IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 January 2023, accepted 28 February 2023, date of publication 14 March 2023, date of current version 20 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3257041

RESEARCH ARTICLE

A New Method Based on Encoding Data Probability Density and Convolutional Neural Network for Rotating Machinery Fault Diagnosis

BOWEN ZHANG, XINYU PANG, PENG ZHAO, AND KAIBO LU
College of Mechanical and Vehicle Engineering, Taiyuan University of Technology, Taiyuan 030024, China

Corresponding author: Xinyu Pang (typangxy@163.com)

This work was supported in part by the Key Research and Development Project of Shanxi Province under Grant 202102010101009 and Grant 202102010101006, and in part by the National Natural Science Foundation of China under Grant 52175108.

ABSTRACT In order to apply the advantages of image recognition for fault diagnosis using convolutional neural network (CNN), it is necessary to convert one-dimensional (1D) signal data into two-dimensional (2D) images. Traditional signal-based conversion methods face the challenges of complex feature extraction process, high dependence on expert knowledge and poor feature repeatability, which hinder timely and accurate fault diagnosis. Therefore, this paper proposes a fault diagnosis method for rotating machinery in virtue of Data Probability Density-Gram Angle Field-Convolutional Neural Network (DPD-GAF-CNN). DPD-GAF first computes the DPD of a 1D time series through parameter-free statistics, and then encodes the DPD into a 2D feature image that directly reflects the mean and standard deviation of the probability distribution. Besides the simplified transformation process, no artificially designed features are required like the original GAF to encode in the process of fault diagnosis. After that, the CNN based on LeNet-5 transformation is used to achieve high-precision fault classification. The proposed fault diagnosis method is verified and compared with other existing intelligent methods using the experimental data generated by the planetary gearbox test bench with various faulty conditions and the bearing data set of Case Western Reserve University. The results show that the presented method can effectively improve the accuracy and stability of fault diagnosis with the classification accuracy of several fault datasets up to 99.9%.

INDEX TERMS Data probability density, GAF, convolutional neural network, fault diagnosis.

I. INTRODUCTION

Fault diagnosis plays a vital role in industrial production. With the automation and complexity of modern industrial systems, however, fault diagnosis methods based on mathematical models [1] and expert experience [2] are often inefficient and complex [3], [4]. In the industrial background of big data, the data-driven fault diagnosis method only needs to collect a large amount of historical data to complete the fault diagnosis [5], avoids formulation of a complex mechanism model and reduces the dependence on expert experience as much as possible. Therefore, data-driven fault diagnosis methods have become the mainstream of research in recent years [6].

At present, deep learning methods based on a large amount of historical data have been widely used in the field of fault diagnosis [7]. The reason is that deep learning overcomes the shortcomings of traditional machine learning-based fault diagnosis methods, and has stronger model generalization ability and simple diagnosis process, the degree of manual intervention is small, and it can process massive data, which can adapt to the high-precision and intelligent requirements of diagnosis in the industrial background of big data [8]. The models of deep learning mainly include Deep Belief Network (DBN), Deep Autoencoder (DAE), Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN). Among them, the CNN is an important branch of the current

The associate editor coordinating the review of this manuscript and approving it for publication was Mehrdad Saif.

deep learning field [9], which has many model architectures, such as LeNet-5 [10], VGGNet [11], AlexNet [12] and GoogleNet [13]. CNNs have attracted much attention from researchers because of their excellent performance in the field of image recognition [14]. The valid data in the field of fault diagnosis is usually 1D time series data, such as vibration signals. While it is possible to directly classify 1D signals by converting existing 2DCNN models to 1DCNNs, CNN models are diverse and increasingly complex.It is easier to convert data into two-dimensional images, and expressing one-dimensional signals with two-dimensional images can intuitively reflect the potential characteristics of sequence signals. Therefore, converting 1D vibration signals into 2D images can effectively apply the advantages of CNNs in the field of vision to mechanical fault diagnosis [15], [16], [17], [18].

Several methods based on signal processing have been applied to convert 1D vibration signals into 2D images, such as the short-time fourier transform [19], continuous wavelet transform [20], wavelet packet transform [21] and other signal processing methods. The generated images are then classified by CNN model to realize fault diagnosis. However, these methods are interfered by human factors: the complex process of feature extraction requires rich expert knowledge, and the time complexity is high and difficult to meet the real-time requirements [22]. Moreover, the features extracted by the signal analysis method using windows have certain limitations: the equipment status information reflected by the signals of different time windows is different, and it is also inevitably affected by different interference factors, resulting in the generation of images based on the signal analysis method [23]. The feature repeatability is poor, which affects the classification accuracy of the CNN model.

In order to simplify the feature extraction process and eliminate the limitations brought by expert knowledge, some researchers directly convert discrete time-domain data into images, and then input them into the CNN model to automatically extract the required features to achieve fault classification [17]. The advantage of this fault diagnosis method is that it does not require expert knowledge or predetermined parameters, has low time complexity, and is easier to achieve real-time high-precision fault diagnosis [24]. Wen [25] et al. converted the 1D signal into a grayscale image through matrix shape change and then input it into a CNN model based on LeNet-5 transformation to achieve fault classification. However, the forced matrix shape transformation destroys the continuity of the signal, which makes CNN have certain limitations in mining the original signal features. Wang and Oates [26] proposed a Gram Angular Field (GAF) encoding method. Different from the forced matrix shape change of grayscale images, GAF encoding constructs a special Gram matrix to generate images from time series data. The constructed matrix not only retains the original time series, but also reflects the correlation between adjacent elements. Thus, it mines the potential information of the data, and more intuitively reflects the difference of images generated by different time series samples, which is conducive to the automatic mining of features by the CNN model to achieve classification. In the field of fault diagnosis, Pang [27] et al. transformed the vibration signal of the planetary gearbox into GASF and GADF feature maps respectively through the GAF coding method, and then input them to the CNN model for classification, and compared and studied the two GAF feature maps. It is found that the GADF feature map is more suitable for planetary gears fault classification. Li [28] et al. converted the time series data into 2D images through the modified GADF encoding method of bearing fault signals, and constructed a simple CNN model to achieve better classification results. Tong [29] et al. transformed bearing fault signals of different lengths into GADF feature maps, and found that the data length had a significant impact on the accuracy of the GADF-CNN diagnostic model. Moreover, the above fault diagnosis cases all require a larger size of the GAF feature map to ensure better classification accuracy, which brings inconvenience to the training of the previous model. Although the idea of local pixel averaging or maximum value can be used to adjust the image size [26], as the data length increases, the information loss will increase exponentially, resulting in excessive image feature loss, which is not conducive to subsequent classification work. Therefore, how to choose an appropriate signal length to make it more conducive to fault classification requires multiple trials in combination with the CNN model, and when the signal length increases to a certain extent, the training process of GAF-CNN will consume a lot of resources. Fortunately, using the statistical feature of Data Probability Density combined with GAF encoding can overcome the shortcomings of the original GAF image encoding. The way of the statistical data density distribution can be based on the simplest histogram algorithm [30], because it is based on the statistical results of non-parametric methods, so there is no artificial intervention information. When the sample size is sufficient, the statistical density feature does not change due to the change of data length, so encoding the 1D data density distribution through GAF will not be limited by the choice of which signal length is better. For long sequences the data can also be transformed into appropriately sized images preserving their density characteristics.

Based on the above analysis, we propose an image conversion method that combines the Data Probability Density and the Gram Angular Field (DPD-GAF) to improve the deficiencies of converting the original signal into a GAF image. Then, the generated DPD-GAF images are input into the CNN model training based on LeNet-5 transformation to realize the classification of fault data and achieve the effect of fault diagnosis. Finally, the proposed fault diagnosis method is verified by the fault data generated by the drivetrain dynamics simulator (DDS) test bench and the bearing test bench of Case Western Reserve University. Comparison with the existing diagnostic models is also performed.

The main contributions of this paper are as follows. First of all, the proposed DPD-GAF image conversion method does not require extensive expert knowledge and too many preset parameters. It can generate images of appropriate size and mine the potential information of the data to highlight the differences among various states, which is convenient for the training and prediction of deep learning algorithm models. DPD eliminates the influence of signal length on image features and simplifies the process of image conversion. Then, the fault diagnosis method combined with DPD-GAF-CNN simplifies the original GAF-CNN fault diagnosis process, and does not need to adjust the length of the original signal multiple times, so that the diagnosis algorithm model can obtain higher recognition accuracy, and can adapt to longer sequences of the sample signal realizes fault diagnosis and has a wider adaptability than the GAF-CNN model. This study uses the well-known Case Western Reserve University bearing dataset and the DDS test bench to simulate the planetary gear fault dataset under constant operating conditions and variable speed and variable load operating conditions to verify the performance of the diagnostic model. And compared with other existing models, it is proved that the proposed method can ensure the accuracy of fault diagnosis and take into account the efficiency of the diagnosis process.

The rest of this paper is organized as follows. Section II presents related work, including the application of DPD in fault diagnosis and the original GAF coding method. Section III presents the proposed DPD-GAF-CNN based fault diagnosis method. It includes the coding method of DPD-GAF, the model structure and parameters of CNN, and the overall diagnosis process. Section IV presents several different fault diagnosis cases and the results of verification and comparison of the proposed fault diagnosis model. Conclusions and outlook are given in Section V.

II. RELATED WORK

This section mainly introduces related work, including two aspects: the application of Data Probability Density (DPD) in fault diagnosis and condition monitoring; and the Gram Angular Field (GAF) image coding method.

