

Research Article

Bearing Fault Identification Method under Small Samples and Multiple Working Conditions

Yuhui Wu ^{1,2}, Licai Liu ¹, Shuqu Qian,¹ and Jianyong Tian ¹

¹Anshun University, Anshun 561000, Guizhou, China

²St. Paul University Philippines, Tuguegarao 3500, Cagayan, Philippines

Correspondence should be addressed to Yuhui Wu; yuhuiwu2121@yeah.net

Received 6 June 2022; Revised 12 July 2022; Accepted 19 July 2022; Published 13 August 2022

Academic Editor: Praveen Kumar Reddy Maddikunta

Copyright © 2022 Yuhui Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aiming at the problem of the low bearing faults identification accuracy of the method based on the deep neural network under small samples and multiple working conditions, a novel bearing fault identification method combined with the coordinate delay phase space reconstruction method (CDPSR), residual network, meta-SGD algorithm, and the AdaBoost technology was proposed. The proposed method firstly calculates the high-dimensional space coordinates of bearing vibration signals using the CDPSR method and uses these coordinates to construct a training set, then learns and updates the parameters of classifier networks using the meta-SGD algorithm with the train set, iteratively trains multiple classifiers, and finally integrates those classifiers to form a strong classifier by AdaBoost technology. The 4-way and 20-shot experiments of artificial and natural bearing faults show that the proposed method can identify the fault samples and nonfault samples with 100% accuracy, and the fault location accuracy is over 90%. Compared with some state-of-the-art methods such as WDCNN and CNN-SVM, the proposed method improves the fault identification accuracy and stability to a certain extent. The proposed method has high fault identification accuracy under small samples and multiworking conditions, which makes it applicable in some practical areas of complex working conditions and difficulty obtaining bearing fault signals.

1. Introduction

From the perspective of the application, the bearing fault identification can be divided into two types, which include single work conditions and multiworking conditions. Bearing fault identification of small samples under multiworking conditions refers to the problem of predicting which fault type the test bearing samples belong to under a few fault training samples collected from complex work conditions. The problem is also called N -way and K -shot multiworking bearing fault identification problem, where the training data include N classes, each class only has K samples, and the K is no more than 20. The movement of the faulty bearing is nonlinear. In order to fit the nonlinear features included in the rolling bearing vibration signals, many studies used the neural network to learn bearing vibration signal features and got good results in bearing fault identification by establishing the mappings between the fault

feature and fault. For example, literature [1] and literature [2] used convolution neural networks (CNNs) to repeatedly learn the time-frequency features of a large number of original bearing vibration signals and realized the precise classification of bearing faults. The fault identification method based on the deep neural network usually requires a large amount of training data, and it is hard to play its ability to fit nonlinear features under small samples. Literature [3] and literature [4] show that the common problem of small sample learning is that the feature extraction effect is not very high and the identification accuracy is poor. Literature [5] shows that the bearing fault identification accuracy based on rolling bearing vibration signal analysis is affected by the bearing conditions. Rotation, load, and faults are the key factors affecting the motion state of the bearing rotor. When the speed of bearing continues to increase, due to the failure occurring in bearing such as touching, the nonlinear features of the rotor will further intensify, and the movement will

even evolve into chaotic motion. The bearing fault identification under multiworking conditions is more difficult than the identification in a single working condition. To improve the identification accuracy of the method based on deep neural networks for multiworking bearing faults, literature [6] and literature [7] used attention, label smoothness, and other auxiliary algorithms to optimize the training processes of the neural network. Under the limited training data condition, the identification performance of the optimized deep CNN has been improved, but the number of each class of training sample still exceeds 20. At present, the bearing fault identification of small samples under multiworking conditions is one of the research hotspots in deep neural network applications.

In recent years, some scholars have conducted research on bearing fault identification under small samples and multiworking conditions from different viewpoints. Literature [8] used CNN to extract the early bearing fault of the motor, and literature [9] used CNN to extract the features in the spectrum image of the bearing vibration signal. Different neural network structures have different capabilities for extracting bearing fault features, and the study of network structures to improve bearing vibration performance has received widespread attention. The training strategy of neural networks has also been the main direction of deep neural network research. MAML and SAMME are outstanding works in this area. Data preprocessing technology generates new training samples through geometric transformation, which was widely applied to prevent the neural network from overfitting in the process of learning small samples. Literature [10] researched data preprocessing methods based on the sparse autoencoder, and literature [11] used sliding sampling to enhance bearing vibration data. Now, fault identification of small samples based on learning strategies and data preprocessing is one of the important research directions for bearing fault diagnosis. In this paper, the fault signals collected from multiple working conditions with different speeds, different loads, and different fault degrees were used as the research objects. We first employed the coordinate delay phase space reconstruction (CDPSR) method to process bearing vibration signals, then designed a deep bearing fault identification neural network based on CNN and residual network to produce some classifiers and finally, through the Adaboost method, implemented an integrating multiply classifiers algorithm for training and integrating these bearing fault classifiers to form a stronger bearing fault classifier.

The rest of this paper is organized as follows. Section 2 introduces related technologies such as the CDPSR method, residual network, meta-SGD algorithm, and the Adaboost technology. Section 3 details our bearing fault identification method, Section 3.1 presents the processes of the data preprocessing and building new training set, Section 3.2 designs the network structure of the bearing fault classifier and its training method, and Section 3.3 develops the integrating step of multiple bearing fault classifiers; Section 4 describes the steps of conducting validation experiments using the artificial bearing fault dataset and the natural bearing fault dataset and discusses the experiment results;

Finally, Section 5 concludes this paper and highlights the significance of the proposed method.

2. Related Work

2.1. Phase Space Reconstruction. The coordinate delay method [12] is a specific implementation of the phase space reconstruction technology, which can be used to calculate the phase space coordinates of the bearing vibration time-series signals. Suppose $\{x(t)\}$ is a one-dimensional time-series signal, $0\tau, 1\tau, \dots, (m-1)\tau$ are delay time, and M is the embedded dimension of high-dimensional phase space. According to the coordinate delay method, the phase coordinates of M -dimensional can be represented as equation (1), where $i=0, \dots, M-1$.

$$X(t) = \begin{Bmatrix} X_0(t) \\ X_1(t) \\ \vdots \\ X_{m-1}(t) \end{Bmatrix} = \begin{Bmatrix} \{x(t), \dots, x(t+(m-1)\tau)\} \\ \{x(t+1), \dots, x(t+1+(m-1)\tau)\} \\ \vdots \\ \{x(t+m-1), \dots, x(t+m-1+(m-1)\tau)\} \end{Bmatrix}. \quad (1)$$

Solving the phase space coordinates of the time sequence requires τ and M . Calculating delay time τ has a variety of methods such as correlation coefficients and other methods based on mutual information judgment. The related function method for solving τ is simple and effective. The autocorrelation function of the time sequence $R(\tau) = (1/N)\sum_{i=1}^{N-\tau} x(i) \cdot x(i+\tau)$ is a function that changes with τ . When $R(\tau)$ drops to the $R(0) * (1 - e^{-1})1 - e^{-1}$, the time is the best delay time τ for reconstructing phase space [13].

