, 70% are randomly selected for training and the rest (30%) are used for validation. The validation process of the transfer model is shown in Fig. 9. The classification accuracy of the validation data reaches 98.15%, illustrating that the transfer network obtained by fine-tuning the pre-trained network can achieve a high diagnosis rate. Furthermore, it is higher than 97.36% of planet gear, which illustrates that the proposed approach can improve the ability of the target

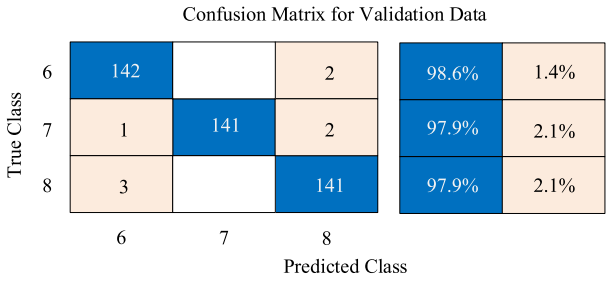


Fig. 10. Validation data confusion matrix of transfer learning.

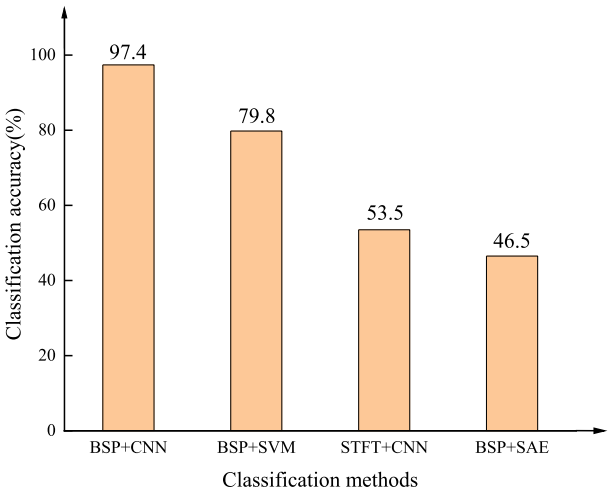


Fig. 11. Comparison of classification accuracy.

prediction function, agreement with the proposed analysis made in Section II. C. Compared with the verification process of the pre-trained network, it can be seen that the pretrained network can achieve convergence at iteration number of 120, whereas the transfer learning can converge after 30 times iteration, indicating that the transfer network consumes less validation time than that of pretrained network. This shows that the transfer network has a higher identification rate and better efficiency. The confusion matrix is shown in Fig. 10, where the horizontal axis represents the predicted fault types of the sun gear, and the vertical axis represents the real fault types of the sun gear. As can be seen from Fig. 10, the classification accuracy of tooth root crack, tooth wear, and tooth breakage of the sun gear is 98.6%, 97.9%, and 97.9%, respectively.

D. Comparative Studies

To further confirm the performance of proposed method, several different diagnosis methods are compared in the application of planet gear fault recognition. The BSP analysis of signals of the planet gear in five different states is carried out and the BSP image features are extracted. These features are respectively fed into a CNN, stack autoencoder (SAE), and SVM learning models for testing. The recognition rates of comparison methods are shown in Fig. 11. As can be seen, the accuracy of BSP + CNN is the highest, reaching 97.4%, showing that with

TABLE VII
COMPARISON OF TRANSFER LEARNING AND OTHER ALGORITHMS

Methods	Accuracy	Time
Deep CNN-TL	98.15%	46s
CNN	96.30%	95s
SVM	88.19%	60s

BSP as input, the CNN method has the best recognition effect than traditional machine learning. The main reason is that the sparsity of SAE results in loss of important information, SVM is more suitable to process dataset with small samples as shallow learning method, while CNN is a powerful tool in pattern recognition and data mining [24]. Additionally, the vibration signals are processed and imported into the CNN for testing with the STFT time-frequency diagram for comparison. Fig. 11 shows that the recognition rate of STFT + CNN is 53.5%, far lower than that of BSP+CNN. Because the fixed window width of STFT cannot satisfy the time and frequency resolution requirements simultaneously. Which also indicates that the BSP can better reflect the features of different faults than STFT. These results confirm that BSP+CNN can accurately classify planetary gearbox faults. Based on the above analysis, it is the CNN model with BSP analysis to be selected as the pre-trained network for TL.