A. APPLICATION OF DATA PROBABILITY DENSITY (DPD) IN FAULT MONITORING AND DIAGNOSIS

The Data Probability Density (DPD) gives the probability of the signal taking different amplitudes and can reflect the waveform characteristics of the signal. The shape of the density distribution curve is often different for different types of signals. As early as 1978, Dyer and Stewart [31] detected the damage of rolling bearings through the probability density function (PDF) analysis of statistical vibration signals. The PDF curve of the signal of the faulty bearing follows a non-Gaussian probability distribution, and as the bearing is damaged to varying degrees, the tail of the distribution curve significant changes will occur. In order to accurately describe the change of tail characteristics, the local statistic kurtosis is used to measure the non-Gaussian character of PDF curve, and a kurtosis-based bearing fault detection method is proposed. Yu [32] et al. proposed a new statistical modeling and detection method based on alpha stable distribution for the non-Gaussian statistical characteristics of bearing fault signals. alpha stable distribution is suitable for modeling random variables with heavy-tailed probability density functions. Compared with The kurtosis of the local statistic is a complete parametric probability distribution model. Xu [33] et al. introduced a nonparametric kernel density estimation (KDE) to describe the bubble size PDF distribution, and designed a detection filter based on the output PDF to achieve reagent addition in industrial froth flotation by threshold criteria determined by the system stability conditions required fault monitoring. Li [34] et al. calculated the probability density of the vibration signal of the connected parts, and meshed the probability density curve to generate a probability matrix, and then combined the principal component analysis method to realize the identification of bolt loosening. Su [35] et al. collected the ultrasonic signals and vibration signals generated by the bearing vibration measuring instrument, used the simplest histogram method to estimate the probability density of the signals, and studied and compared the distribution of ultrasonic amplitude histogram and vibration amplitude histogram in different fault states. Therefore, the use of this important feature of DPD can be used as an important basis for fault monitoring and diagnosis.

Based on the histogram algorithm, the probability density of the data can be quickly obtained. The histogram is the simplest non-parametric probability density estimation method. It consists of a series of vertical bars or lines of different heights, and the complexity of the algorithm is much lower than that of parameter estimation, KDE [36], etc. The basic idea of the histogram algorithm is to discretize continuous floating-point eigenvalues into K integers and construct a histogram with a width of K [37]. As the data is traversed, statistics are accumulated in a histogram based on the discretized value as an index. After traversing the data once, the statistical information accumulated by the histogram can be normalized to represent the probability density of the statistical data. If the DPD feature extraction process based on histogram algorithm is applied to fault diagnosis, the complexity of the algorithm will be reduced compared with some traditional frequency domain or time-frequency feature extraction methods. In addition, the histogram algorithm is also a method of compressing data, which can compress the original redundant data samples into shorter-length DPD feature vectors, making subsequent data processing faster and more convenient. However, the disadvantage of the histogram algorithm is that it needs to ensure that the sample size of the statistical data is sufficient, but this is easy to meet in the actual signal acquisition process. Setting a higher sampling frequency or increasing the sampling time of the samples can make the amount of acquired data samples sufficient.

B. GRAM ANGULAR FIELD (GAF) CODING METHOD

Gram Angular Field (GAF) converts 1D time series data into 2D images [26]. First, scale the 1D time series $X = \{x_1, x_2, \ldots, x_n\}$ to the interval $[-1,1]$. The calculation formula is as Equation (1):

$$\tilde{x}_{-1}^{i} = \frac{(x_i - \max(X) + (x_i - \min(X))}{max(X) - min(X)} \tag{1}$$

Then, use Equation (2) to re-convert to the time series of polar coordinates, where t_i is the time stamp, and N is the constant factor of the normalized polar coordinate system generation space.

$$\begin{cases} \emptyset = arc\ cos\left(\tilde{x}_i\right), & -1 \le \tilde{x}_i \le 1, \tilde{x}_i \in \tilde{X} \\ r = \dfrac{t_i}{N}, & t_i \in \mathbb{N} \end{cases} \tag{2}$$

After the original time series is transformed as described above, GAF generates images by constructing two different matrices. Equation (3) defines the Gram Angle Sum Field (GASF), and Equation (4) defines the Gram Angle Difference Field (GADF). The difference between the two matrices is the transformation of trigonometric functions, where GASF is based on the cosine function and GADF is based on the sine function.

$$GASF = \begin{bmatrix} cos\left(\theta_1 + \theta_1\right) & \cdots & cos\left(\theta_1 + \theta_n\right) \\ \vdots & \ddots & \vdots \\ cos\left(\theta_n + \theta_1\right) & \cdots & cos\left(\theta_n + \theta_n\right) \end{bmatrix} \tag{3}$$

$$GADF = \begin{bmatrix} sin\left(\theta_1 - \theta_1\right) & \cdots & sin\left(\theta_1 - \theta_n\right) \\ \vdots & \ddots & \vdots \\ sin\left(\theta_n - \theta_1\right) & \cdots & sin\left(\theta_n - \theta_n\right) \end{bmatrix} \tag{4}$$

Therefore, the core idea of the GAF coding method is to construct a special Gram matrix composed of polar angle combinations from a polar coordinate perspective. However, it is not a Gram matrix in the strict sense. It uses a combination of sine and cosine functions to replace the inner product elements in the traditional Gram matrix.

III. DPD-GAF-CNN FAULT DIAGNOSIS METHOD

In this section, the DPD-GAF-CNN fault diagnosis method is introduced: Firstly, the DPD-GAF coding method is introduced and some cases of analog conversion are listed; Then introduce the parameter structure of the CNN model based on LeNet-5 transformation; Finally, the detailed fault diagnosis process of DPD-GAF-CNN is given.

A. DPD-GAF CODING METHOD

The Data Probability Density-Gram Angle Field (DPD-GAF) transformation method converts 1D time series data into images. The conversion process is shown in Fig. 1.

Step 1: Convert the original floating point data to form Bins data. Then traverse the Bins data, count the number of identical Bins units, and construct an amplitude frequency histogram.

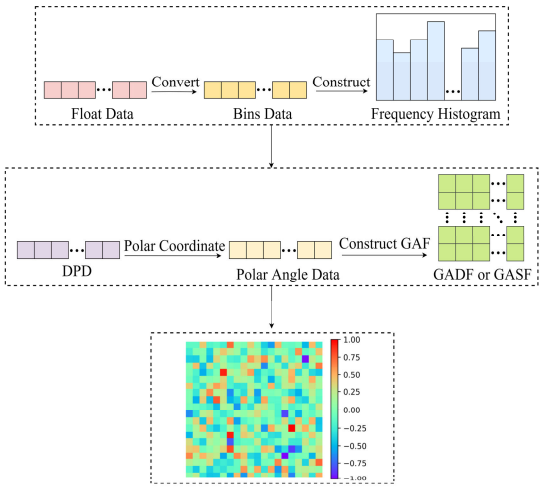


FIGURE 1. DPD-GAF encoding process.

Step 2: The DPD can be obtained by normalizing the amplitude frequency histogram obtained in the first step. According to the GAF coding method, the DPD is scaled, polar angle data is converted, and a GADF or GASF matrix is constructed.

Step 3: The GAF matrix of the second step is assigned red, green and blue colors according to the size of the element value, and is converted into a GAF feature map with multiple colors.

In order to visually demonstrate the effect of GAF in encoding DPD, we simulated the GADF and GASF feature maps encoded by the change of Gaussian distribution parameters based on Python language. Although it can be known that the DPD of the vibration signals collected for faulty bearings and gears does not follow the Gaussian probability distribution. Usually, the acceleration random signal collected by the rotating mechanical equipment in the normal state obeys the Gaussian distribution, accompanied by the generation of faults, the random signal is doped with the pulse signal generated by the fault impact, therefore, the statistical characteristics of the fault signal have non-Gaussian distribution characteristics (that is, the original Gaussian distribution has waveform distortion). However, the probability distribution of the fault signal is still evolved from the Gaussian distribution, as the most important distribution of probability statistics, the Gaussian distribution is the most important distribution in probability and statistics. Understanding the changes of the GAF feature map caused by the change of the Gaussian distribution parameters has important guiding significance for the subsequent understanding of the DPD-GAF feature map of the fault signal. As shown in Fig. 2, there are a total of 6 simulated transformation cases.