When the delay time τ is known, the Cao algorithm [14] can be used to solve the embedding dimension M , and it can calculate the embedding dimension with a small amount of data. Denote the i -th d -dimensional reconstruction vector as $y_i(d) = \{x_i, x_{i+\tau}, \dots, x_{i+(d-1)\tau}\}$, $i = 1, \dots, N - (d-1)\tau$, and define $a(i, d) = ((\|y_i(d+1) - y_{n(i,d)}(d+1)\|) / (\|y_i(d) - y_{n(i,d)}(d)\|))$, $i = 1, \dots, N - \tau$, where $y_i(d+1)$ is the reconstruction vector of the i -th $(d+1)$ dimension. Since $n(i, d) \in \{1, \dots, n - d\tau\}$, $y_{n(i,d)}(d)$ is in the d -dimensional phase space, the mean of $a(i, d)$ is $E(d) = (1/N - d\tau)\sum_{i=1}^{N-d\tau} a(i, d)$, and $E_1(d) = E(d+1)/E(d)$. If $E_1(d)$ stops changing from a certain d_0 , then $d_0 + 1$ is the minimum embedding dimension to be found.

2.2. Convolution Neural Network. Convolution Neural Network [15] is made up of neurons that have learnable weights and biases; it uses the convolution layer and non-linear activation functions to abstract the original data layer by layer as features required for specific tasks to achieve mapping of features and targets. The CNN is a sequence of

layers that mainly includes the convolution layer, the pooling layer, and the fully connected layer. We can stack these layers to form a full CNN architecture. Since CNN has good nonlinear fitting capabilities, it was widely used in fields such as image feature extraction and voice feature analysis. There are many excellent CNN applications in the field of bearing fault identification [16, 17].

2.3. Residual Network. The residual refers to the gap between the observed value and the prediction value; literature [18] applied it to the neural network and proposed the concept of the residual block whose structure is shown in Figure 1.

Here, x_l and x_{l+1} represent the input and output of the residual blocks, $h(x)$ represents the observed value, and the residual block can be expressed as
$$\begin{cases} y_l = h(x_l) + F(x_l, w_l) \\ x_{l+1} = f(y_l) \end{cases}$$
.

If $h(\cdot)$ and $f(\cdot)$ are directly mapping functions, for example, $h(x) = x$ and $f(y_l) = \text{ReLU}(y_l)$, the gradient of the L -layer can pass to any layer that is shallower than it [19]. Therefore, as the number of network layers increases, the residual network will not degenerate, and the ability of fitting nonlinear features becomes stronger. So far, many studies were using residual networks to extract features of business data. For example, literature [20] studied using the SELU activation function to optimize the deep residual network, used the deep residual network model to analyze information dissemination in wireless networks, and obtained a complete media information dissemination prediction of wireless networks.

Resnet is a residual network structure stacked by the residual blocks, which has performed well in image classification applications [21, 22]. Literature [23] researched the advantage of Resnet-18 by comparing Resnet-18/50, VGG-19, and Googlenet and found that Resnet-18 has the advantages of short training time and high accuracy, developed a deep learning model base on ResNet-18 to diagnose the fan blade surface damage, and got good recognition effect. Literature [24] proposed a dual attention residual network that uses a residual module from Resnet18 to detect oil spills of various shapes and scales. Resnet-18 is an implementation of the Resnet model, and it consists of five parts including Conv1_x, Conv2_x, Conv3_x, Conv4_x, and Conv5_x. Each part has a convolution and constant block. Resnet-18 was designed for the classification of thousands of pictures, its network structure is complicated, and it needs to take a long time to train. Compared to the classification of thousands of pictures, the task of bearing fault identification under dozens of working conditions is small. Therefore, this paper deleted the Conv4_x and Conv5_x network structures in the Resnet-18, modified the size of the convolution kernel, reduced the complexity of the Resnet-18, and still retained the advantages of the residual network.

2.4. Meta-SGD Algorithm. The Model-Agnostic Meta-Learning (MAML) algorithm [25] proposed by Finn et al. can be used to train the model that can be optimized by gradient descent algorithm; it is an excellent algorithm in the field of meta-learning. The difference between the MAML

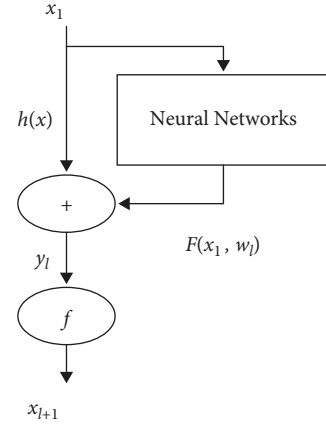


FIGURE 1: Structure of residual block.

and other optimization algorithms such as the stochastic gradient descent (SGD) algorithm [26] is that the optimization process in MAML is divided into two steps, which first assumes that n tasks ($T_i, i = 1, \dots, n$) are selected from the supporting dataset and each task is used to calculate the gradient based on the current neural network parameter $\theta^{(t)}$ to get n updated neural network parameter sets $\theta'_1, \dots, \theta'_n$, $\theta'_i = \theta^{(t)} - \alpha \nabla_{\theta^{(t)}} L(\theta^{(t)}, D_{T_i}^{\text{Support}})$, second calculates the update gradient on the query dataset, and determines the new neural network parameter $\theta^{(t+1)}$ according to $\theta^{(t+1)} = \theta^{(t)} - \beta \sum_{T_i \sim T} \nabla_{\theta^{(t)}} L(\theta'_i, D_{T_i}^{\text{Query}})$.

As the α and β hyperparameters are fixed in the MAML, they cannot be adjusted with the change of the network, which makes the training process fluctuate. The Meta-SGD algorithm [27] automatically adjusts the α parameter to increase the stability, whose execution steps are as follows.

Meta-SGD can learn task-agnostic features rather than simply adapt to task-specific features [28]. Literature [29] used the learning rate of the meta-learning SGD to predict streaming time-series data online and achieved good results.

2.5. Multiclass AdaBoost Method. AdaBoost [30] is a multimodel integration method, and its final output is the weighted sum of the results of the integrated multiple classifiers. AdaBoost was originally raised by Freund to research the binary classification problem. By introducing multiclassification index loss in the front-directional model, AdaBoost also can be applied to the multiclass problem. The sample weight is constantly updated in subsequent AdaBoost training, and samples that have not been identified correctly will be set up with greater weights. This process of AdaBoost training is shown in Figure 2.

SAMME [31] is an implementation algorithm of the multiclass AdaBoost, its specific implementation steps are as shown in Algorithm1.

