Received April 21, 2021, accepted May 6, 2021, date of publication May 11, 2021, date of current version May 21, 2021.
Digital Object Identifier 10.1109/ACCESS.2021.3079286

Feature Classification Method of Frequency Cepstrum Coefficient Based On Weighted Extreme Gradient Boosting

MINGSI QI[1], RUI ZHOU[1], QIANQIAN ZHANG[2], AND YONGSHENG YANG[1]
[1]School of Mechanical Engineering, North University of China, Taiyuan 030051, China
[2]School of Automation and Software, Shanxi University, Taiyuan 030006, China
Corresponding author: Mingsi Qi (qms0712@163.com)

This work was supported in part by the Shanxi Provincial Natural Science Foundation of China under Grant XJJ201802.

ABSTRACT Cepstrum coefficient and decision tree classification provide a feasible method for mechanical fault diagnosis. However, this method has the problems of low accuracy of feature extraction and low reliability of feature classification. The efficiency and classification results of this method are not ideal. Therefore, this paper proposes a feature classification method of frequency cepstrum coefficients based on weighted extreme gradient boosting. The proposed method is mainly divided into two parts: feature extraction and feature classification. In the part of feature extraction, Considering the limitation of Mel frequency scale in MFCC, an adaptive frequency cepstrum coefficient (AFCC) method is proposed, which can adaptively obtain the frequency scale of the corresponding signal, thus improving the feature accuracy extracted by MFCC method. In the aspect of feature classification, XGboost algorithm has strong learning ability, which easily leads to over fitting and affects the classification accuracy, A weighted extreme gradient boosting (WXGB) method is proposed, and an improved distance evaluation method (IDE) is introduced to modify the feature weights learned by XGboost, so as to improve the prediction accuracy of the classification model. In order to verify the effectiveness of the method, the vibration signals of rolling bearings with different fault types are taken as samples to carry out feature extraction and classification. The results of feature extraction and classification of different methods are compared, and the effectiveness of the proposed feature classification method based on weighted extremum gradient boosting frequency cepstrum coefficient in rolling bearing fault diagnosis is verified.

INDEX TERMS Feature extraction, fault diagnosis, MFCC, XGboost.

I. INTRODUCTION

Rotating machinery structure is a common mechanical equipment structure in modern industrial production. Rotating machinery structure is the part with high damage rate in all kinds of mechanical structure faults [1]. Timely detection of rotating machinery faults can avoid property losses and personal safety accidents, so it is of great significance to analyze rotating machinery faults [2]. Rotating machinery is mainly composed of gearbox, rotor, shaft and rolling bearing. Rolling bearing is the most widely used part in rotating machinery [3]. Rolling bearing has rich fault types and many analysis methods. The fault analysis of rolling bearing can quickly and accurately determine the possible fault problems in rotating machinery [4]. In recent years, the fault diagnosis of rolling bearing has been improved the research of obstacle diagnosis has attracted much attention [5].

There are a large number of time-frequency features of vibration signals. The methods based on time-frequency feature classification are usually feature selection, feature fusion and weighting [6]. The characteristics of these methods are to collect a large number of time-frequency vibration features as samples, and then select sensitive features based on mechanical vibration characteristics [7]. For example, Fatima et al for the multi-channel vibration signal features of bearing signal, the compensation technology is used to select the sensitive features, and finally a variety of SVM support vector machines are used to classify the signal features [8]. A large number of feature types and the selection of sensitive features make the use of fault classification method based on

The associate editor coordinating the review of this manuscript and approving it for publication was Cesar Vargas-Rosales.

time-frequency features limited and inefficient [9]. Among a large number of feature extraction methods, a feature extraction method based on speech vibration signal, Mel frequency cepstrum coefficient (MFCC), is noticed [10]. MFCC is widely used in speech denoising and human voice feature recognition [11], MFCC can extract features of a variety of vibration signals, this method has high efficiency in feature extraction, and in most cases, this kind of features has high accuracy in classification [12]. In the field of fault diagnosis, Jiang *et al.* proposed the method of feature extraction and classification of rolling bearing vibration signals by MFCC method [13]. Due to the limitation of Mel frequency scale, the result of feature extraction of vibration signals other than speech signals by MFCC is not ideal [14]. In order to effectively extract signal features by MFCC, Wang *et al.* used the second-order difference of MFCC features to obtain[15]. However, the MFCC method is limited to the frequency conversion relationship [16]. In order to make full use of MFCC method to extract vibration signal features, we need to improve its Mel frequency scale [17]. Pliwak proposed an adaptive frequency scale (SFCC) based on human voice to improve the MFCC feature extraction method [18], and achieved good results compared with the traditional MFCC feature extraction results. For the rolling bearing vibration signal under complex conditions, the use of MFCC is limited, so this paper proposes an adaptive frequency cepstrum coefficient (AFCC) method to extract the MFCC features of rolling bearing vibration signal, so as to effectively extract the features of rolling bearing vibration signal.

There are many classification methods for fault features, and MFCC features are suitable for a variety of classification methods; Bakhshi *et al.* used genetic optimization algorithm to classify MFCC features [19]; Ahmad *et al.* used k-nearest neighbor SVM support vector machine to classify MFCC features of heart sound signal [20], and obtained 92.5% classification accuracy. Abdul used SLTM classifier to classify MFCC features of bearing vibration signal [21]. In 2014, Chen *et al.* proposed to use CNN convolutional neural network to classify the features of MFCC, and obtained 98% classification accuracy [22]. In these classification methods, the method with high accuracy often relies on a large number of data samples [23], which limits the data types that are difficult to obtain a large number of data and need to be quickly classified XGboost algorithm can quickly and accurately complete the construction of two-dimensional data classification model [24], and MFCC features can be transformed into two-dimensional matrix data through discrete cosine change [25]. Therefore, using XGboost to classify MFCC features has significant advantages. Harar *et al.* tried to use XGboost limit gradient lifting and RF random forest method to classify MFCC features, which proved XGboost [26]. In order to solve this problem, this paper proposes the weighted extreme gradient boosting (WXGB) method [27], which can improve the classification ability of MFCC features. WXGB method through the introduction of improved distance evaluation (IDE) to solve the over fitting problem of XGboost learning model [28], obtains the importance weight between features through IDE, which is used to modify the importance of features obtained by XGboost self-learning, so as to improve the classification effect of XGboost method on MFCC features. In this paper, firstly, AFCC method is proposed based on the principle of MFCC method, which can be applied to the time-frequency feature extraction of vibration signal of rotating machinery. In addition, XGboost algorithm is improved combined with distance evaluation, considering the different importance of different dimension features to the results. Compared with other fault diagnosis methods, the flow of this method is more efficient and the results are more accurate.