First, define a Gaussian probability distribution with different mean μ and standard deviation σ in the integer interval of 0-6, and then use 256 sampling points to averagely sample the generated curve in the interval, and finally generate it according to the encoding process of DPD in Fig. 1 GADF

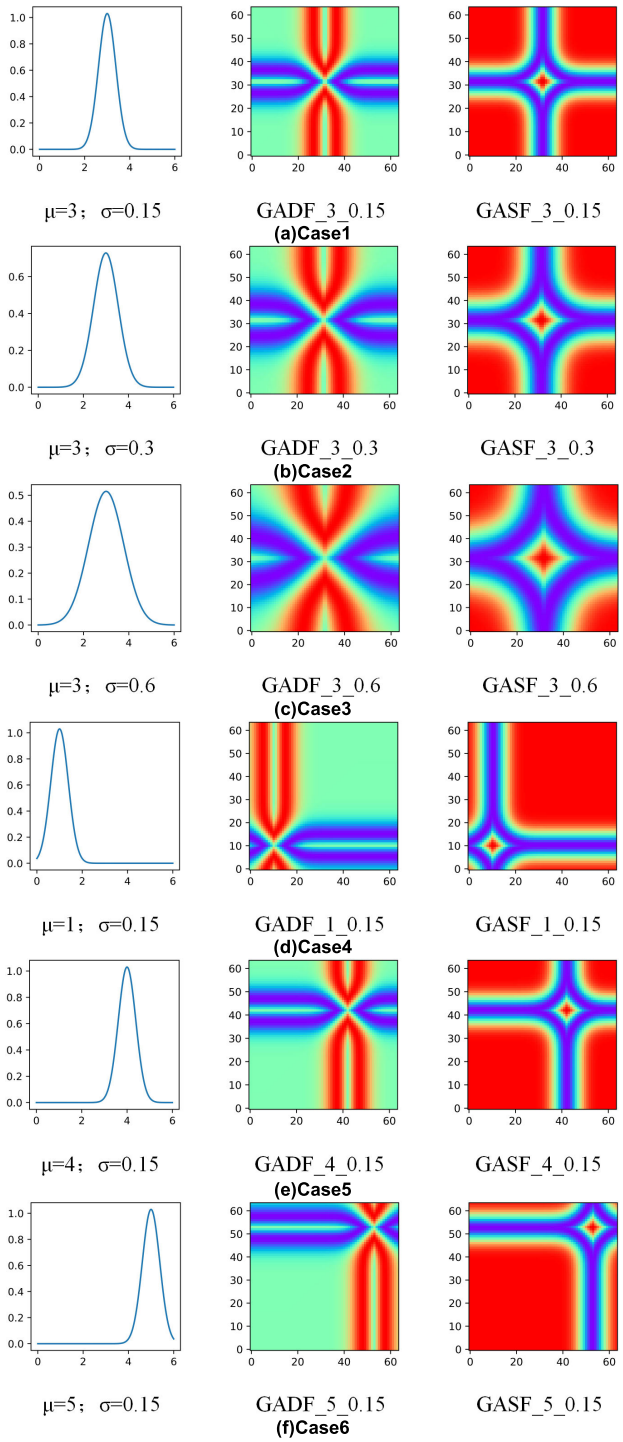


FIGURE 2. Simulation conversion case.

and GASF feature maps of size 64×64. Cases 1, 2, and 3 show GAF feature maps with mean $\mu = 3$ and standard deviation σ of 0.15, 0.3, and 0.6, respectively. Cases 4, 5, and 6 show GAF feature maps with standard deviation $\sigma = 0.15$ and mean μ of 1, 4, and 5, respectively. Obviously, with the change of the standard deviation and mean of the probability distribution, the feature contours similar to ''cross'' in the GADF and GASF feature maps will produce discrete and offset changes.

From the above analysis, it can be seen that GAF coding method is a good DPD visualization method. Compared with the grayscale method, the continuity of the data is not destroyed, and the characteristic changes of 1D data are preserved in the diagonal area of the image from the lower left corner to the upper right corner, and the non-diagonal area reflects the correlation between adjacent elements based on the sine and cosine functions. Compared with the traditional method based on signal processing, GAF has low time complexity and does not need to extract complex features by artificial design.

Therefore, the coding method based on DPD-GAF can intuitively and efficiently represent the probability distribution of vibration signal data. Combined with effective image recognition methods, the classification of different fault signals can be achieved. In the following part, we will use DPD-GADF, a specific coding method, to convert the vibration signals of gears and bearings to images, and combine CNN to realize fault diagnosis of gears and bearings. In order to train the model efficiently, we convert all kinds of DPD sequences generated based on the histogram algorithm into DPD-GADF feature maps of 64 × 64 size.

B. PROPOSED CNN STRUCTURE

CNN has powerful feature extraction capabilities, which is widely used in the field of image recognition in virtue of many model architectures. In recent years, many scholars have applied various models of CNN to the field of fault diagnosis [17]. Among them, the more popular 2DCNN model architectures mainly include LeNet-5, VGGNet, AlexNet, ResNet and GoogleNet. Literature [38], [39], [40], [41] used different 2DCNN models (including AlexNet, VGG-19, ResNet-50 and GoogleNet) to realize the fault diagnosis of bearings. The experimental results show that they have good performance in the field of bearing fault diagnosis. However, these network models have complex model structures and more hyperparameters compared with LeNet-5, which will lead to a longer duration of model training and testing and is not conducive to practical engineering applications. Considering the real-time detection requirements of fast fault diagnosis and early faults in practical engineering, the low-complexity LeNet-5 shallow CNN is selected for improvement. It mainly consists of convolutional layers, pooling layers and fully connected layers. Among them, the convolution layer is a feature extraction layer, and multiple convolution kernels can be used to extract different features of the image. After the convolution operation is completed, it is usually necessary to perform a nonlinear transformation on the features using a rectified linear unit (ReLU) activation function. The pooling layer performs downsampling on the feature map extracted by convolution, which can effectively reduce the computational complexity of the network and suppress overfitting. The combination of several convolutional

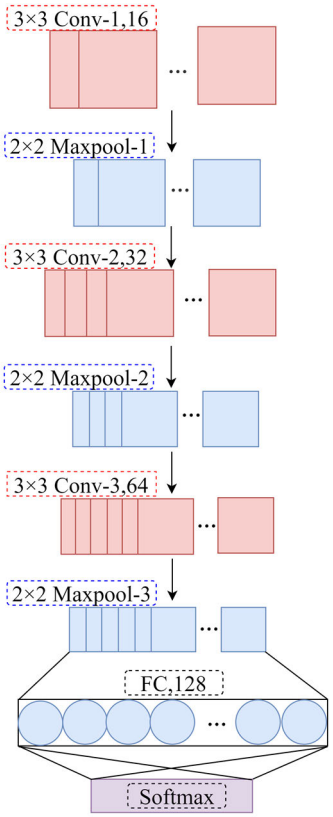


FIGURE 3. CNN model structure.

layers and pooling layers can effectively increase the depth of the model to extract high-dimensional features [42], [43]. After multiple convolution pooling operations, fully connected layers are used for classification.

As shown in Fig. 3, it is the constructed CNN model structure. The CNN model used in this study is based on the most classic LeNet-5 transformation. The size of the image identified in LeNet-5 is 32 × 32, which contains 2 convolutional layers and pooling layers. To accommodate the fault diagnosis case of this study, we use 3 convolutional layers and 3 pooling layers to extract 64 × 64 DPD-GADF image features. However, unlike LeNet-5, we adopt the zero-padding [44] method to keep the feature map size unchanged after the convolution operation, the convolution kernels are unified to 3×3 size, and the sliding step size is 1. The number of convolution kernels of the three convolutional layers is respectively are 16, 32 and 64. A 2 × 2 max pooling method with a sliding stride of 1 is used. In order to reduce the model parameters, one of the two fully connected layers of the LeNet-5 structure is removed, only one fully connected layer is used to implement feature mapping, and the dropout probability is set to 0.4. Finally, the image samples are classified by the softmax classification layer.

In addition, CNN models usually require a large amount of data to achieve better training results, so the sample set data needs to be enhanced. For the 1D data of fault diagnosis,

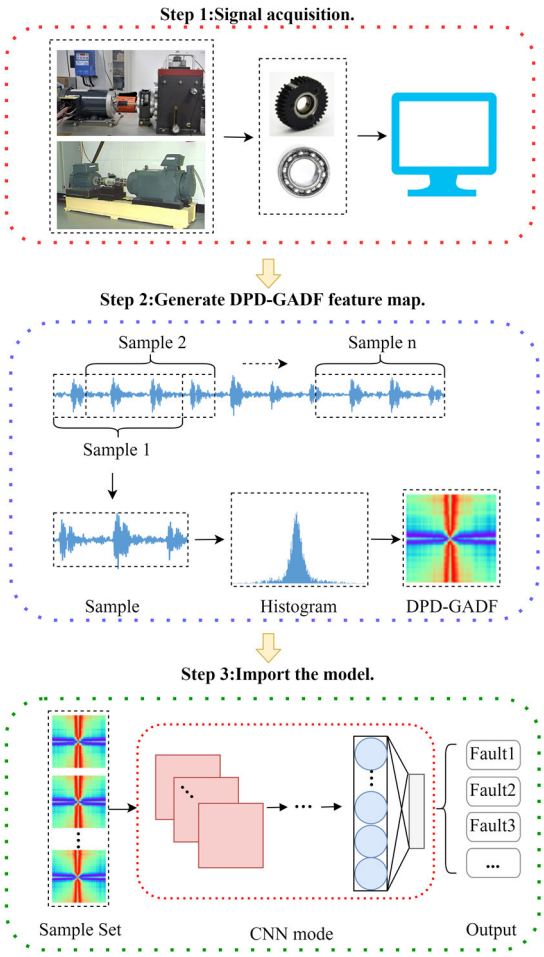


FIGURE 4. Process of fault diagnosis.