Literature [32] hybridized the AdaBoost with a linear support vector machine model and developed a diagnostic system to predict hepatitis disease; the results demonstrate that the strength of a conventional support vector machine model is improved by 6.39%. Literature [33] studied the use

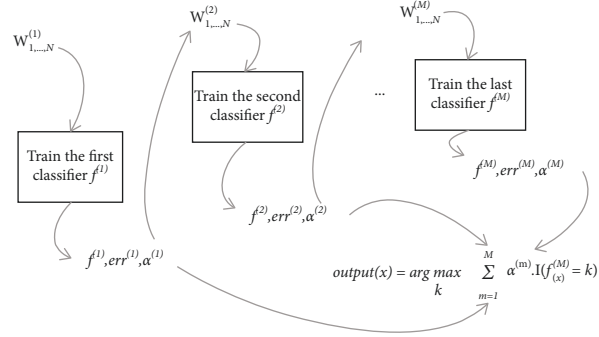
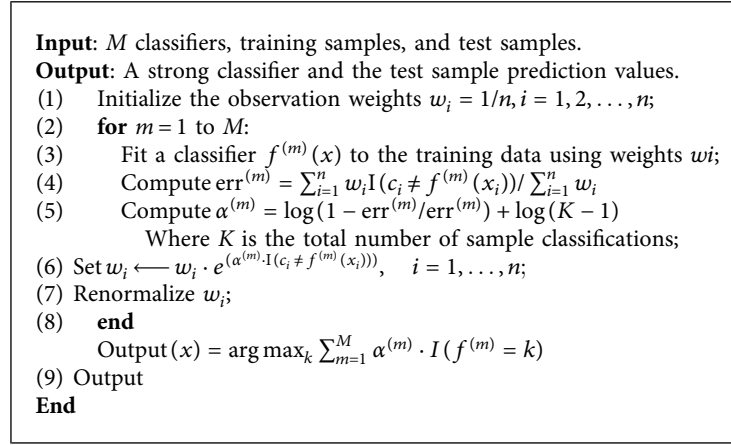


FIGURE 2: AdaBoost model structure.



ALGORITHM 1: SAMME.

of the AdaBoost framework to integrate other methods and proposed a wind turbine fault feature extraction method based on the AdaBoost framework and SGMD-CS. Experiments show that AdaBoost is an effective multimodal integration framework. To realize bearing fault identification, this paper studied the use of the AdaBoost framework to integrate multiple fault identification classifier models designed based on CNN and residual networks.

3. Our Method

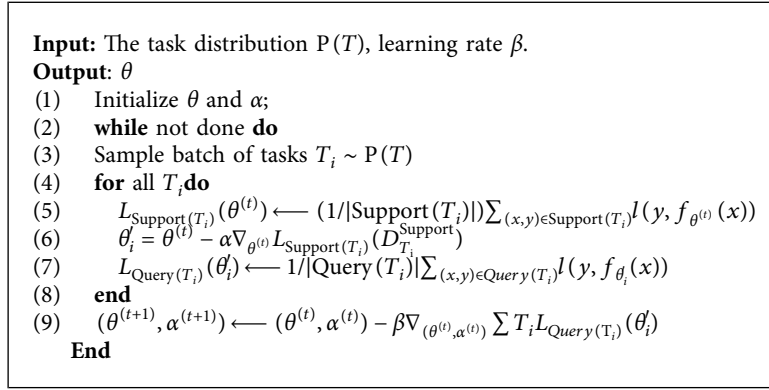
3.1. Data Preprocessing and Building New Feature Set. Use min-max normalization (2) to handle the training samples including N_{Train} samples and test samples including N_{Test} samples. Divide the training samples into a training set and verification set according to the ratio of 5:1. Select one sample separately from the training set, verification set, and test set as an input of Algorithm 2 to calculate the best value of time delay τ and phase space dimension M . Assume that the training sample has L data points. According to (1), we can get $(5/6)N_{\text{Train}} (L - M + \tau F / \tau FM)$ coordinates for the training set, $(1/6)N_{\text{Train}} (L - M + \tau F / \tau FM)$ coordinates for the verification set, and $N_{\text{Test}} (L - M + \tau F / \tau FM)$ coordinates for the test set, when the sampling frequency of the bearing vibration signal is F and the phase space vector is not reused.

$$\bar{x} = \frac{x - \min(x)}{\max(x) - \min(x)}. \quad (2)$$

Combine the reconstructed phase space coordinates with the labels of the original signals to build new training samples. Since the phase space has the same topological properties as the original bearing vibration signal system, the regularity of the bearing time-series signal in the high-dimensional space is restored in the coordinates. Therefore, any coordinate in phase space represents the state of the original bearing vibration signal system and contains corresponding features. Compared with the features included in the original signals, the features in the phase space coordinates are more obvious and easier to be identified by the classifier.

3.2. Bearing Fault Classifier and Its Training. Our bearing fault classifier is designed based on CNN and the residual block. The classifier uses 7 network layers, of which the Conv_x network consists of convolutional layers and ReLU activation operations. The full connection layer used 100 neurons, and the output layer used 4 neurons. The first layer of our network uses a larger convolution kernel and then reduces the size of the convolution kernel. The detailed structure is shown in Figure 3.

Our bearing fault classifier takes the phase space coordinates of the bearing vibration signal as input. In the classifier, the input matrix first flows through the Conv1_x



ALGORITHM 2: Mata-SG D.

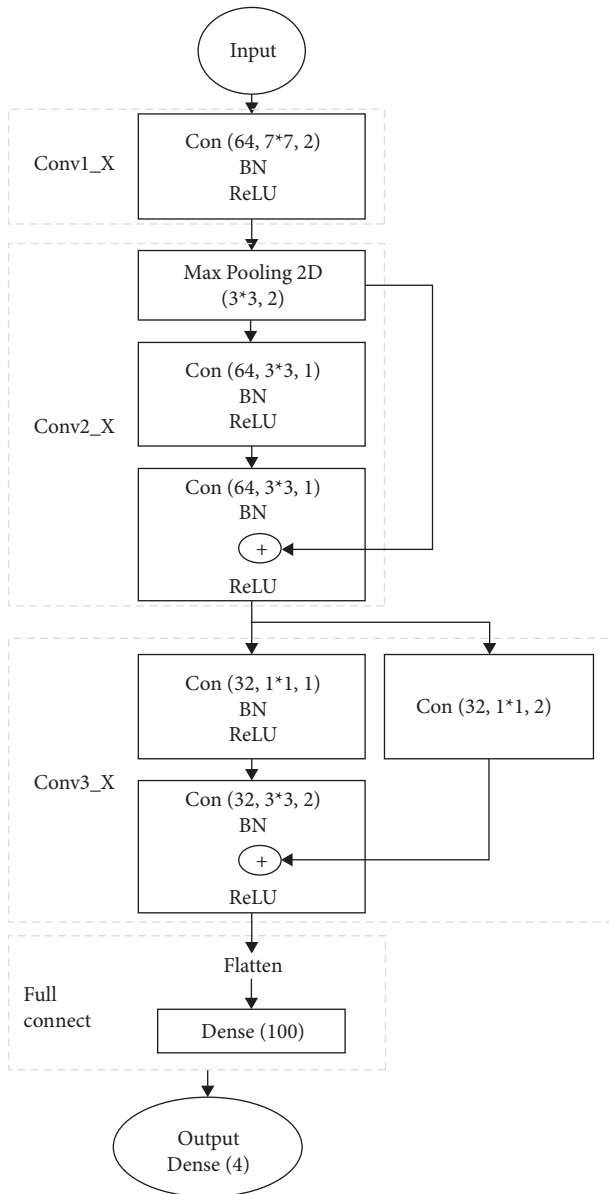


FIGURE 3: The detailed structure of the bearing fault classifier.