In the second part of this paper, the principle and improvement method are introduced in detail. In the third part of this paper, two groups of experimental cases are used to verify the proposed improvement method. In the first case, the improved method is used to process four types of rolling bearing fault vibration signal samples, and the feasibility of the method is verified. In the second case, the method is verified by the experimental results four types of rolling bearing faults are obtained by setting control experiments, and the improved method is used to deal with them, which shows the ability of this method in practical application. The fifth part is the conclusion.

II. THEORY AND IMPROVED METHOD

A. MEL FREQUENCY CEPSTRUM COEFFICIENT

Mel frequency cepstrum coefficient (MFCC) is a feature extraction method of vibration signal in time and frequency domain. Firstly, the vibration signal is divided into frames and windowed, and then the frequency spectrum of each frame is obtained by short-time Fourier transform. The short frame spectrum is divided by Mel frequency scale to highlight the signal characteristics in the short frame spectrum. Finally, the frequency spectrum is divided by discrete cosine transform, the spectral feature is reduced to a low dimensional feature form, which is called MFCC feature. By taking the number of frames as the abscissa of the coordinate axis and the extracted spectrum feature information as the ordinate of the coordinate axis, the frequency cepstrum feature information of the signal can be represented by a two-dimensional graph.

The conversion relationship between Mel scale and actual frequency as:

$$\mathrm{Mel(f)} = 1125^{*}\ln(1 + \mathrm{f}/700) \tag{1}$$

Mel frequency scale is an empirical conversion formula based on the spectrum characteristics of speech signal, which mainly corresponds to the frequency domain characteristics of speech signal. However, for rolling bearing fault diagnosis, MFCC method is limited. Mel scale based on speech signal is not suitable for the frequency distribution of rolling bearing vibration signal, so a frequency scale division corresponding to rotating machinery vibration signal is needed. In addition, MFCC also has some shortcomings in denoising and

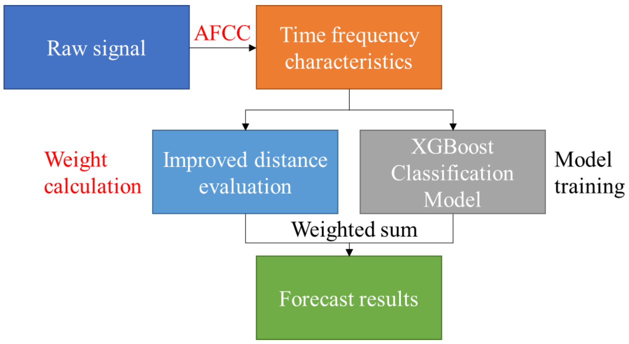


FIGURE 1. Method flow chart.

feature selection. In order to meet the requirements of feature extraction of rolling bearing vibration signal, MFCC method is improved: firstly, on the basis of MFCC, the impact component of the spectrum is improved, which is used to improve the peak value of the spectrum of the feature part, improve the signal-to-noise ratio, and ensure the effective feature extraction, in order to solve the problem of the unadaptability of the rate scale in Mel frequency, an adaptive method to obtain the distorted frequency scale is proposed to replace the Mel scale, and the filter is reconstructed by the obtained distorted frequency. Therefore, the time-frequency feature extraction of rolling bearing vibration signal can be completed adaptively by the improved method.

The flow chart of the proposed improvement method is shown in Figure 1.

B. AN ADAPTIVE FREQUENCY CEPSTRUM COEFFICIENT IS PROPOSED

Based on the feature extraction method of rolling bearing vibration signal by MFCC, this paper proposes an improved method of adaptive frequency scale acquisition, so as to improve the accuracy of feature extraction.

The signal feature extraction process is shown in Figure 2.

1)The preprocessed signal is divided into frames and windowed.

Firstly, the signal is divided into N frames, and each frame is windowed as:

$$w(n) = \frac{1}{2}[1 + \cos(\frac{2\pi n}{N-1})] \tag{2}$$

Equation (2) is Hamming window function, Where n belongs to $[-\frac{(N-1)}{2}, \frac{(N-1)}{2}]$;

$$u_n = w(n)^* y \tag{3}$$

In equation (3), u_n is the time domain signal of the nth frame after windowing;

The signal is divided into several small segments and processed with short-time Fourier transform. The reason is that the signal is converted into time-frequency domain. The requirement of framing is that each frame should contain at least one period of the signal. The actual rotation speed of the signal can be obtained during the analysis, so as to obtain the periodic information of the signal. In order to prevent the loss of information, frame shift operation is needed. After framing, Hamming is added to each segment of the signal, The purpose of windowing is to split the longer signal segment, so as to facilitate the signal processing.

2) Short time Fourier transform:

$$U_n(\omega) = \int_{-\infty}^{+\infty} u_n(t)e^{-i\omega t}dt \tag{4}$$

where $u_n(t)$ is the time-domain information of the nth frame of the signal, $U_n(\omega)$ is the frequency domain information corresponding to the nth frame.

By transforming the processed time domain signal segment into the corresponding frequency domain, the time-frequency characteristic spectrum of the signal can be obtained.