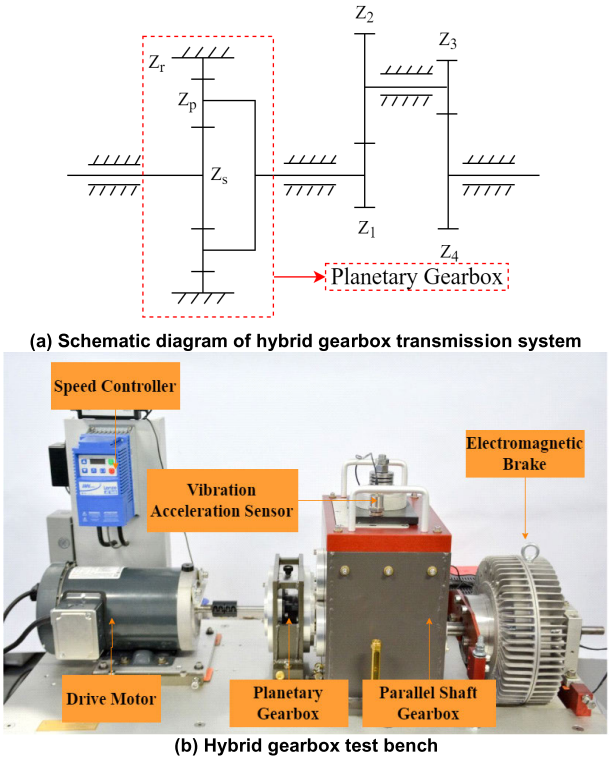
the method of overlapping sampling [45] can usually be used to realize the expansion of the data. The specific operation method is to set the sampling window size of the sample, and each time a sample is collected, the next sampling sample is obtained by sliding a fixed step size from the starting point of the previous sample.

C. DPD-GADF-CNN FAULT DIAGNOSIS PROCESS

The main process of fault diagnosis based on DPD-GADF-CNN is shown in Fig. 4, which is mainly divided into three steps:

Step1: Signal acquisition. Preset parts with different faults on the gear test bench or bearing test bench, and use the vibration acceleration sensor to set the appropriate sampling frequency to collect the gear and bearing signals of different fault modes;

Step2: Generate DPD-GADF feature map. Select the collected vibration signals to expand the data volume by overlapping sampling, and then convert the signal samples into DPD-GADF feature maps based on the process shown in Fig. 1;



(a) Schematic diagram of hybrid gearbox transmission system

(b) Hybrid gearbox test bench

FIGURE 5. Fault diagnosis comprehensive test bench.

Step3: Import the model. The DPD-GADF sample set is divided into training set and test set according to a certain proportion, input to the constructed CNN model and adjust the parameters, so that it can adaptively analyze image features and output fault diagnosis results.

IV. CASE STUDY AND EXPERIMENTAL RESULTS

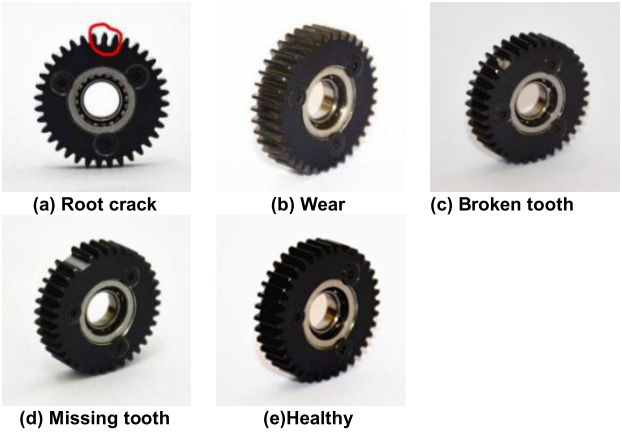
In this section, three different fault datasets are selected to verify and compare the proposed fault diagnosis method based on DPD-GADF-CNN.

The development of the fault diagnosis model is based on the built-in IDLE environment of Python 3.6, third-party libraries: Keras 2.3.1, Tensorflow 2.0.0, which are implemented by CPU for calculation.

A. CASE1: CONSTANT WORKING CONDITION PLANETARY GEARBOX DATA

Case 1 selects the planetary gearbox fault data generated by the DDS power transmission fault diagnosis comprehensive test bench manufactured by SpectmQuest Company for model testing, as shown in Fig. 5(a) for its transmission system diagram, relevant parameters: $Z_s = 28$, $Z_p = 36$, $Z_r = 100, Z_1 = 29, Z_2 = 100, Z_3 = 36, Z_4 = 90$. Fig. 5(b) is a physical diagram of the test bench, which is mainly composed of a drive motor, a speed controller, a planetary gearbox, a parallel shaft gearbox, and an electromagnetic brake.

The planetary gearbox of the DDS test bench has 4 planetary gears, the sun gear is the structural input member, he planet carrier is the structural output member, and is a first-level reduction device. The reduction ratio of the entire planetary gear train is 4.571. Five different fault types are preset for planetary gears: root crack, tooth surface wear, broken tooth, missing tooth and normal, so the fault diagnosis of planetary gearbox can be regarded as a 5-level classification task. Fig. 6 shows the physical map of the planetary gear with 5 fault states.



(a) Root crack (b) Wear (c) Broken tooth

(d) Missing tooth (e)Healthy

FIGURE 6. Physical map of faulty gear.

The vibration signal of the planetary wheel is collected by the vibration acceleration sensor, the sampling frequency is 5000Hz, and two constant operating conditions are set, the input shaft speed is 30Hz-system load 0N m (30-0) and the input shaft speed 50Hz-system load 0N m (50-0), other experimental conditions were the same. The vibration acceleration signals of the five faults under two operating conditions were collected respectively, and 150,000 points were collected in each operating condition, that is, the sampling time was 30 seconds.

1) DPD-GADF IMAGE CONVERSION

The length of the sample sampling window used for training and testing the diagnostic model is set to 5120, that is, the sampling time of each sample is about 1 second. The data volume is augmented by overlapping sampling with a sliding window step size of 128, and 1100 samples are generated for each fault type. According to the conversion process in Fig. 1, the samples are generated based on the histogram algorithm to generate a DPD with a length of 256, and then converted into a 64×64 DPD-GADF sample set.

As shown in Fig. 7, the DPD-GADF feature map generated for the 30-0 constant operating condition fault signal sample. Combined with the six cases in which the simulated Gaussian probability distribution is transformed into the GAF feature map in Fig. 2, the main difference between the DPD-GADF maps of different gear faults is reflected in the discrete degree of the ''cross'' feature contour, that is, the difference in the standard deviation of the data probability distribution. For obvious faults such as broken teeth and missing teeth, the characteristic contour is relatively convergent. For healthy

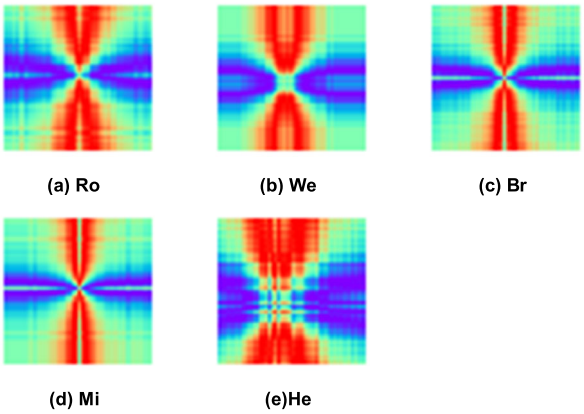


FIGURE 7. 30-0 gear DPD-GADF characteristic diagram.

gears the performance is relatively discrete. For weak faults such as cracks and wear, the discrete degree of characteristic contours is between that of healthy gears and obviously faulty gears.

2) MODEL TRAINING AND TESTING RESULTS

The ratio of training and testing samples of the DPD-GADF sample set of Case1 is 7:3. The corresponding labels and quantities are shown in Table 1.

TABLE 1. Sample structure of gear.

Working condition	Number of train set samples	Number of test set samples	Label
Root crack	770	330	Ro
Wear	770	330	We
Broken tooth	770	330	Br
Missing tooth	770	330	Mi
Healthy	770	330	He

The CNN model is the same as Fig. 3, the training period is set to 100 epochs, the optimizer selection is Adam, and the batch_size is set to 50. Input the DPD-GADF training set samples into the constructed CNN model. During the training process, the built-in ModelCheckpoint [46] callback function of the Keras library is used to realize the function of "continuous training at breakpoints". During the training process, a model weight is saved for each epoch of training, and save the corresponding optimal model weights according to the performance of the model on the validation set when iterating to 100 epochs.

After the model training is completed, load the saved optimal model weights to perform fault prediction on 1650 DPD-GADF test samples. The fault diagnosis accuracy of the test set samples for the 30-0 and 50-0 working conditions is 99.93% and 99.98%, respectively.

As shown in Fig. 8(a) and (b), the DPD-GADF test samples of two constant working conditions are input to the CNN diagnostic model classification confusion matrix. The prediction accuracy of the model for single-type faults can be obtained from the classification confusion matrix. According to the data in Fig. 8, the proposed diagnostic model achieves a high accuracy of more than 99% for any single type of gear fault diagnosis.