and then through Conv2_x, Conv3_x, the fully connected layer, and the output layer in sequence. In the Conv1_x module, the input matrix performed convolution operations according to (3) using 64 different convolution kernels with a size of $7 * 7$. To avoid the deviation of the result distribution after the convolution operation, the batch normalization (BN) technique is used to standardize the convolution result, so that the convolution result obeys a normal distribution with a mean of 0 and a variance of 1. Then, the result of BN is nonlinearly transformed using the ReLU activation function ($f(x) = \max(0, x)$). Different from the Conv1_x module, the Conv2_x module first uses the maximum pooling technique that takes the maximum value in a fixed-size sliding window to reduce the density of data features and then uses two convolution layers with smaller kernels to perform convolution operations on the input. After the second BN operation, Conv2_x uses the sum of the result of max-pooling (as the observed value of residual block) and the result of BN as the input to the nonlinear activation function. Conv3_x also has two convolution layers, the difference is that to make the observed value of the residual block of Conv3_x own the same shape as the second convolution result of Con3_x, the observed value is processed with a convolution of size $1 * 1$. In the fully connected layer and the output layer, the input is processed in the same way as equation (4).

$$y_{u,\mu}^{(l+1)} = \sum_{n=1}^K \sum_{m=1}^K \{x^{(l)}[n+u, m+\mu] \times w_i^{(l)}[n, m]\} + b^{(l)}, \quad (3)$$

$$y_i^{(l+1)} = \text{ReLU} \left(\sum_{j=1}^J x_j^{(l)} \times w_{i,j}^{(l)} + b_i^{(l)} \right). \quad (4)$$

In (3), $x^{(l)}$ represents the input matrix of the l -th layer, K is the size of the convolution kernel, and $w_i^{(l)}$ and $b^{(l)}$ represent the connection weight and activation parameter of the i -th convolution kernel of l -th layer (all convolution kernels of the same convolution layer shared the same activation value). In equation (4), i represents the serial number of neurons in the full connection layer, j means the

position number of the input matrix, $w_{i,j}^{(l)}$ represents the weight of the i -th neurons to the j -th value of the input, and $b_i^{(l)}$ is the bias of the i -th neurons of l -th layer.

To prevent overfitting in the training process, we added the following judgment statements after step 9 of Algorithm 2 and used the modified Algorithm 2 to update the network parameters of our bearing fault classifier.

```

If Acc_of_Train  $\geq$  0.9 and Acc_of_Train - Acc_of_
Validation > 0.1:
Number_of_ Overfitting ++;
else:
Number_of_ Overfitting = 0;
If Number_of_ Overfitting > 5:
break;

```

3.3. The Integrating Step of Multiple Bearing Fault Classifiers.

Combining multiple weak bearing fault identification classifiers designed in Section 3.2 can generate a stronger bearing fault identification classifier using the AdaBoost algorithm; however, unlike the SAMME, our method divides the training data into a support set and a query set, calculates each sample weight of the support and query set, and updates the classification error with the sample weights. The integrating steps of our bearing fault classifiers are shown in Algorithm 3.

This algorithm uses the Meta-SGD learning strategy to decrease the value of the objective loss function and update the network parameters; it not only has the characteristics of the meta-learning strategy to quickly converge but also adjusts the learning rate based on the learning task. The initialization of neural network parameters has an important impact on training. In order to make the network get good initialization parameters, the algorithm initializes the parameters of the next classifier in the parameters of the previous classifier during the iterative training process.

The algorithm needs to determine 2 hyperparameters in advance. They are the learning rate β of Algorithm 2 and the number of data points of the training sample. The two parameters have a great influence on the algorithm. First, the learning rate β will affect the learning speed of meta-SGD; second, the number of data points of the training sample determines the shape of the classifier input data, and the execution effect of the coordinate delay method algorithm will be affected by this parameter. If it is too large, it will increase the interference data, and if it is too small, the high-dimensional phase space of the bearing vibration time-series signal cannot be accurately established. In this paper, we set β with 0.001, and the number of the data points is 1024.

4. Experiment

The fault identification accuracy under different working conditions is an important indicator for measuring the rolling bearing fault identification method. This paper verified the effectiveness of the proposed method by calculating the test accuracies of bearing fault identification on

the artificial and natural bearing fault dataset collected from different loads, different speeds, and different fault conditions.

The experiments were conducted on the Tensorflow CPU 2.7 platform programming with python. The hardware and software environments included core i7-4790K 4.0 GHz processor, 16 G memory, and Windows Server 2018 operating system.

To eliminate the impact of accidentality, each experiment in this paper was performed 5 times independently, and the average value of 5 test accuracy was used as the experimental results.

4.1. Experiments on the CWRU Dataset. Artificial bearing fault data set CRWU [34] consists of bearing vibration time serial signals in the state of normal, internal circle fault, outer circle fault, and rolling bodies fault. The CWRU dataset is a representative data set in the field of bearing fault diagnosis. Many scholars got positive results when they used the CWRU to perform simulation experiments [35, 36].

4.1.1. Experimental Data. The bearing vibration signals used in this case were sampled from the 6205-2RSJEM SKF rolling bearings at a sampling frequency of 12 kHz. The fault types of these signals are associated with 4 different bearing damage diameters, recorded as A (0), B (0.007 inches), C (0.014 inches), and D (0.021 inches). These signals were cut into several segments. Each contained 1024 data points and was used as an experimental sample. The experiment samples are selected from the 16 kinds of work conditions described in Table 1 and are used to carry out three groups of small sample experiments. Three groups of experiments randomly selected 10 samples per class for testing, and these testing samples have the same labels and same working conditions as training samples.

4.1.2. Experiment and Result Analysis. According to Algorithm 3, three group experiments were conducted. The first group experiment was a variable power experiment. The samples of normal, inner circle fault, outer circle fault, and rolling element fault are randomly selected from 8 working conditions numbered 1, 5, 9, 13, 2, 6, 10, and 14 in Table 1 to carry out 4-way and 20-shot experiments. Similar to the first group experiment, the second group experiment was a variable fault degree experiment, in which samples were randomly selected under four operating conditions numbered 1, 2, 3, and 4. The third group of experiments was the variable power and fault degree experiments, and samples of five operating conditions of 1, 3, 4, 5, 10, and 14 were randomly selected. The sample distribution of the three experiments is shown in Table 2. Calculated the test accuracy and the test results are shown in Table 3.

In the three groups of experiments, the test accuracies of the proposed method are greater than 95% and the standard deviations of the accuracy of the three experiments are within 3 when the number of training samples of normal, inner ring fault, outer ring fault, and rolling element fault is

Inputs: Number of classifiers, training and test samples composed of rolling bearing vibration signals, and the sample labels.
Outputs: Bearing fault identification classifier and the prediction values of the test samples.