3) The distortion frequency scale of the signal is obtained.

Firstly, the average spectrum of each segment is calculated, and the new frequency distribution is obtained through the average spectrum:

$$F(\omega) = \frac{1}{n}\sum\nolimits_{n=1}^{N} U_n(\omega) \tag{5}$$

Frequency division is performed on the average power spectrum, calculate the area A of the average power spectrum, divide the interval according to the number f_n of sampling points on the spectrum, calculate the average interval area $\bar{a}$, match the spectrum area according to the average area, and get the reallocated frequency distribution:

$$A = \int F(\omega)d\omega \tag{6}$$

$$\bar{a} = \frac{A}{f_n - 1} \tag{7}$$

Sampling points are taken according to the average area:

$$\bar{a} = \int_{p_i}^{p_{i+1}} \overline{X}(\omega)d\omega, \quad i = 0, 1, \ldots, f_n \tag{8}$$

By $p_1 = 0$, it can be calculated in turn the value of p_i. Taking the permutation as the abscissa and p_i as the ordinate respectively, the frequency distribution of redistribution is obtained:

$$P = [p_1, p_2, p_3, \ldots, p_{f_n},] \tag{9}$$

The filter obtained according to the adaptive frequency conversion relationship is shown in Fig. 3.

4) The filter bank is reconstructed according to the obtained frequency scale.

Firstly, the center frequencies of M filters are obtained from the frequency scale P according to the set number m of filters, the center frequency sequence of each filter $Z(m)$:

$$Z(m) = [z_1, z_2, z_3, \ldots, z_m] \tag{10}$$

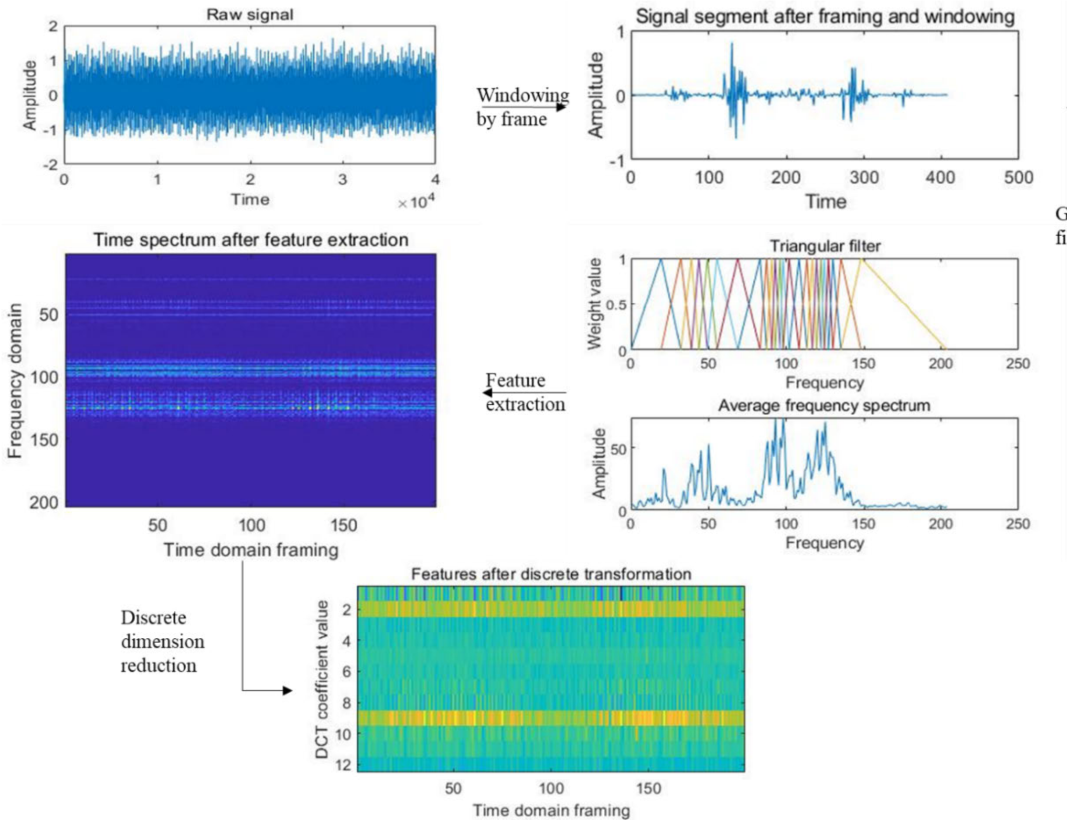


FIGURE 2. Feature extraction process.

$$H_m(k) = \begin{cases} 0, & k < Z(m-1) \\ \dfrac{2(k - Z(m-1))}{(Z(m+1) - Z(m-1))(Z(m) - Z(m-1))}, & Z(m-1) \le k < Z(m) \\ \dfrac{2(Z(m+1) - k)}{(Z(m+1) - Z(m-1))(Z(m) - Z(m-1))}, & Z(m) \le k < Z(m+1) \\ 0, & k > Z(m+1) \end{cases} \quad (11)$$

$H_m(k)$ is the transfer function of the filter, where $\sum_{m=1}^{M} H_m(k) = 1$;

5)The logarithmic energy of each frame spectrum is calculated according to the obtained filter:

$$s(m) = \ln(\sum_{k=0}^{K-1} |U_n(k)|^2 {}^* H_m(k)), \quad 0 \le m \le M \quad (12)$$

6) The amount of logarithmic energy data extracted is large, which is not conducive to signal feature classification. Therefore, the dimension of feature is reduced by discrete cosine transform to obtain a limited number of signal features:

$$C(n) = \sum_{k=0}^{K-1} s(m)^* \cos(\frac{\pi n(m-0.5)}{M}), \quad n = 1, 2, \ldots, L \quad (13)$$

L is the number of coefficients obtained by DCT, and $C(n)$ is the characteristic of cepstrum coefficients.