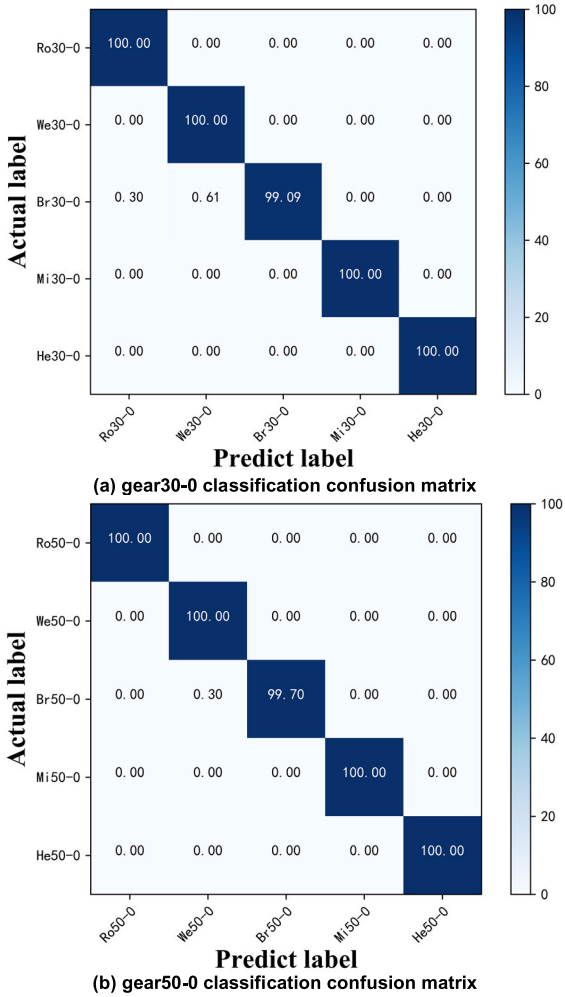


FIGURE 8. Gear classification confusion matrix.

3) COMPARED WITH OTHER METHODS

In order to verify the advantages of the combination of DPD-GADF and CNN, the classic machine learning classifier algorithms such as support vector machine (SVM), K nearest neighbor (KNN) and multi-layer perceptron (MLP) were further used in the 30-0 working condition planet gearbox dataset for experimental comparison. In addition, the corresponding GADF-SVM, GADF-KNN, GADF-MLP and GADF-CNN ablation experiments were set up. The experimental results are shown in Fig. 9. In order to avoid accidental errors, the accuracy rate is the average of 5 experiments.

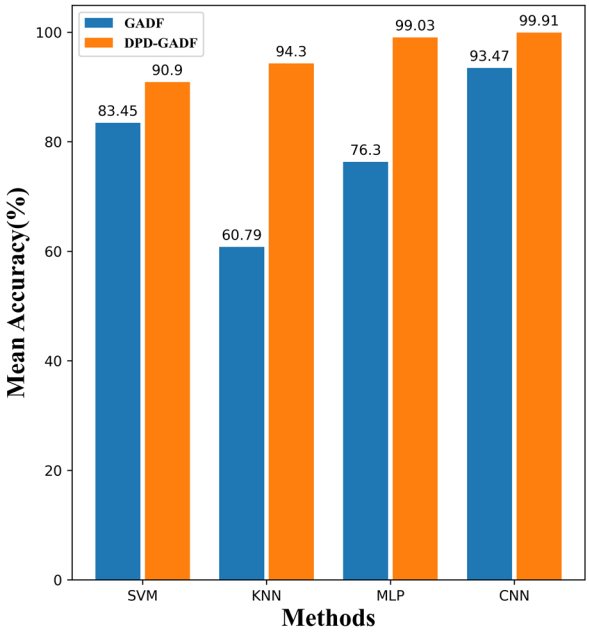


FIGURE 9. Comparison chart of different fault diagnosis methods.

Analyzing the experimental results in Fig. 9, it can be seen that whether DPD-GADF or GADF is used to encode the 30-0 working condition planetary gearbox data, the classification model based on CNN has higher classification accuracy than the classic machine learning classifier algorithm. The main reason is that the CNN model has certain automatic feature extraction capabilities compared with traditional machine learning algorithms. It is not difficult to find that the DPD-GADF encoding method has a certain contribution to the improvement of the classification accuracy of the model by setting the GADF encoding ablation experiment. Even a simple MLP classification algorithm can achieve a classification effect of more than 99%, which benefits from the DPD extraction process. It effectively highlights the characteristics of different fault states, making the subsequent pattern recognition process more accurate. On the contrary, if the original GADF encoding method is adopted, due to the redundancy of the original data samples, the dimension of the encoded GADF matrix is too large, which is not conducive to the expression of fault characteristics. According to the experimental results in Fig. 9, it can be found that the classification accuracy of GADF-SVM, GADF-KNN and GADF-MLP models is not good, which is 83.45%, 60.79% and 76.3% respectively. Even if the CNN model with feature extraction ability is used, the classification accuracy is only 93.47%.

B. CASE2: MIXED WORKING CONDITION PLANETARY GEARBOX DATA

In case 2, the data generated by the DDS-type test bench is still used for model testing. The difference is that the sampling frequency of the vibration acceleration sensor is 10000Hz, and 2 kinds of rotational speed (30Hz, 40Hz) and 2 kinds of torque (0N m, 40N m) are set to form 30-0, 30-40, 40-0 and 40-40 mixed operating conditions. The vibration acceleration signals of 5 faults under 4 operating conditions were collected respectively, and 120,000 points were collected for each condition, that is, 480,000 sampling points under mixed conditions were collected for each fault, and the sampling time was 48 seconds.

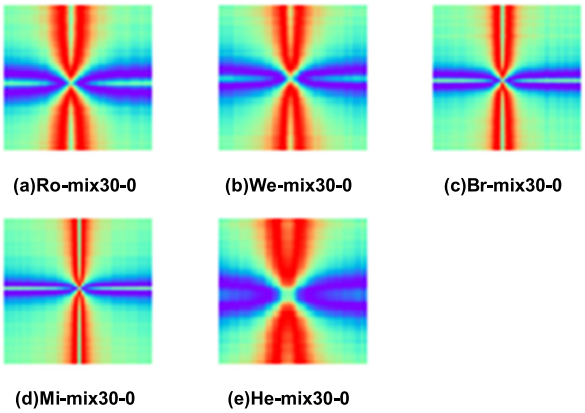


FIGURE 10. Mix30-0 gear DPD-GADF characteristic diagram.

1) DPD-GADF IMAGE CONVERSION

In order to verify the encoding effect of DPD-GADF on long sequence samples, the sample sampling window length for training and testing the diagnostic model in this case is set to 10240, but the sampling time of each sample is still about 1 second. 1400 samples were generated by overlapping sampling of 4 different operating conditions for each fault. The corresponding DPD sequence length is increased to 512, and then converted into a 64×64 DPD-GADF sample set.

Fig. 10 shows the DPD-GADF characteristic map generated from the 30-0 gear signal in the mixed condition. The changing law of the characteristic contour is similar to that in Fig. 7, but the original 1D signals of the two are different. The sampling frequency of the gear 30-0 signal is 5000, and the length of a single sample is 5120. The sampling frequency of the gear mix30-0 signal is 10000, and the length of a single sample is 10240. Therefore, even the data of long sequence samples can resist the redundancy of information through DPD-GADF and generate small-sized feature maps to facilitate the training of CNN models. Even, since the data of the long sequence samples are more sufficient, the generated DPD-GADF feature outline is clearer.

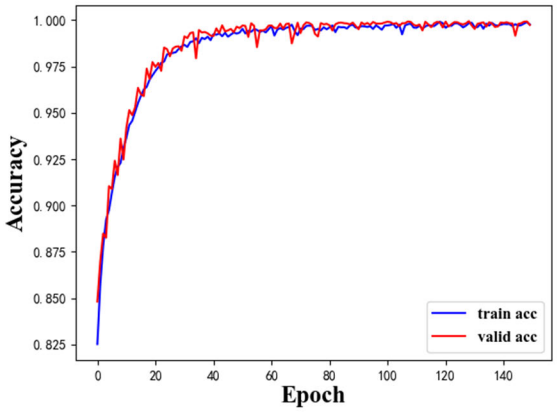
2) MODEL TRAINING AND TESTING RESULTS

The proportion of Case2 samples is the same as that of Case1, and the corresponding labels and quantities are shown in Table 2.

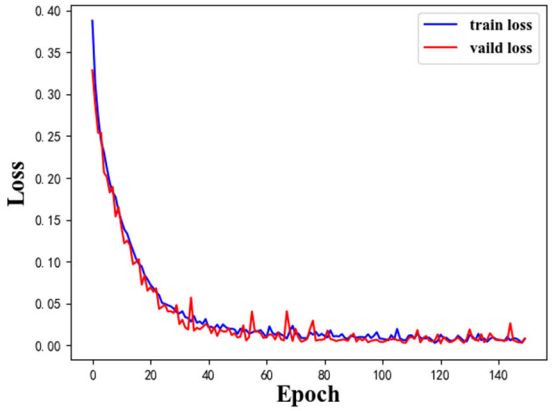
In view of the large number of data samples in Case2, the sliding step size of the first layer convolution of the CNN

TABLE 2. Sample structure of gear-mix.

Working condition	Number of train set samples	Number of test set samples	Label
Root crack	3920	1680	Ro-mix
Wear	3920	1680	We-mix
Broken tooth	3920	1680	Br-mix
Missing tooth	3920	1680	Mi-mix
Healthy	3920	1680	He-mix



(a)The accuracy curve of gear-mix



(b) The loss function of gear-mix

FIGURE 11. The curve of training accuracy and loss function of the gear-mix diagnostic model.

model of Case1 is changed to 2, and the batch_size is set to 128 to speed up the model training. The training period is set to 150 epochs, and other settings are consistent with the Case1 model. When the iteration reaches 150 epochs, the corresponding optimal model weights are saved according to the performance of the model on the validation set.

Fig. 11(a), (b) shows the change curve of the accuracy rate and the loss function during CNN model training. After iterating to 80 epochs, the change of the curve tends to be stable and the fitting state is good. After finishing training at 150 epochs, load the best model weights to perform fault prediction on 8400 DPD-GADF test samples. The accuracy on the test set is 99.92%.