- (1) Set β with 0.001 for Algorithm 2;
- (2) **For** all classifiers **do**:
- (3) Decompose the bearing vibration signals and construct the new training set, verification set, and test set according to the data preprocessing step described in Section 3.1;
- (4) Divide the training set into support and query sets with the ratio of 1:1;
- (5) Initialize the sample weight of the support and query set with $w_i^{\text{support}} = (1/\text{Num}_s)$, $i = 0, \dots, \text{Num}_s - 1$; $w_{ji}^{\text{query}} = (1/\text{Num}_q)$, $j = 0, \dots, \text{Num}_q - 1$, where Num_s and Num_q are the numbers of the sample of the support set and query set;
- (6) Set the target loss function with the cross-entropy ($l_{CE}(p, q) = -\sum (p(x)\log(q(x)) + (1-p(x))\log(1-q(x))) \times w(x)$), where p is the predictive value, q is the true value, and w is the sample weight);
- (7) Update the parameters of the first classifier using the support set and query set according to the modified Meta-SGD;
- (8) Calculate the identification error rate of the training set according to equation (2);
- (9) Calculate the weight coefficient of the classifier according to equation (3);
- (10) Update the sample weight of the training sample, and normalize these weights according to (4);
- (11) Use the network parameters of the previous classifier to initialize the network parameters of the next classifier;
- (12) Train the next classifier using the training set updated with the new weights according to the modified Meta-SGD;
- (13) **end**
- (14) Calculate the prediction values of the test sample according to equation (5), output the prediction values and the integrated classifier.

End

ALGORITHM 3: Our bearing fault identification algorithm.

20 and in a variety of working conditions composed of three different factors of bearing fault degree, speed, and load. Figure 4 is the convergence process of loss value and validation accuracy under the 4-way and 20-shot experiments with variable power and fault degree.

During the training process of the first classifier, the value of the target loss function continued to decrease, the test accuracy rate continued to increase, and the value of the network parameters was continuously optimized. The test accuracy of this classifier exceeded 90%. After the first classifier, the four classifiers were fine-tuned according to the training dataset with different sample weights in the training of the second, third, fourth, and fifth four classifiers. Finally, five different classifiers were obtained, which complemented each other, their test accuracy was all above 90%, and the final test accuracy of our method reached 100%.

The above N -way and K -shot experiments show that the proposed method can filter the influence of the three factors of speed, load, and fault degree on the bearing fault identification to a certain extent whether it is applied in the constant or variable condition. The proposed method has good accuracy and stability for the fault identification of small sample bearings with different speeds, loads, and fault degrees.

4.2. Experiments on XJTU-SY Dataset. To further verify the effectiveness of the proposed method, the natural bearing fault data set, XJTU-SY [37], was used for our experiments.

4.2.1. Experimental Data. The sampling frequency is 25.6kHz in the XJTU-SY experiment, and the XJTU-SY dataset also includes the faults raised in the outer circle, inner circle, and rolling bodies. The bearing vibration signals

sampled from eight different conditions were randomly selected to form training sets and test sets. Every experiment sample has 1024 data points, and its type and its labels are shown in Table 4.

4.2.2. Experimental Results and Analysis. The 4-way and 20-shot experiment (experiment 4) is carried out by randomly selecting normal, inner fault, outer fault, and rolling element fault samples from the working condition numbered 18 in Table 4. The test results are shown in Table 5.

The results of experiment 4 show that the identification accuracy of the proposed method for natural bearing fault is still high under the small number of training samples, and the test accuracy of the 4-way and 20-shot is more than 96% when the training samples and test samples are in the same working condition. Figure 5 shows the confusion matrixes of the predicted value of the test sample in the 5 experiments.

From the confusion matrix, the predicted values of label 0 were always consistent with the actual values, and the prediction accuracy was 100%; the prediction accuracy of other labels fluctuated slightly, and the error mainly came from the wrong prediction of label 1 as a label 3. In the end, the prediction results of each classifier were excellent. The prediction accuracies of label 1 were above 90% for 4 consecutive times, the prediction accuracies of label 2 exceeded 99%, and the prediction accuracies of label 3 exceeded 90% four times. The recognition rate of 100% fault samples and nonfault samples means that the proposed method can accurately distinguish between fault samples and nonfault samples, and the fault location accuracy rate of 94.7% shows that the method also has a good ability to identify natural bearing faults under known working conditions.

TABLE 1: Work conditions of the CWRU experiment.

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
Load (hp)		1				2				3				0			
Motor speed (rpm)		1772				1750				1730				1797			
Damage degree	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	

TABLE 2: The bearing status, label, and quantity of the experimental samples.

No.	Condition no.	Normal		Inner fault		Outer fault		Rolling fault	
		Label	Quantity	Label	Quantity	Label	Quantity	Label	Quantity
1	1	0	5						
	5	0	5						
	9	0	5						
	13	0	5						
	2			1	5	2	5	3	5
	6			1	5	2	5	3	5
	10			1	5	2	5	3	5
	14			1	5	2	5	3	5
2	1	0	20						
	2			1	7	2	7	3	7
	3			1	7	2	7	3	7
	4			1	6	2	6	3	6
3	1	0	10						
	5	0	10						
	10			1	5	2	5	3	5
	14			1	5	2	5	3	5
	3			1	5	2	5	3	5
	4			1	5	2	5	3	5

TABLE 3: Results of the three group experiments.

Experiment number	4-way and 20-shot	
	Test accuracy	Standard deviation
1	$(92.5 + 97.0 + 95.0 + 100 + 97.5)/5 = 96.4$	2.5
2	$(97.5 + 92.5 + 100 + 92.5 + 95.0)/5 = 95.5$	2.9
3	$(92.5 + 100 + 95.0 + 92.0 + 95.0)/5 = 95.0$	2.7

Our method can be viewed as a combination of CDPSR data preprocessing, residual network, Meta-SGD, and AdaBoost, which is insensitive to changes in bearing load, rotational speed, and failure degree. To analyze each part in the proposed method, we performed the ablation experiments. First, we analyzed the contribution of the residual network designed in Section 3.2 by comparing two different types of combinations of CNN + CDPSR + Meta-SGD + AdaBoost and ResNet + CDPSR + Meta-SGD + AdaBoost, in which CNN consisted of a stack of simple convolutional layers and max-pooling layers. Second, the effect of the Meta-SGD learning strategy method is analyzed by comparing ResNet + CDPSR + Meta-SGD + AdaBoost and ResNet + CDPSR + SGD + AdaBoost, and the use of SGD was described in reference [23]. Then, by comparing Resnet + Meta-SGD + AdaBoost and ResNet + CDPSR + Meta-SGD + AdaBoost, we discussed the influence of the reconstructed bearing vibration timing signal on the method. Finally, the influence of AdaBoost on this method was analyzed by comparing ResNet + CDPSR + Meta-SGD and ResNet + CDPSR + Meta-SGD + AdaBoost. The test results of the ablation experiment are shown in Figure 6.