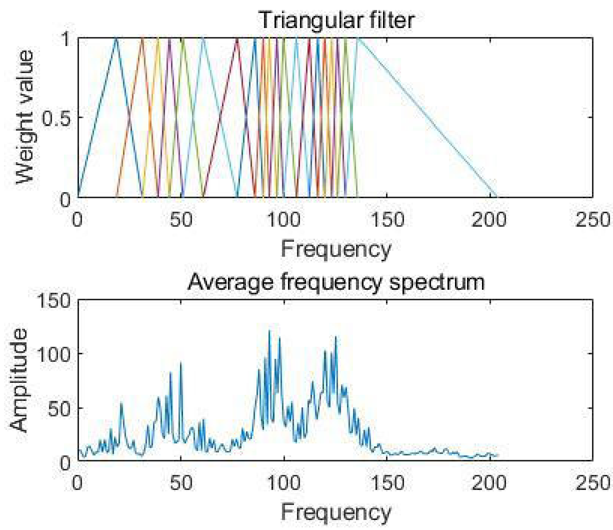


FIGURE 3. Improvement effect of triangular filter.

C. PROPOSED XGboost FEATURE CLASSIFICATION METHOD BASED ON DISTANCE EVALUATION

XGboost is used to train and classify the AFCC results, and an improved distance evaluation method is used to calculate the weight of the coefficients in the feature to improve the possible over fitting caused by XGboost.

XGboost can adaptively complete the classification learning of samples with a certain number of features, but the strong learning ability of XGboost often leads to over fitting, which affects the classification results of samples. Therefore, additional conditions are needed to limit the learning ability of XGboost model. The improved distance evaluation algorithm can determine the distance by calculating the differences between features and sample categories. The improved distance evaluation method is used to obtain the influence weight of sample features in the model tree, so as to improve the over fitting learning of XGboost and improve the classification effect of XGboost.

The flow chart of the proposed method is shown in Figure 3. XGboost can quickly complete the classification of sample features by decision tree binary classification. Compared with other decision tree models, the accuracy and calculation speed of this model are faster. The combination of the weights between features obtained by the improved distance evaluation algorithm and the weight parameters in XGboost classification tree model can ensure the correlation between data, so as to prevent over fitting in XGboost method. The calculation process of the parameters in the model is as follows:

1) Input the feature data set, the sample category in the data set is Y, the number of samples in each category is N, and the number of features in each sample is M:

$$D = \{X_{y,n,m}, Y\} \tag{14}$$

where $X_{y,n,m}$ is the N -th sample in the Y -class of the dataset, which contains M features, and Y is the category of the corresponding sample.

2) The weight of the improved distance evaluation coefficient is as follows.

Calculate the standard deviation between the data in the coefficient:

$$\delta_{y,m} = \sqrt{\frac{\sum_{n=1}^{N}(x_{y,n,m} - u_{y,m})^2}{N-1}} \tag{15}$$

where $u_{y,m}$ is the average value of the data in the m-th feature in the dataset:

$$u_{y,m} = \frac{1}{N}\sum_{n=1}^{N} X_{y,n,m} \tag{16}$$

The standard deviation within the coefficient can be obtained:

$$clt_m^{inner} = \frac{1}{Y}\sum_{y=1}^{Y} \delta_{y,m} \tag{17}$$

The differences between different types of feature data, where y and c indicate that the data belong to different classes:

$$f_m^{inner} = \frac{\max(clt_{y,m}^{inner})}{\min(clt_{c,m}^{inner})} \tag{18}$$

Then the data differences between features are calculated and the standard deviation between features is calculated first:

$$\tau_{y,m} = \sqrt{\frac{\sum_{i=1}^{M-1}(cd_{n,r,y,m} - d_{y,m})^2}{N(N-1)-1}} \tag{19}$$

In equation (19), $cd_{n,r,y,m} = \left|X_{y,n,m} - X_{y,r,m}\right|$,

$$d_{y,m} = \frac{1}{N(N-1)}\sum_{n,r=1}^{N} cd_{n,r,y,m}.$$

Calculate the average value of standard deviation between features:

$$clt_m^{outer} = \frac{\sum_{y,c=1}^{Y}(\tau_{y,m} - \tau_{c,m})^2}{Y(Y-1)} \tag{20}$$

The differences between the features are as follows:

$$f_m^{outer} = \frac{\max(clt_{y,m}^{outer})}{\min(clt_{c,m}^{outer})} \tag{21}$$

The distance weight coefficient of each coefficient can be calculated as:

$$\eta_m = \frac{1}{\frac{f_m^{inner}}{\max(f_m^{inner})} + \frac{f_m^{outer}}{\max(f_m^{outer})}} \cdot \frac{clt_m^{outer}}{clt_m^{inner}} \tag{22}$$

3) Building XGboost model.

The prediction results of single sample $X_{m,y}$ are as follows:

$$\widehat{y}_x = \phi(X_{m,y}) = \sum_{m=1}^{M} \eta_m \cdot f_m(X_{m,y}) \tag{23}$$

$\widehat{y}_x$ is the prediction result of decision tree, $X_{m,y}$ It represents a sample with m feature classes as y class, and f_m is the m -th tree model. Through XGboost training, each tree can obtain the corresponding weight value w and tree structure parameter q according to the feature learning. In addition to the weight of the tree model obtained through training, this method adds the weight value η_m of the feature to modify the result of each tree model, so as to improve the accuracy of the model degree.

Objective function of XGboost decision tree:

$$L_m(\phi) = \sum_i l(y_i, y_i) + \sum_m \Omega(f_m) \tag{24}$$

where l is the loss function of the model tree, which is used to reduce the error between the predicted value and the real value and form the basic tree model structure. Ω is the regular term of the model tree, which is used to control the complexity of the tree model, so that the learner can avoid over fitting as much as possible.