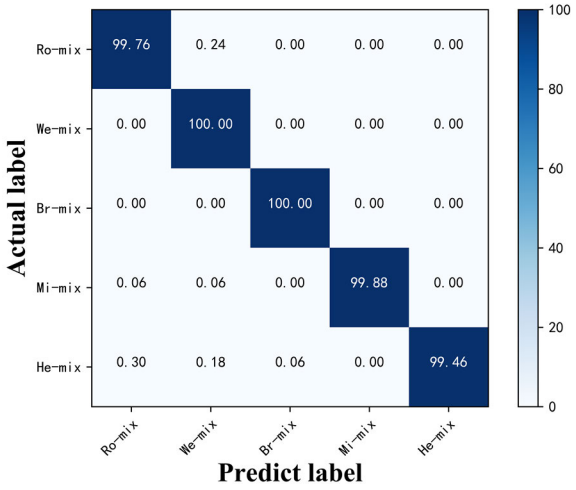


FIGURE 12. Gear-mix classification confusion matrix.

Fig. 12 shows the classification confusion matrix of the faulty gear under mixed working conditions. The same as the case 1's constant working condition gear fault diagnosis, the prediction accuracy for five types of faulty gears is also maintained above 99%. The effectiveness of the proposed method for gear fault diagnosis is proved.

3) COMPARED WITH OTHER METHODS

Further, the diagnostic model based on DPD-GADF is compared with other existing methods. As shown in Table 3, the literature [27] adopted the original GADF encoding method to encode the faulty gear data of DDS mixed conditions into a feature map of size 300×300, and then input it into a CNN model containing three layers of convolution and pooling. The diagnostic accuracy on the test set was 98.8%. Therefore, compared with the method based on GADF, DPD-GADF reduces the size of the gear feature map to speed up the training of the CNN model, and on the other hand improves the accuracy of fault diagnosis to 99.92%. The literature [47] generates a bi-spectral feature map (BSP) for faulty gears under different working conditions based on the method of bi-spectral analysis, and then uses the CNN model to achieve fault classification, and the fault diagnosis accuracy rate is 97.2%. Compared with such methods based on signal analysis, the method based on DPD-GADF has obvious advantages, and has a certain improvement in the accuracy of each type of gear fault diagnosis. And no complicated calculation is involved in the image conversion process.

C. CASE3: BEARING DATA OF CWRU

Case 3 is verified by using the data generated by the bearing test bench of Case Western Reserve University. Specifically, we selected 12KHz sampling frequency, 2HP motor load and drive end bearing data for verification. The bearing model is SKF 6205-2RS deep groove ball bearing. Single point faults are arranged on the inner ring, outer ring and rolling element of the rolling bearing. The fault dimensions are 0.007inch,

TABLE 3. Fault diagnosis results of gears in mixed working conditions.

Method	Ro-mix (%)	We-mix (%)	Br-mix (%)	Mi-mix (%)	He-mix (%)	Mean (%)
Proposed	**99.76**	**100**	**100**	**99.88**	**99.46**	**99.92**
[27]GADF-CNN			-			98.8
[47]BSP-CNN	96.5	97.2	98.6	95.1	99.3	97.4

TABLE 4. Sample structure of bear.

Working condition	Number of train set samples	Number of test set samples	Label
7 mils ball faults	630	270	B007
14 mils ball faults	630	270	B014
21 mils ball faults	630	270	B021
7 mils Inner race faults	630	270	IR007
14 mils Inner race faults	630	270	IR014
21 mils Inner race faults	630	270	IR021
7 mils outer race faults	630	270	OR007
14 mils outer race faults	630	270	OR014
21 mils outer race faults	630	270	OR021
Healthy	630	270	No

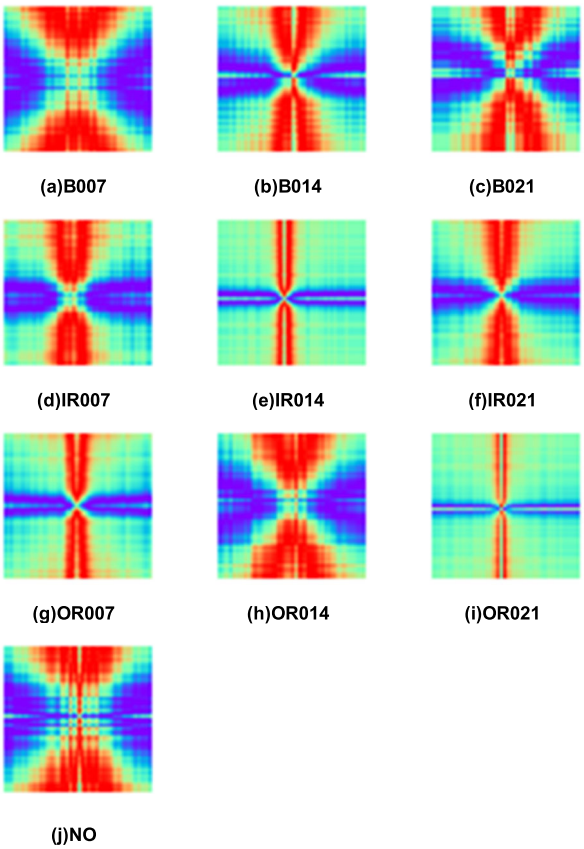


FIGURE 13. Bearing DPD-GADF feature map.

0.014inch and 0.021inch, plus a normal fault-free type, a total of 10 types of bearings.

1) DPD-GADF IMAGE CONVERSION

The length of the sample sampling window for training and testing the diagnostic model is set to 4096, and the data volume was expanded by overlapping sampling. 900 samples are generated for each fault type, and the length of the DPD is 256, and then converted into a 64 × 64 bearing DPD-GADF sample set, as shown in Fig. 13.

Observing the feature map generated by the CWRU bearing data through the DPD-GADF coding method in Fig 13, we can find a very interesting phenomenon: for the faulty bearings of 007 and 021, from the rolling element fault to the inner ring fault to the outer ring fault (B007-IR007-OR007 and B021-IR021-OR021), the ''cross''-type feature profile of the feature map is gradually convergent from dispersion. But for the faulty bearing of 014 (B014-IR014-OR014), this rule is not followed. Even, OR014 seems to be an abnormal label sample from the surface features. From the rules summarized by 007 and 021, the actual label of OR014 should be classified as a rolling element fault (B007, B014 or B021), which is in line with intuitive experience.

Aiming at this special phenomenon, the literature on analyzing the signal characteristics of CWRU bearings was consulted. Smith and Randall [48] applied multiple fault diagnosis methods based on signal processing to the entire CWRU dataset. It was found that the fault data set of 014 was the most difficult to diagnose, because it was found that the pulse shaft speed modulation phenomenon of the bearing fault frequency (BPFI) appeared in the signal. Consistent with the conclusion in the literature [38], even if the fault 1D data of 014 is encoded by DPD-GADF image, it is difficult to make an accurate judgment on the faulty bearing type of 014 with the intuitive experience summarized. Obviously, the identification of OR014 is easily confused with B007, B014 or B021, and the identification of IR014 is easily confused with OR021. Therefore, accurately distinguishing ten bearing states still requires the help of high-dimensional and abstracts knowledge extracted by CNN.

2) MODEL TRAINING AND TESTING RESULTS

The bearing signal DPD-GADF sample set of Case3 is also used to train the CNN model according to 70%, and 30% is used to verify and test the fault diagnosis model. The corresponding labels and quantities are shown in Table 4.

The CNN model and hyperparameter settings are the same as Case1, except that the CWRU fault data is a 10-classification task. Fig. 14(a), (b) shows the change curve of the accuracy rate and the loss function during the training of the CNN model. After iterating to 60 epochs, the curve

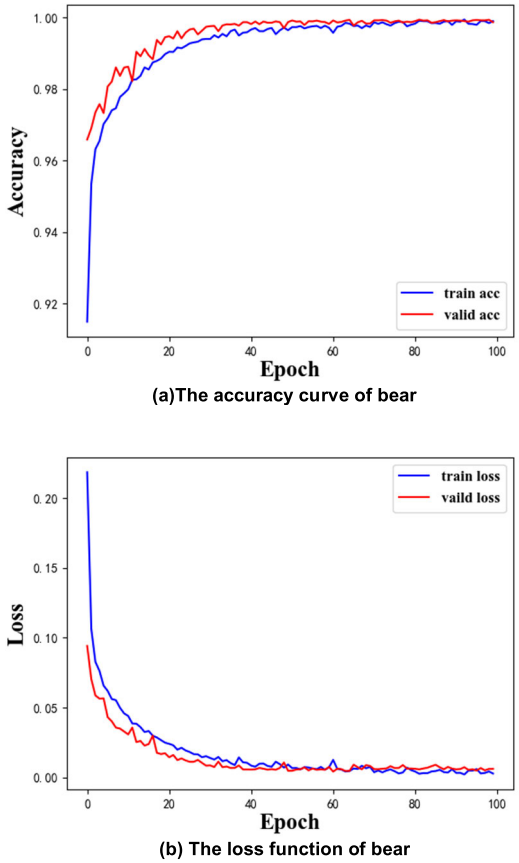


FIGURE 14. The curve of the training accuracy and loss function of the bear diagnostic model.