From the results in Figure 6, the bearing fault recognition accuracy of the classifier using the residual network was 6% higher than that of the CNN classifier, Meta-SGD had a greater improvement in bearing fault recognition than SGD, and the CDPSR method played a positive influence on the fault identification. From the variance of the test accuracy of ResNet + Meta-SGD + CDPSR and ResNet + Meta-SGD + CDPSR + AdaBoost, we can see that AdaBoost also played a positive role as a stabilizer for the proposed method. Hence, the four parts of CDPSR, residual network, Meta-SGD, and AdaBoost had made positive contributions to the proposed method. Residual network and Meta-SGD can effectively improve the test accuracy; CDPSR data preprocessing and Adaboost have significant impacts on the stability of the method.

To analyze the advantages of our method, the comparison experiment was carried out under the same conditions as experiment 4. The test accuracy of the proposed method, the WDCNN method [16], and the CNN-SVM method [17] would be compared.

The WDCNN method proposed by Zhang et al. in 2017 included a specific deep network (WDCNN network), which

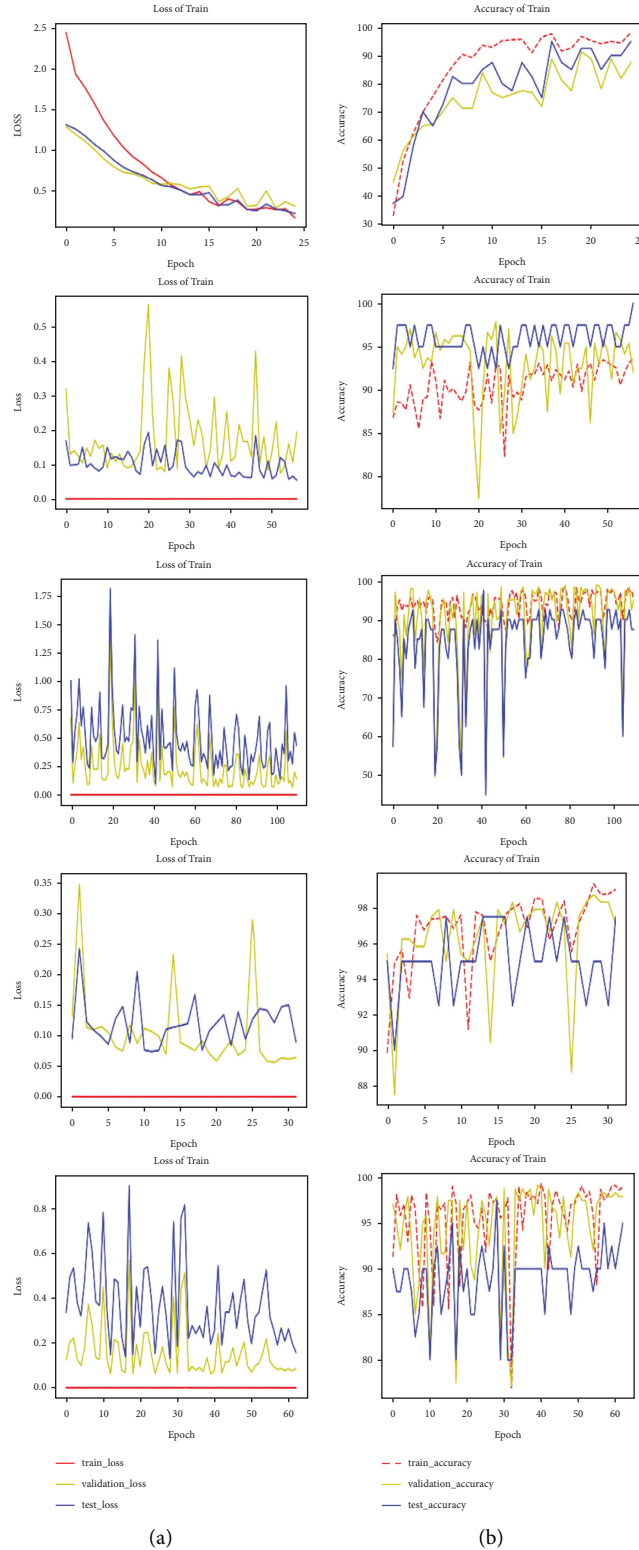


FIGURE 4: Convergence process of loss value and validation accuracy metrics in one training session. (a) Convergence process of loss value. (b) Convergence process of verification accuracy.

takes the wide kernels in the first as convolutional layers and small convolutional kernels in the preceding layers. The method slices the training samples with overlap to obtain huge amounts of data, then uses raw vibration signals as

input, and trains the WDCNN network using the back-propagation algorithm detailed in reference [16].

The CNN-SVM method combines both the merits of CNN and SVM, which firstly uses the 2D representation of

TABLE 4: Bearing fault state and its label of XJTU-SY dataset.

Condition no.	Load (kN)	Rotating speed (rpm)	Bearing state			
			Normal	Inner fault	Outer fault	Rolling element fault
17	10 (<i>h3</i>)	2400		1	2	
18	11 (<i>h2</i>)	2250	0	1	2	3
19	12 (<i>h1</i>)	2100	0			3

TABLE 5: Test results of experiment 4.