Solve the objective function. Firstly, the loss function is expanded to the second order by Taylor expansion:

$$\begin{aligned} Obj = \sum_{i=1}^{n} (l(y_i, y_i^{(t-1)}) + g_i f_t(x_i) \\ + \frac{1}{2} h_i f_t^2(x_i)) + \Omega(f_t) + const \end{aligned} \tag{25}$$

Define XGboost tree structure:

$$f_t(x) = \omega_{q(x)}, \omega \in \mathbb{R}^T, q : \mathbb{R}^d \to \{1, 2, \ldots, T\} \tag{26}$$

Define the complexity of a tree:

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \tag{27}$$

The final objective function is as follows:

$$Obj = \sum_{j=1}^{T} [G_j \omega_j + \frac{1}{2}(H_j + \lambda)\omega_j^2] + \gamma T \tag{28}$$

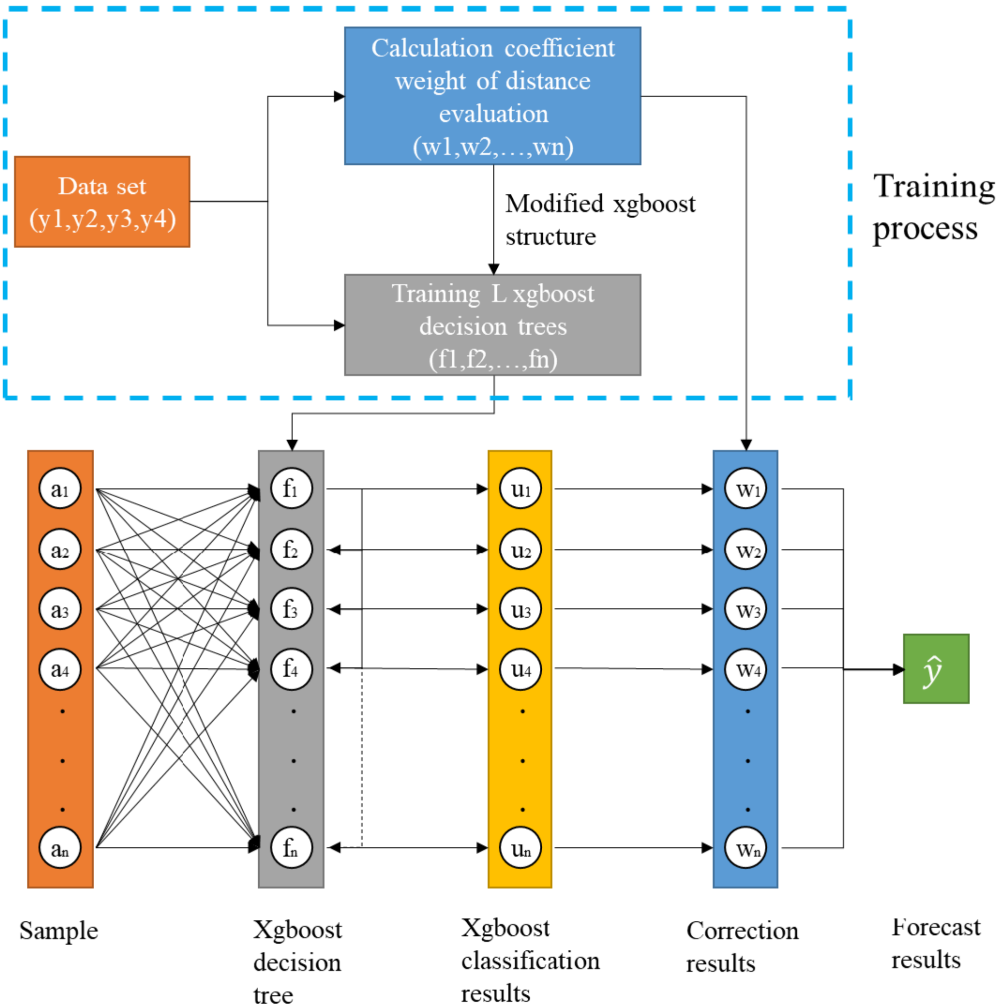


FIGURE 4. Weighted extreme gradient boosting method structure.

The optimal partition points are as follows:

$$\omega_j^* = -\frac{G_j}{H_j + \lambda}, Obj = -\frac{1}{2}\sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (29)$$

The basis of leaf node division is as follows:

$$Gain = \frac{1}{2}\left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda}\right] - \gamma \quad (30)$$

III. EXPERIMENTAL ANALYSIS

A. CASE 1

The open data set of Case Western Reserve University bearing center is taken as the sample for feature extraction [29]. The data used is the vibration signal collected by the driving end of rolling bearing under different fault conditions. The test bench for collecting data is shown in the figure 5, and the data samples are shown in the table 1. In addition, in order to verify the anti noise performance of the proposed method, noise is added to the data sets, and the SNR of noise is 10dB.

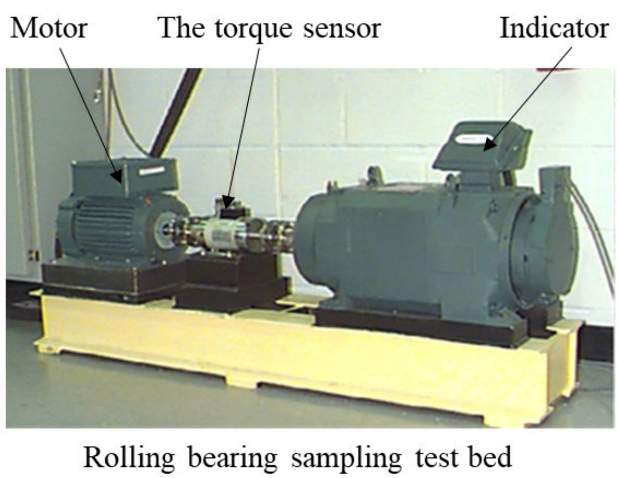


FIGURE 5. Test bed for data collection.