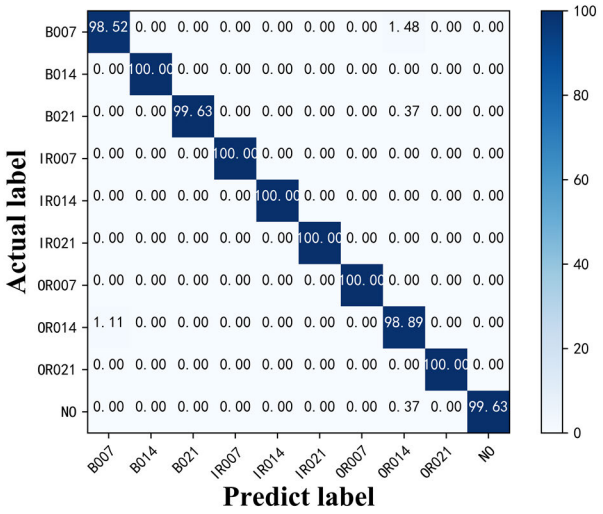


FIGURE 15. Bear classification confusion matrix.

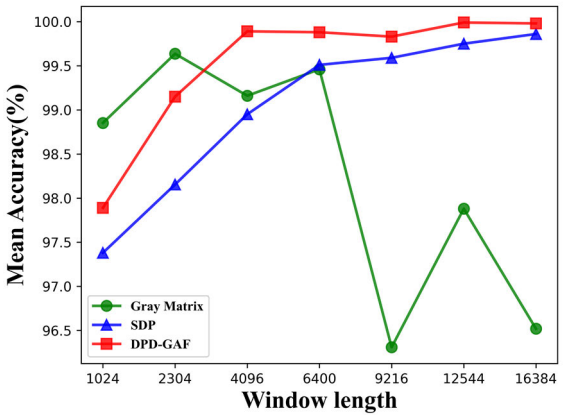


FIGURE 16. Comparison chart of redundant data processing by different encoding methods.

changes tend to be stable and the fitting state is good. After 100 epochs of training, load the best model weights to predict the failure of 2700 DPD-GADF test samples, and the prediction accuracy of the test set is 99.93%.

Fig. 15 shows the classification confusion matrix of the bearing data set, which is consistent with the results we analyzed according to Fig. 13: the source of fault misdiagnosis in the CWRU data set is mainly due to OR014, and there is a misdiagnosis rate of about 1.5% between it and B007. However, based on the features extracted by the CNN model, the diagnostic accuracy of OR014 and B007 faults can reach 98.89% and 98.52%, respectively, which has a good classification effect. The remaining fault types maintain a high diagnostic accuracy of more than 99.6%. Overall, the proposed method has good applicability for bearing fault diagnosis.

3) COMPARED WITH OTHER METHODS

The results of the DPD-GADF-CNN fault diagnosis method tested on the CWRU dataset are compared with other existing methods. At the same time, in order to verify the advantages of DPD-GADF-CNN in processing redundant data, a sample set with different window lengths was set to test the stability of the algorithm model. As a comparison, the gray matrix coding of literature [25] and the SDP coding method of literature [22] are used to process sample sets with different window lengths. Since the gray matrix coding method requires the sample length to be the square of an integer, the set sample length They are 1024, 2304, 4096, 6400, 9216, 12544 and 16384 in turn. The image sizes encoded by DPD-GADF and SDP are both 64×64, and the selected CNN model is consistent with the model of Case 3.

The experimental results are shown in Fig. 16. In order to avoid accidental errors, the accuracy rate is the average of 5 experiments. From the change curve of the average accuracy rate with the length of the sample window in Fig. 16, it can be seen that the combination of DPD-GADF encoding method and CNN can achieve high-precision diagnosis for redundant samples with a length greater than 4096, which is mainly due to the probability density Statistical features remain unchanged for different degrees of redundant samples, so that the subsequent CNN algorithm model can maintain stable performance. However, for the sample sets with

window lengths of 1024 and 2304, the probability density characteristics of the histogram algorithm may be lost to a certain extent due to the relatively under-redundant sample size, making the final diagnostic accuracy lower than other relatively redundant sample sets. As shown in Fig. 17, the accuracy change curve of the verification set for the first 50 Epochs of DPD-GADF-CNN model training intuitively shows the convergence performance of the model for sample sets with different window lengths. Similar conclusions can be drawn from the experimental results in Fig. 16: for redundant sample sets with a length of 4096 or more, the diagnostic model can eventually converge to high precision and remain stable, but for sample sets with a window length of 1024 and 2304, the loss of statistical features makes The model converges poorly.

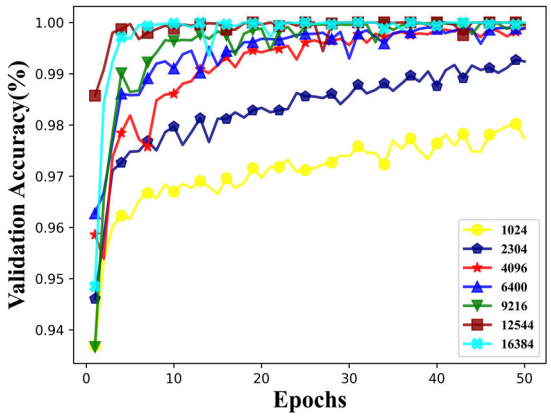


FIGURE 17. DPD-GADF-CNN model verification set accuracy change curve.

However, compared with the other two encoding methods, DPD-GADF still maintains a relatively good performance: although the gray matrix encoding method has high diagnostic accuracy for 1024 and 2304 window length sample sets, it is not suitable for other relatively redundant signals. It cannot maintain stable performance, mainly because as the redundancy of samples increases, the image encoded by the gray matrix has insufficient ability to represent the characteristics of the sample, which makes the final diagnosis effect of the CNN model unsatisfactory. The diagnostic effect of the SDP coding method gradually increases with the increase of sample redundancy, but its accuracy is slightly lower than that of the DPD-GADF coding method. Therefore, the two methods of gray matrix coding and SDP coding have certain deficiencies in the parameter selection of the sample window length. The parameter selection of DPD-GADF coding samples is more convenient and more adaptable, which enables the CNN model to maintain stable and high-precision diagnostic performance when processing redundant samples.

D. DISCUSSION

A total of three representative fault datasets are used in the above case studies to verify the superiority of the proposed DPD-GADF-CNN based fault diagnosis method. Converting 1D signals with lengths of 5120, 10240, and 4096 into DPD-GADF feature maps of 64×64 size can highlight the probability distribution characteristics of different fault signals on the one hand, and reduce the parameter adjustment work of the CNN model on the other hand. The prediction accuracy of the proposed fault diagnosis method on these three datasets is above 99.9%, and the results are better than other results verified on the same datasets, which proves that the proposed method has a wide range of adaptability.

V. CONCLUSION

This study proposes a novel fault diagnosis method based on DPD-GAF-CNN. Its main contribution is to represent the probability density distribution of the signal with a GAF image combined with the CNN based on LeNet-5 transformation to achieve high-precision classification of fault signals. The proposed diagnosis method is verified on the data of the planetary gearbox under constant working conditions, the data for a planetary gearbox under mixed working conditions and the CWRU bearing data. It has been proved that the prediction accuracy on these three data sets is above 99.9%. Compared with other deep learning and machine learning methods, the DPD-GAF-CNN method can reduce the work of CNN in input data and model hyperparameter adjustment, adapt to different degrees of redundant signals, and improve the efficiency and accuracy of fault diagnosis.

REFERENCES

[1] M. Yu, D. Wang, and M. Luo, "Model-based prognosis for hybrid systems with mode-dependent degradation behaviors," *IEEE Trans. Ind. Electron.*, vol. 61, no. 1, pp. 546–554, Jan. 2014.

[2] K. Wawryn and W. Zinka, "A prototype expert system for fault diagnosis in electronic devices," in *Proc. Eur. Conf. Circuit Theory Design*, 1989, pp. 677–680.

[3] W. Gong, H. Chen, Z. Zhang, M. Zhang, R. Wang, C. Guan, and Q. Wang, "A novel deep learning method for intelligent fault diagnosis of rotating machinery based on improved CNN-SVM and multichannel data fusion," *Sensors*, vol. 19, no. 7, p. 1693, Apr. 2019.

[4] Y. Long, W. Zhou, and Y. Luo, "A fault diagnosis method based on one-dimensional data enhancement and convolutional neural network," *Measurement*, vol. 180, Aug. 2021, Art. no. 109532.

[5] J. MacGregor and A. Cinar, "Monitoring, fault diagnosis, fault-tolerant control and optimization: Data driven methods," *Comput. Chem. Eng.*, vol. 47, pp. 111–120, Dec. 2012.

[6] W. Cui, G. Meng, A. Wang, X. Zhang, and J. Ding, "Application of rotating machinery fault diagnosis based on deep learning," *Shock Vibrat.*, vol. 2021, pp. 1–30, Dec. 2021, Art. no. 3083190.

[7] A. K. Ozcanli, F. Yaprakdal, and M. Baysal, "Deep learning methods and applications for electrical power systems: A comprehensive review," *Int. J. Energy Res.*, vol. 44, no. 9, pp. 7136–7157, Jul. 2020.

[8] H. Ding, R. X. Gao, A. J. Isaksson, R. G. Landers, T. Parisini, and Y. Yuan, "State of AI-based monitoring in smart manufacturing and introduction to focused section," *IEEE/ASME Trans. Mechatronics*, vol. 25, no. 5, pp. 2143–2154, Oct. 2020.