Test accuracy	Standard deviation
$(95 + 96.25 + 97.5 + 96.5 + 95)/5 = 96.05$	1.2

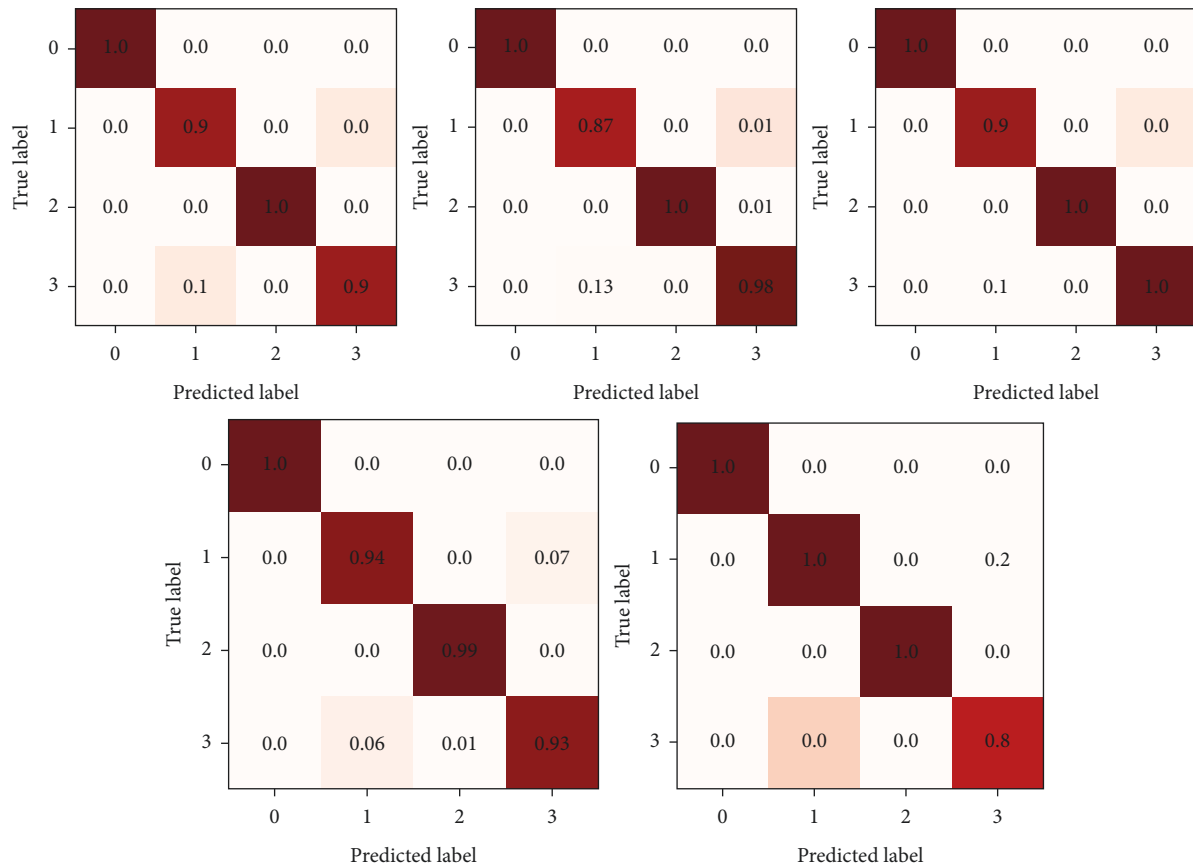


FIGURE 5: Confusion matrix of test samples in 4-way and 20-shot of experiment 4.

raw vibration signals as input, then trains the original CNN with the output layer for several epochs until the training process converges using stochastic gradient descent, and finally replaces the output layer with the SVM, which included the radial basis function (RBF) kernel. The optimum scheme used in this method was elaborated in reference [17].

The WDCNN and the CNN-SVM method are typical methods in the application of fault-bearing recognition. The fault identification results of the comparison experiment are shown in Figure 7.

In the comparison, the residual network designed for our method and the network used in the WDCNN method have the same number of network layers and the same scale of learning parameters. The average result of the five CNN-SVM experiments was 82.5%, and the average result of the WDCNN was 90.5%. The test accuracies of both the WDCNN and our method were above 90%, and our result was 96%. Compared with the test accuracy of CNN-SVM and WDCNN, the test accuracy of our method was higher. Moreover, the test results of our method were 95 to 97.5, the

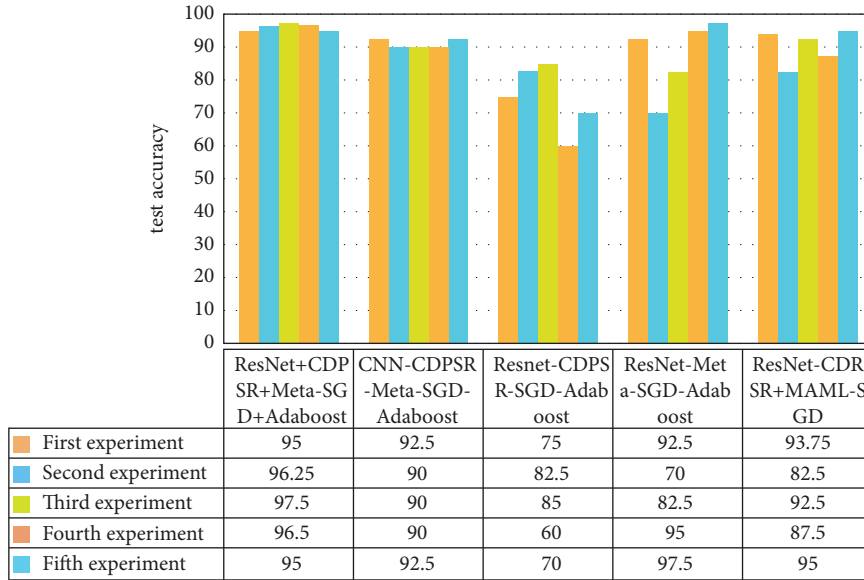


FIGURE 6: The results of the ablation experiments.

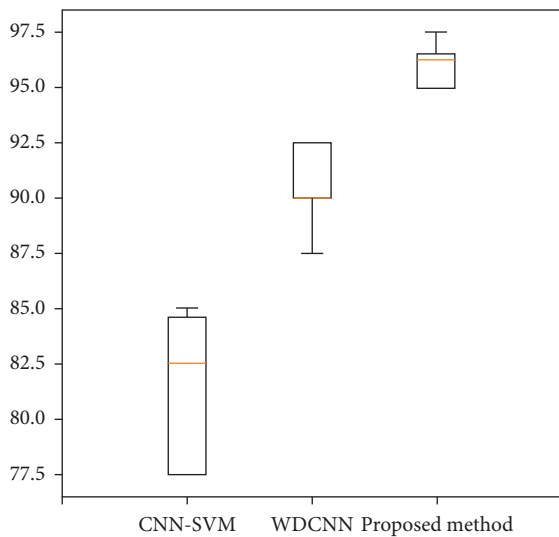


FIGURE 7: Results of the comparison experiment.

value range was relatively narrow, while the value range of the other two methods was relatively much wider. Therefore, the proposed method has a certain improvement in the accuracy of bearing fault identification, and its stability is better.

5. Conclusions

A novel bearing fault identification method for multiconditions and small samples was proposed to challenge the problems of lacking fault data and poor performance. To verify the effectiveness of the proposed method, the artificial and natural bearing fault signals were taken to experiment with as a case study. The result shows that the proposed method realized accurate fault signal identification under multiple working conditions and small samples, and its accuracy rate of bearing fault positioning exceeds 90%. Benefitting from the

reconstruction of high-dimensional space of bearing vibration time series by coordinate delay construction method, extraction of the phase space features using the convolutional neural network, the transmission of the gradient to other layers by residual block, updation of the classifier parameters by Meta-SGD, and integration of multiple classifiers by AdaBoost method, the proposed method gets excellent bearing fault feature extraction and high fault identification ability. Finally, compared with other advanced methods, the proposed method also has certain advantages. From these cases, the proposed method is very effective.

The proposed method can accurately identify bearing faults under small samples and multiworking conditions without manually setting fault features. Therefore, the proposed method has a certain value in some application areas with complex working conditions and difficulty obtaining a large number of bearing fault samples, such as aviation bearings.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Growth Project of Young Scientific and Technological Talents in Guizhou for Colleges and Universities (grant no. Qian Jiao He KY2020137) and the tripartite joint fund project of Guizhou Provincial Department of Science and Technology (grant no. LH2016 7275).