1) FEATURE EXTRACTION

Comparing MFCC and AFCC under the same parameters, the number of filters is 24, and the extracted discrete cosine

TABLE 1. Data sets used.

Data category	Sampling frequency	Speed /rpm	Samples number
No fault			50
0.007 inch inner ring fault			25
0.014 inch inner ring fault			25
0.007 inch rolling element fault	12K	1750	25
0.014 inch rolling element fault			25
0.007 inch outer ring fault			25
0.014 inch outer ring fault			25

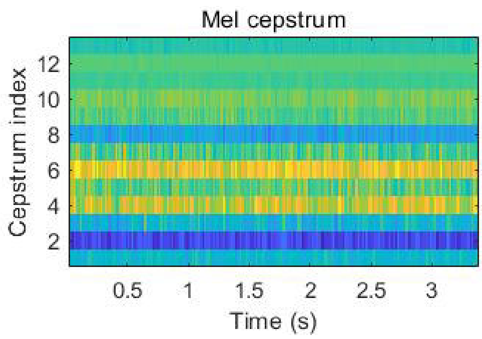


FIGURE 6. MFCC fault free feature.

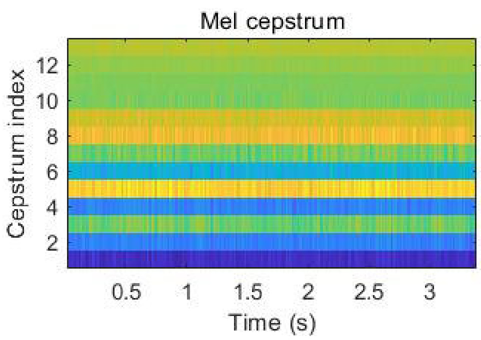


FIGURE 7. Fault characteristics of MFCC inner ring.

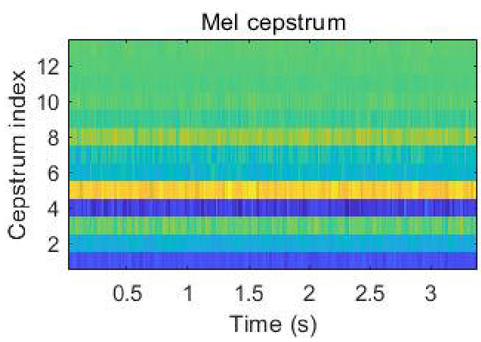


FIGURE 8. Fault characteristics of MFCC rolling element.

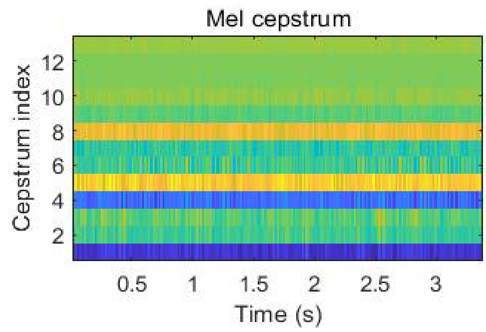


FIGURE 9. Fault characteristics of MFCC outer ring.

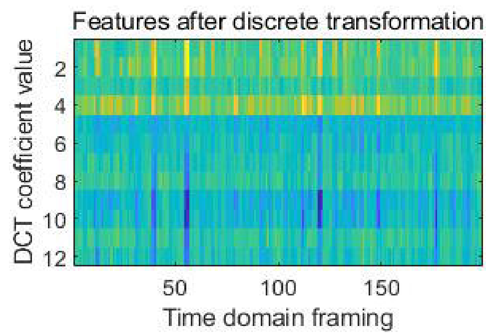


FIGURE 10. AFCC fault free feature.

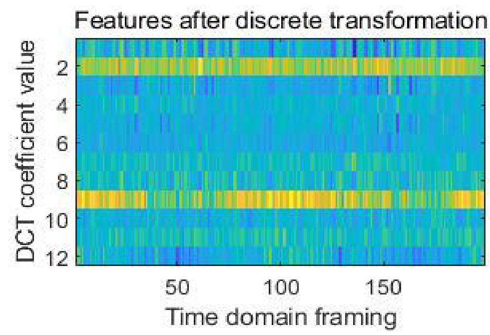


FIGURE 11. Fault characteristics of AFCC inner ring.

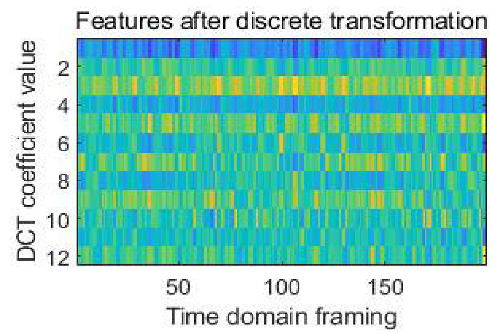


FIGURE 12. Fault characteristics of AFCC rolling element.

coefficient is 12. The other conditions are the same. The comparison of results is shown in Figure 6-13.

By comparing the results of feature extraction, it is found that MFCC method in feature extraction, because of the unreasonable distribution of Mel scale filter in the spectrum, can not correctly obtain the information of features, resulting in the loss of feature information; AFCC can extract the corresponding features, and get better feature results, which provides favorable conditions for feature classification.

2) FEATURE CLASSIFICATION

The improved XGboost model is used to judge the classification accuracy of AFCC feature extraction. The above

TABLE 2. Prediction results and related parameters of classification model.

Methods	Model parameters				Score
	Maximum depth	Branch basis	Iterations	Learning rate	
DT	5	Gini	420	*	0.90
GBDT	5	Cart	100	0.5	0.97
XGboost	5	Gbtree	160	0.4	0.95
WXGB	5	Gbtree	160	0.4	0.98

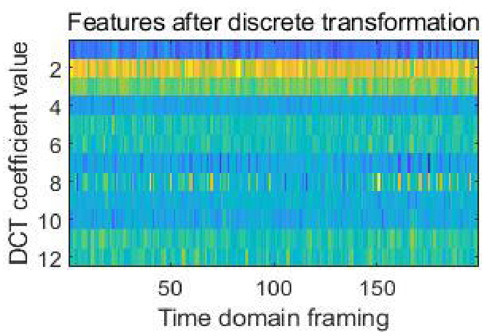


FIGURE 13. Fault characteristics of AFCC outer ring.