To verify the advantages of deep CNN-TL, three methods are compared to identify the sun gear faults. These methods are deep CNN-TL, CNN, and SVM. Deep CNN-TL is the transfer-learning model through pre-trained the CNN of the planet gear. The identification results are given in Table VII. It can be seen that the identification rate of deep CNN-TL is higher than that of the other two methods, and the computational time is less. The comparison results show that deep transfer learning can effectively share and transfer information between planet and sun gears.

To the further verify the superiority of the proposed model, comparisons were also made with respect to approaches proposed in [24] and [27]. The accuracy of these two models is 85.37% and 92.45%, and the time consumed is 132s and 72s. This should not come as a surprise. As mentioned in [24], correctness of training samples will affect results significantly. Raw vibration signals usually consist of high random noises and interferences. CNN trained using such data would results in poorer identification accuracy even with harder training efforts, which will also lead to lower performance for the transfer learning. As the proposed approach in [27] is based on the two multiwavelet scaling functions, there does not seem to be a consensus on what configuration of parameters such as the scale factor and shift factor to use for gear fault diagnosis. The results indicate that the proposed approach shows higher accuracy and needs less computational time in comparison with than these two models dealing with the same transfer diagnosis task. Therefore, it can be concluded that the proposed approach shows better satisfactory outcomes in the regard of identification accuracy and computational efficiency for planetary gearbox fault diagnostics.

V. Conclusion

In this article, a BSP-based deep CNN is developed to diagnose the faults in a planetary gearbox which has a much complicated dynamic responses. Because BSP can enhance the diagnostic features, the proposed approach greatly improves the training and performance of CNN and TL processes. Consequently, it allows a number of common fault cases including wear, tooth breakages in a planetary gearbox to be diagnosed with higher degree of accuracy. The corrected rate is 97.4% for direct fault diagnosis and 98.15% for transfer learning based diagnostics. Comparatively, other compound methods such as stack autoencoders, SVM combined with STFT produces a much lower correct rate.

References

[1] J. McNames, "Fourier series analysis of epicyclic gearbox vibration," *J. Vib. Acoust.*, vol. 124, no. 1, pp. 150–152, Jan. 2001.

[2] W. Bartelmus and R. Zimroz, "Vibration condition monitoring of planetary gearbox under varying external load," *Mech. Syst. Signal Process.*, vol. 23, no. 1, pp. 246–257, Jan. 2009.

[3] A. S. Sait and Y. I. Sharaf-Eldeen, "A review of gearbox condition monitoring based on vibration analysis techniques diagnostics and prognostics," *Shock Vib.*, vol. 5, pp. 307–324, Mar. 2011.

[4] R. Bajrić and D. Sprečić, "Review of vibration signal processing techniques towards gear pairs damage identification," *Int. J. Eng. Technol.*, vol. 11, no. 4, pp. 124–128, Jan. 2011.

[5] S. Yang, B. Tian, and R. Zhou, "A jamming identification method against radar deception based on bispectrum analysis and fractal dimension," *Jiaotong Univ*., vol. 50, no. 12, pp. 128–135, Dec. 2016.

[6] N. Mahmoodian, A. Schaufler, A. Pashazadeh, A. Boese, and M. Friebe, "Proximal detection of guide wire perforation using feature extraction from bispectral audio signal analysis combined with machine learning," *Comput. Biol. Med.*, vol. 107, pp. 10–17, Apr. 2019.

[7] S. Li and Y. Liu, "Feature extraction of lung sounds based on bispectrum analysis," in *Proc. Int. Symp. Inf. Process.*, Qingdao, China, 2010, pp. 393–397.

[8] E. B. Assi, L. Gagliano, S. Rihana, D. K. Nguyen, and M. Sawan, "Bispectrum features and multilayer perceptron classifier to enhance seizure prediction," *Sci. Rep.*, vol. 8, pp. 1–8, Oct. 2018.

[9] H. Li, Z. Tian, H. Yu, and B. Xu, "Fault prognosis of hydraulic pump based on bispectrum entropy and deep belief network," *Meas. Sci. Rev.*, vol. 19, no. 5, pp. 195–203, Oct. 2019.

[10] I. M. Howard, "Higher-order spectral techniques for machine vibration condition monitoring," *J. Aerosp. Eng.*, vol. 211, no. 4, pp. 211–219, Jun. 1997.

[11] L. Jiang, Y. Liu, X. Li, and S. Tang, "Using bispectral distribution as a feature for rotating machinery fault diagnosis," *Measurement*, vol. 44, no. 7, pp. 1284–1292, Jun. 2011.