References

- [1] Y. Xin, S. Li, C. Cheng, and J. Wang, "An Intelligent Fault Diagnosis Method of Rotating Machinery Based on Deep Neural Networks and Time-Frequency Analysis," *Journal of Vibroengineering*, vol. 20, no. 6, pp. 2321–2335, 2019.
- [2] D. Kolar, D. Lisjak, M. Pajak, and M. Gudlin, "Intelligent fault diagnosis of rotary machinery by convolutional neural network with automatic hyper-parameters tuning using bayesian optimization," *Sensors*, vol. 21, no. 7, p. 2411, 2021.
- [3] Y. U. You, L. Feng, and G. Wang, "Few-shot learning method based on deep network," *Journal of Chinese Computer Systems*, 2019.
- [4] C. Lu, Z. Wang, and B. Zhou, "Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification," *Advanced Engineering Informatics*, vol. 32, pp. 139–151, 2017.
- [5] H. Shao, M. Xia, G. Han, Y. Zhang, and J. Wan, "Intelligent fault diagnosis of rotor-bearing system under varying working conditions with modified transfer CNN and thermal images," *IEEE Transactions on Industrial Informatics*, vol. 17, pp. 3488–3496, 2020.
- [6] X. Li, W. Zhang, and Q. Ding, "Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism," *Signal Processing*, vol. 161, pp. 136–154, 2019.
- [7] C. Wang and Z. Xu, "An Intelligent Fault Diagnosis Model Based on Deep Neural Network for Few-Shot Fault Diagnosis," *Neurocomputing*, vol. 456, pp. 550–562, 2021.
- [8] P. Kumar and A. Shankar Hati, "Convolutional neural network with batch normalisation for fault detection in squirrel cage induction motor," *IET Electric Power Applications*, vol. 15, no. 1, pp. 39–50, 2021.
- [9] K. K. Krishnan and K. P. Soman, "CNN Based Classification of Motor Imaginary Using Variational Mode Decomposed EEG-Spectrum Image," *Biomedical Engineering Letters*, vol. 11, pp. 1–13, 2021.
- [10] M. Y. Qiao, X. Y. Tang, and S. H. Yan, "Bearing fault diagnosis based on improved sparse filtering and deep network fusion," *Journal of Zhejiang University*, vol. 54, no. 12, 2020.
- [11] Y. Xin and S. Li, "Novel data-driven short-frequency mutual information entropy threshold filtering and its application to bearing fault diagnosis," *Measurement Science and Technology*, vol. 30, no. 11, p. 115006, 2019.
- [12] J. F. Gibson, J. Doyne Farmer, M. Casdagli, and S. Eubank, "An analytic approach to practical state space reconstruction," *Physica D: Nonlinear Phenomena*, vol. 57, no. 1-2, pp. 1–30, 1992.
- [13] M. T. Rosenstein, J. J. Collins, and C. J. De Luca, "A Practical Method for Calculating Largest Lyapunov Exponents from Small Data Sets," *Physica D: Nonlinear Phenomena*, vol. 65, 1993.
- [14] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D: Nonlinear Phenomena*, vol. 110, no. 1-2, pp. 43–50, 1997.
- [15] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [16] W. Zhang, G. L. Peng, C. H. 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, p. 425, 2017.
- [17] J. Xu, L. Ma, W. Zhang, Q. Yang, X. Li, and S. Liu, "An improved hybrid CNN-SVM based method for bearing fault diagnosis under noisy environment," in *Proceedings of the 31st Chinese Control and Decision Conference CCDC*, pp. 4660–4665, Nanchang China, June 2019.
- [18] K. He, X. Zhang, S. Ren, and J. Sun, "Identity Mappings in Deep Residual Networks," in *Proceedings of the European Conference on Computer Vision*, Springer, Cham, Switzerland, September 2016.
- [19] A. Veit, M. Wilber, and S. Belongie, "Residual networks behave like ensembles of relatively shallow networks," *Advances in Neural Information Processing Systems*, 2016.
- [20] X. Lv and D. Chun, "Media information dissemination model of wireless networks using deep residual network," *Mobile Information Systems*, vol. 2021, no. 3, pp. 1–10, Article ID 1711944, 2021.
- [21] M. Sulc, D. Mishkin, and J. Matas, "Very deep residual networks with MaxOut for plant identification in the wild," in *Proceedings of the Working Notes of CLEF 2016 - Conference and Labs of the Evaluation Forum*, vol. 1609, Evora, Portugal, September 2016.
- [22] M. O. Ramkumar, S. Catharin, V. Ramachandran, and A. Sakthikumar, "Cercospora identification in spinach leaves through resnet-50 based image processing," *Journal of Physics: Conference Series*, vol. 1717, no. 1, p. 012046, 2021.
- [23] H. Zhang and F. Wang, "Fault Identification of Fan Blade Based on Improved ResNet-18," vol. 2221, IOP Publishing Ltd, Article ID 012046, 2022.
- [24] X. Li, X. Liu, Y. Xiao, Y. Zhang, X. Yang, and W. Zhang, "An improved U-net segmentation model that integrates a dual attention mechanism and a residual network for transformer oil leakage detection," *Energies*, vol. 15, no. 12, p. 4238, 2022.
- [25] C. Finn, P. Abbeel, and S. Levine, "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks," vol. 70, pp. 1126–1135, 2017.
- [26] S. Bonnabel, "Stochastic gradient descent on Riemannian manifolds," *IEEE Transactions on Automatic Control*, vol. 58, no. 9, pp. 2217–2229, 2013.
- [27] Z. Li, F. Zhou, and C. Fei, "Meta-SGD: Learning to Learn Quickly for Few-Shot Learning," 2017, <https://arxiv.org/abs/1707.09835>.
- [28] T. Starshak, "Negative Inner-Loop Learning Rates Learn Universal Features," *Machine Learning*, arXiv e-prints, arXiv e-prints. 2022, <https://arxiv.org/abs/2203.10185>.
- [29] W. Zhang, "POLA: Online Time Series Prediction by Adaptive Learning Rates," in *Proceedings of the ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, Canada, June 2021.
- [30] Y. Freund and R. E. Schapire, "Experiments with a New Boosting Algorithm," *Experiments with a new boosting algorithm*, 1996.
- [31] T. Hastie, S. Rosset, J. Zhu, and H. Zou, "Multi-class AdaBoost," *Statistics and Its Interface*, vol. 2, no. 3, pp. 349–360, 2009.
- [32] W. Akbar, W. P. Wu, S. Saleem et al., "Development of hepatitis disease detection system by exploiting sparsity in linear support vector machine to improve strength of AdaBoost ensemble model," *Mobile Information Systems*, vol. 2020, Article ID 8870240, 9 pages, 2020.
- [33] H. Li, F. Li, R. Jia, F. Zhai, L. Bai, and X. Luo, "Research on The fault feature extraction of rolling bearings based on SGMD-CS and the AdaBoost framework," *Energies*, vol. 14, no. 6, Article ID 1555, 2021.
- [34] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: a

- benchmark study,” *Mechanical Systems and Signal Processing*, vol. 64, pp. 100–131, 2015.
- [35] D. Neupane and J. Seok, “Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: a review,” *IEEE Access*, vol. 8, pp. 93155–93178, 2020.
- [36] T. W. Rauber, F. D. A. Boldt, and F. M. Varejao, “Heterogeneous feature models and feature selection applied to bearing fault diagnosis,” *IEEE Transactions on Industrial Electronics*, vol. 62, no. 1, pp. 637–646, 2015.
- [37] B. Wang, Y. Lei, and N. Li, “A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings,” *IEEE Transactions on Reliability*, vol. 69, pp. 401–412, 2018.