200 groups of samples are used as the classification training set, and 40 groups of samples are randomly selected as the test set to determine the classification accuracy. The corresponding accuracy of several classification methods is given as follows:

3) ANALYSIS OF EXPERIMENTAL RESULTS

a. The general decision tree classifies the data directly through the differences between the data. The problem of this classification is that the model structure formed is only suitable for the classification of the current data set, and the fault tolerance rate is low. For the classification set with more features, it may ignore the influence of some features on the results in use. This kind of tree model can be found by calling the feature weight. In the program only a part of the features are used as the classification standard, while the XGboost algorithm constructs the tree model for each type of features, and comprehensively considers the influence of each feature on the results. Therefore, the XGboost algorithm has higher classification accuracy than the general decision tree.

b. GBDT is a gradient lifting decision tree. Compared with GBDT, XGboost has the advantage of adopting the second derivative in the loss function and adding the regular term. The second derivative of the loss function helps the algorithm to get a more accurate optimal tree structure. The regular term is used to control the complexity of the number, so as to prevent the structure from being complicated due to the strong learning ability of the model.

c. Using XGboost to train the dataset, we can get a better classification tree model compared with the previous several schemes. In the training, we consider the influence of each feature comprehensively, and finally reduce the result of each tree to weighted sum, so as to get a suitable classification model.

TABLE 3. Sample type used.

Fault type	Sampling
NF	A.normal;B.normal
SF	A.normal;B.spalling
CF	A.crackle;B.normal
MF	A.crackle;B.spalling

d. Through the improved distance evaluation algorithm, the feature importance of the data set is affected by the weight when the XGboost tree model is output, and the influence weight of each feature in the model is further modified, so as to ensure the relevance between the XGboost model and the data set, which can be used to modify the prediction results of the model and improve the accuracy of the model.

4) THE XGBOOST PREDICTION PROCESS IS SHOWN IN FIGURE 14

B. CASE 2

In order to verify the practical application effect of the proposed method, a four types rolling bearing fault is set to verify the method.

The selected test bench is shown in Figure 15. Two rolling bearings can be installed on the test bench at the same time, which can be used to set different control experimental groups. As shown in Figure 16, rolling bearing peeling fault and rolling bearing crack fault are set artificially. In order to set four types of faults, one end peeling fault, one end crack fault, compound fault with two kinds of faults and two types of faults are set respectively and a case of no fault, the experimental group is set as shown in table 3.

1) Feature extraction

The results of feature extraction for four types of data are as follows:

2) Feature classification

In order to verify the performance of the proposed WXGB method, general decision tree (DT), random forest (RF), gradient increasing decision tree (GBDT) and limit gradient increasing decision tree (XGboost) are set respectively to compare the classification results. The comparison method is decision tree method. This kind of method is easy to obtain the training model, and the more data in the training set, the more classification model can be obtained It has better performance. In this paper, each method uses 200 groups of

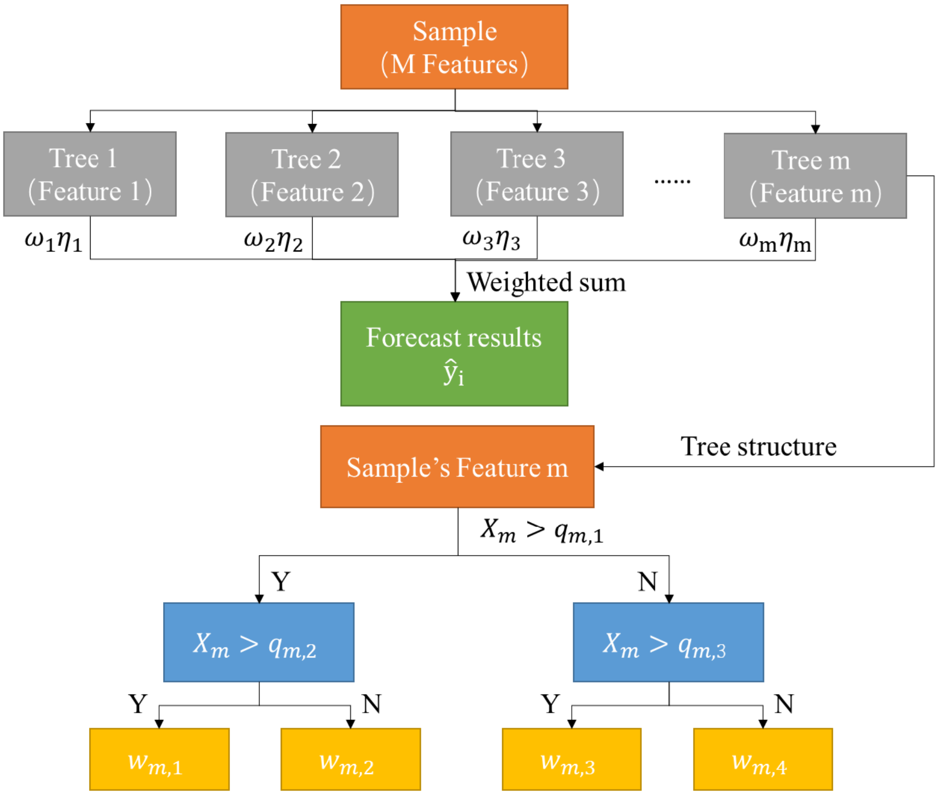
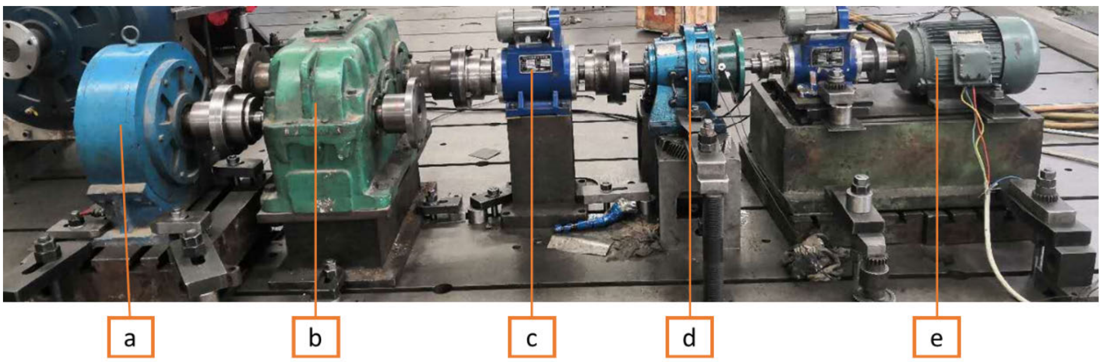