[12] S. Abbasion, A. Rafsajani, A. Farshidianfar, and N. Irani, "Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine," *Mech. Syst. Signal Process.*, vol. 21, no. 7, pp. 2933–2945, Oct. 2007.

[13] J. B. Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech, "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals," *Appl. Acoust.*, vol. 89, pp. 16–27, Mar. 2015.

[14] X. Yan and M. Jia, "A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," *Neurocomputing*, vol. 313, no. 3, pp. 47–64, Nov. 2018.

[15] G. Pan, S. Li, H. Du, and Y. Zhu, "Feature extracting method for gearbox tooth breakage under impact based on the S-transform time-frequency spectrum combined with the denoising by SVD," *J. Vib. Shock*, vol. 38, no. 18, pp. 256–263, Oct. 2019.

[16] S. Zhang, M. Wang, W. Li, and J. Luo, "Deep learning with emerging new labels for fault diagnosis," *IEEE Access*, vol. 7, pp. 6279–6286, Jan. 2019.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, pp. 84–90, Jun. 2017.

[18] N. Hu, H. Chen, L. Zhang, and Y. Zhang, "Fault diagnosis for planetary gearbox based on EMD and deep convolutional neural networks," *J. Mech. Eng.*, vol. 55 no. 7, pp. 9–18, Jun. 2019.

[19] D. Verstraete, A. Ferrada, E. L. Droguett, V. Meruane, and M. Modarres, "Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings," *Shock Vib.*, vol. 6, no. 6, pp. 1–17, 2017.

[20] Y. M. Hsueh, V. R. Ittangihal, W. Wu, H. Chang, and C. Kuo, "Fault diagnosis system for induction motors by CNN using empirical wavelet transform," *Symmetry*, vol. 11. no. 1212, pp. 1–11, Sep. 2019.

[21] D. K. Appana, A. Prosvirin, and J. M. Kim, "Reliable fault diagnosis of bearings with varying rotational speeds using envelope spectrum and convolution neural networks," *Soft Comput*, vol. 22, no. 20, pp. 6719–6729, Oct. 2018.

[22] M. Xia, T. Li, L. Xu, L. Liu, and C. W. Silva, "Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks," *IEEE-ASME Trans. Mech.*, vol. 23, no. 1, pp. 1083–4435, Feb. 2017.

[23] L. Wang, X. Zhao, J. Wu, Y. Xie, and Y. Zhang, "Motor fault diagnosis based on short-time fourier transform and convolutional neural network," *Chin. J. Mech. Eng.*, vol. 30, no. 6, pp. 1357–1368, Jun. 2017.

[24] P. Cao, S. Zhang, and J. Tang, "Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning," *IEEE Access*, vol. 6, pp. 26241–26252, May 2018.

[25] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[26] M. J. Hasan and J. M. Kim, "Bearing fault diagnosis under variable rotational speeds using Stockwell transform-based vibration imaging and transfer learning," *Appl. Sci-Basel*, vol. 8, no. 12, pp. 1–15, Dec. 2018.

[27] Z. He, H. Shao, X. Zhang, J. Cheng, and Y. Yang, "Improved deep transfer auto-encoder for fault diagnosis of gearbox under variable working conditions with small training samples," *IEEE Access*, vol. 7, pp. 115368–115377, Aug. 2019.

[28] L. Saidi, "The deterministic bispectrum of coupled harmonic random signals and its application to rotor faults diagnosis considering noise immunity," *Appl. Acoust.*, vol. vol. 122, pp. 72–87, Mar. 2017.

[29] L. Jiang, Y. Liu, X. Li, and S. Tang, "Using bispectral distribution as a feature for rotating machinery fault diagnosis," *Measurement*, vol. 44, no. 7, pp. 1284–1292, Jun. 2011.

[30] K. Adem, "Exudate detection for diabetic retinopathy with circular Hough transformation and convolutional neural networks," *Expert Syst. Appl.*, vol. 114, pp. 289–295, Dec. 2018.

[31] D. Zhao, T. Wang, and F. Chu, "Deep convolutional neural network based planet bearing fault classification," *Comput. Ind.*, vol. 107, pp. 59–56, May 2019.

[32] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[33] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[34] O. Kechagias, "Target recognition for synthetic aperture radar imagery based on convolutional neural network feature fusion," *J. Appl. Remote Sens.*, vol. 13, no. 1, Jan. 2019.