[9] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[10] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2014, pp. 1409–1556.
[13] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.
[14] E. Li, J. Xia, P. Du, C. Lin, and A. Samat, "Integrating multilayer features of convolutional neural networks for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5653–5665, Oct. 2017.
[15] J. Zhao, S. P. Yang, Q. Li, Y. Q. Liu, X. H. Gu, and W. P. Liu, "A new bearing fault diagnosis method based on signal-to-image mapping and convolutional neural network," *Measurement*, vol. 176, May 2021, Art. no. 109088.
[16] Q. Wang, B. Zhao, H. Ma, J. Chang, and G. Mao, "A method for rapidly evaluating reliability and predicting remaining useful life using two-dimensional convolutional neural network with signal conversion," *J. Mech. Sci. Technol.*, vol. 33, no. 6, pp. 2561–2571, Jun. 2019.
[17] J. Jiao, M. Zhao, J. Lin, and K. Liang, "A comprehensive review on convolutional neural network in machine fault diagnosis," *Neurocomputing*, vol. 417, pp. 36–63, Jan. 2020.
[18] S. Tang, S. Yuan, and Y. Zhu, "Convolutional neural network in intelligent fault diagnosis toward rotatory machinery," *IEEE Access*, vol. 8, pp. 86510–86519, 2020.
[19] L. H. Wang, X. P. Zhao, J. X. Wu, Y. Y. Xie, and Y. H. Zhang, "Motor fault diagnosis based on short-time Fourier transform and convolutional neural network," *Chin. J. Mech. Eng.*, vol. 30, pp. 1357–1368, Nov. 2017.
[20] M.-F. Guo, X.-D. Zeng, D.-Y. Chen, and N.-C. Yang, "Deep-learning-based earth fault detection using continuous wavelet transform and convolutional neural network in resonant grounding distribution systems," *IEEE Sensors J.*, vol. 18, no. 3, pp. 1291–1300, Feb. 2018.
[21] G. Li, C. Deng, J. Wu, Z. Chen, and X. Xu, "Rolling bearing fault diagnosis based on wavelet packet transform and convolutional neural network," *Appl. Sci.*, vol. 10, no. 3, p. 770, Jan. 2020.
[22] H. Wang, J. Xu, R. Yan, and R. X. Gao, "A new intelligent bearing fault diagnosis method using SDP representation and SE-CNN," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 52, pp. 2377–2389, May 2020.
[23] L. Ciabattoni, F. Ferracuti, A. Freddi, and A. Monteriù, "Statistical spectral analysis for fault diagnosis of rotating machines," *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4301–4310, May 2018.
[24] W. Gong, H. Chen, Z. Zhang, M. Zhang, and H. Gao, "A data-driven-based fault diagnosis approach for electrical power DC–DC inverter by using modified convolutional neural network with global average pooling and 2-D feature image," *IEEE Access*, vol. 8, pp. 73677–73697, 2020.
[25] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990–5998, Jul. 2018.
[26] Z. Wang and T. Oates, "Imaging time-series to improve classification and imputation," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, 2015, pp. 1–7.
[27] X.-Y. Pang, Y. Tong, and Z.-H. Wei, "A fault diagnosis method for planetary gearbox based on GAF-CNN," *Beijing Ligong Daxue Xuebao*, vol. 40, no. 11, pp. 1161–1167, 2020.
[28] C. Li, J. Xiong, X. Zhu, Q. Zhang, and S. Wang, "Fault diagnosis method based on encoding time series and convolutional neural network," *IEEE Access*, vol. 8, pp. 165232–165246, 2020.
[29] Y. Tong, X. Pang, and Z. Wei, "Fault diagnosis method of rolling bearing based on GADF-CNN," *J. Vib. Shock*, vol. 40, no. 5, pp. 247–253, 2021.
[30] N. Park and S. Kim, "FlexSketch: Estimation of probability density for stationary and non-stationary data streams," *Sensors*, vol. 21, no. 4, p. 1080, Feb. 2021.
[31] D. Dyer and R. M. Stewart, "Detection of rolling element bearing damage by statistical vibration analysis," *ASME J. Mech. Des.*, vol. 100, no. 2, pp. 229–235, 1978.
[32] G. Yu, C. Li, and J. Zhang, "A new statistical modeling and detection method for rolling element bearing faults based on alpha–stable distribution," *Mech. Syst. Signal Process.*, vol. 41, nos. 1–2, pp. 155–175, Dec. 2013.
[33] C. Xu, W. Gui, C. Yang, H. Zhu, Y. Lin, and C. Shi, "Flotation process fault detection using output PDF of bubble size distribution," *Minerals Eng.*, vol. 26, pp. 5–12, Jan. 2012.
[34] Y.-G. Li, X.-N. Kong, and Y.-Y. Gao, "Method for detecting bolt looseness based on probability density of vibration signals of two connected parts and principal component analysis," *J. Vib. Shock*, vol. 34, no. 1, pp. 63–67, 2015.
[35] L. C. Su, Y. E. Shi, and X. L. Li, "Fault diagnosis of bearing based on the ultrasonic amplitude histogram," *Adv. Mater. Res.*, vol. 479, pp. 1361–1364, Feb. 2012.
[36] Y.-J. Oyang, S.-C. Hwang, Y.-Y. Ou, C.-Y. Chen, and Z.-W. Chen, "Data classification with radial basis function networks based on a novel kernel density estimation algorithm," *IEEE Trans. Neural Netw.*, vol. 16, no. 1, pp. 225–236, Jan. 2005.
[37] S. Yang and H. Zhang, "Comparison of several data mining methods in credit card default prediction," *Intell. Inf. Manag.*, vol. 10, no. 5, pp. 112–115, 2018.
[38] X. Shi, "An improved bearing fault diagnosis scheme based on hierarchical fuzzy entropy and AlexNet network," *IEEE Access*, vol. 9, pp. 61710–61720, 2021.
[39] L. Wen, X. Li, X. Li, and L. Gao, "A new transfer learning based on VGG-19 network for fault diagnosis," in *Proc. IEEE 23rd Int. Conf. Comput. Supported Cooperat. Work Design (CSCWD)*, May 2019, pp. 205–209.
[40] L. Wen, X. Li, and L. Gao, "A transfer convolutional neural network for fault diagnosis based on ResNet-50," *Neural Comput. Appl.*, vol. 32, pp. 6111–6124, Feb. 2019.
[41] C. Grover and N. Turk, "A novel fault diagnostic system for rolling element bearings using deep transfer learning on bispectrum contour maps," *Eng. Sci. Technol., Int. J.*, vol. 31, Jul. 2022, Art. no. 101049.
[42] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Jul. 2016.
[43] S. Shao, S. McAleer, R. Yan, and P. Baldi, "Highly accurate machine fault diagnosis using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2446–2455, Apr. 2019.
[44] Y. Qian and P. C. Woodland, "Very deep convolutional neural networks for robust speech recognition," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 12, pp. 481–488, Dec. 2016.
[45] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, pp. 425–446, 2017.
[46] S. N. Zisad, M. S. Hossain, and K. Andersson, "Speech emotion recognition in neurological disorders using convolutional neural network," in *Proc. Int. Conf. Brain Informat.*, 2020, pp. 287–296.
[47] X. Pang, X. Xue, W. Jiang, and K. Lu, "An investigation into fault diagnosis of planetary gearboxes using a bispectrum convolutional neural network," *IEEE/ASME Trans. Mechatronics*, vol. 26, no. 4, pp. 2027–2037, Aug. 2021.
[48] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," *Mech. Syst. Signal Process.*, vol. 64, pp. 100–131, Dec. 2015.

BOWEN ZHANG was born in Yan'an, Shaanxi, China, in 1998. He received the bachelor's degree in mechatronic engineering from China University of Geosciences, Wuhan, China, in 2020. He is currently pursuing the degree with the College of Mechanical and Vehicle Engineering, Taiyuan University of Technology, Taiyuan, China.

His research interests include image encoding, time series analysis, and deep learning.

Mr. Zhang received the National Inspirational Scholarship, in 2017 and 2018.

XINYU PANG was born in Lüliang, Shanxi, China, in 1976. She received the Ph.D. degree in mechanical design and theory from 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. Since 2020, she has been a Professor with the College of Mechanical and Vehicle Engineering, Taiyuan University of Technology. She published an academic article titled "An Investigation into Fault Diagnosis of Planetary Gearboxes Using A Bispectrum Convolutional Neural Network" in the IEEE/ASME TRANSACTIONS ON MECHATRONICS, in 2021. Her main research interests include mechanical vibration and testing, mechanical fault diagnosis, oil monitoring, and life prediction.

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

PENG ZHAO was born in Harbin, Heilongjiang, China, in 1998. He received the bachelor's degree in mechanical design manufacture and automation major from Northeast Petroleum University, in 2020. He is currently pursuing the master's degree in mechanical engineering with Taiyuan University of Technology, Taiyuan, China.

His research interests include mechanical life prediction and machine vision.

Mr. Zhao won the third prize in the National College Students Mathematics Competition, in 2019.

KAIBO LU was born in Yuncheng, Shanxi, China, in 1984. 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 Taiyuan University of Technology, Taiyuan, China, where he is currently an Associate Professor with the College of Mechanical and Vehicle Engineering. He was an Academic Visitor with the Center for Efficiency and Performance Engineering, University of Huddersfield, in 2019, which has been sponsored by the China Scholarship Council. He has authored or coauthored 21 academic articles, out of which 11 are included in SCI and EI. In 2018, he has authored an article titled "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, fault diagnosis, and dynamics in machining processes.