[35] S. Aysum and S. Hasan, "Deep convolutional neural network-based automatic classification of neonatal hip ultrasound images: A novel data augmentation approach with speckle noise reduction," *Ultrasound Med. Biol.*, vol. 46, pp. 735–749, Dec. 2019.

[36] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[37] V. J. Kadam, S. M. Jadhav, and K. Vijayakumar, "Breast cancer diagnosis using feature ensemble learning based on stacked sparse autoencoders and softmax regression," *J. Med. Syst.*, pp. 43–263, Aug. 2019.

[38] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proc. 19th Int. Conf. Comput. Stat*., 2010, pp. 177–186.

[39] H. Daume and D. Marcu, "Domain adaptation for statistical classifiers," *J. Artif. Intell. Res.*, vol. 26, pp. 101–126, Sep. 2011.

[40] F. Wei, J. Zhang, C. Yan, and J. Yang., "FSFP: Transfer learning from long texts to the short," *Appl. Math. Inform. Sci.*, vol. 8, no. 4, pp. 2033–2040, Jul. 2014.

[41] H. Liu, J. S. Dhupia, and S. G. Sheng, "An explanation of frequency features enabling detection of faults in equally spaced planetary gearbox," *Mech. Mach. Theory*, vol. 73, ppp. 169–183, Mar. 2014.

[42] M. Inalpolat and A. Kahraman, "A theoretical and experimental investigation of modulation sidebands of planetary gear sets," *J. Sound Vib.*, vol. 323, no. 3, pp. 677–696, Jan. 2009.

Xinyu Pang was born in Lvliang, Shanxi, China, in 1976. She received the Ph.D. degree in mechanical design and theory from the Taiyuan University of Technology, Taiyuan, China in 2011.

From November 2013 to May 2014, she was a Visiting Scholar with the University of Hertfordshire, Hatfield, U.K. From 2014 to 2019, she was employed as an Associate Professor with the College of Mechanical and Vehicle Engineering of Taiyuan University of Technology. Her main research interests include mechanical vibration and testing, mechanical fault diagnosis, oil monitoring and life prediction, etc. She published an academic paper named "A fault feature extraction method for a gearbox with a composite gear train based on EEMD and translation-invariant multiwavelet neighboring coefficients" in the Strojniski Vestnik-Journal of Mechanical Engineering, in 2019.

Dr. Pang's is a Member of the Chinese Society for Vibration Engineering and the Deputy Secretary-General of the Shanxi Society for Vibration Engineering.

Xuanyi Xue was born in Yuncheng, Shanxi, China in 1994. She received the bachelor's degree in mechatronic engineering from the North University of China, Taiyuan, China, in 2017. She is currently a Student with the College of Mechanical and Vehicle Engineering, Taiyuan University of Technology, Taiyuan, China.

Her research direction is mechanical fault diagnosis.

Ms. Xue was a Recipient of a Third-Class Academic Scholarship, in 2018.

Wangwang Jiang was born in January 1995 in Yuncheng, Shanxi Province, China. He received the master's degree in mechanical engineering from Taiyuan University of Technology, Taiyuan, China, in 2020.

Since September 2017, he has published three academic papers, among which he published an academic paper entitled "Investigation on lubrication state of sliding bearings in low-speed rotor system subjected to torque load" in International Journal of Rotating Machinery in 2019. Second, in July 2019, he successfully obtained the financial support from the Education Department of Shanxi Province and undertook a provincial graduate education innovation project named "Study on the influence of side plate mass on the dynamics of frame structure." On January 2019, he also applied for a national invention patent (CN 109948174 A). His main research interests include mechanical vibration and testing.

Mr. Jiang's honors include First Class Scholarship (2019), Third Class Scholarship (2018), and National Inspirational Scholarship (2014–2015, 2013–2014).

Kaibo Lu was born in 1984 in Yuncheng, Shanxi Province, China. He received the Ph.D. degree in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 2014

Since 2014, he has been with the Taiyuan University of Technology, Taiyuan, China. He is currently an Associate Professor with the College of Mechanical and Vehicle Engineering, Taiyuan University of Technology, Taiyuan, China. He has authored or coauthored 21 academic papers, out of which 11 are included in SCI and EI. In 2018, he has authored or coauthored a paper "Model-based chatter stability prediction and detection for the turning of a flexible workpiece" in *Mechanical Systems and Signal Processing*. His research interests include condition monitoring and fault diagnosis, dynamics in machining processes.

Dr. Lu is currently an Academic Visitor with the Center for Efficiency and Performance Engineering of the University of Huddersfield, which has been sponsored by China Scholarship Council.