FIGURE 14. XGboost tree model.



a.Magnetic powder loader; b.Gear speed increaser; c.Torque speed sensor;
d. Three way acceleration sensor; e. Three phase asynchronous motor

FIGURE 15. Data acquisition experimental platform.

a.NJ405 Bearing spalling fault b.NJ210 Bearing crack fault

FIGURE 16. Bearing fault status.

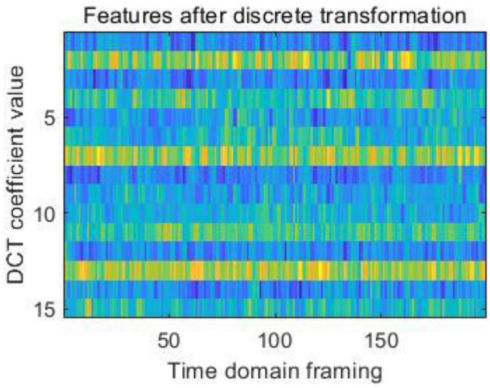


FIGURE 17. Spalling fault characteristics.

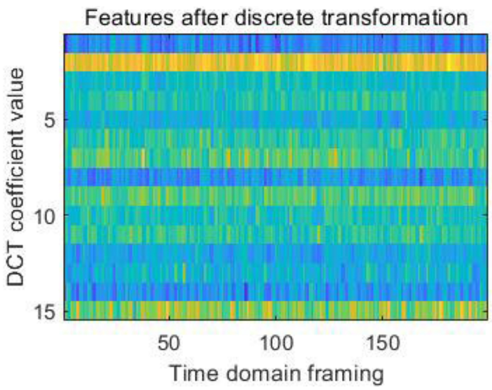


FIGURE 18. Fault characteristics of WEDM.

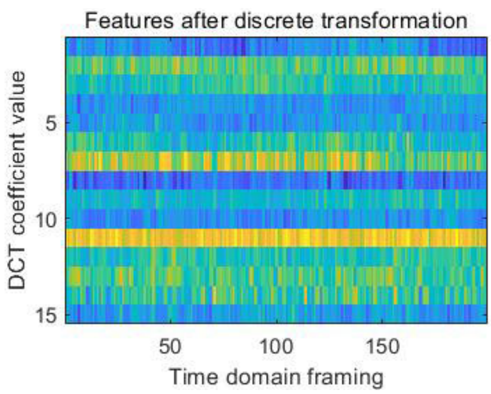


FIGURE 19. Composite fault characteristics.

samples as training set, the number of test sets is set to 40, and four groups of different combinations of test sets are used to test the accuracy of the obtained model. In order to facilitate the observation results, the average value of classification results is calculated, and the classification results are shown in table 4.

According to the classification results of table four, the classification accuracy of the improved method for the same data set is about ninety-seven percent. Compared with the same type of decision tree classification method, the accuracy of the method has been improved to a certain extent, which verifies the advanced nature of the new method proposed in this paper. Compared with the previous group of cases,

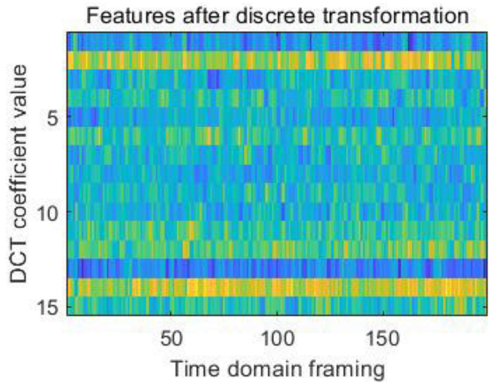


FIGURE 20. Fault free feature.

TABLE 4. Comparison of classification results of four types of faults(%).

Methods	Score1	Score2	Score3	Score4	Average
DT	91.67	85.00	95.00	90.00	90.42
RF	92.50	90.00	92.50	95.00	92.50
GBDT	90.62	92.64	90.71	94.55	92.13
XGboost	95.88	93.76	96.44	96.96	95.76
WXGB	96.64	97.15	97.14	97.91	97.21

the classification accuracy is relatively low, the reason may be the adoption of the method. In addition, composite fault is used as one of the class samples in this group of data. The existence of composite fault data may also be one of the reasons to reduce the accuracy of classification results.

IV. CONCLUSION

Through the comparison between AFCC feature extraction results and MFCC feature extraction results, compared with MFCC feature, AFCC feature can better reflect the feature differences between different types of data, using AFCC feature for rolling bearing fault diagnosis will obtain better classification results. Compared with different classification methods, WXGB method shows better classification accuracy and better classification performance, which verifies the excellent performance of the method.

The AFCC method can effectively extract the time and frequency domain features of rolling bearing at constant speed, and the trained WXGB can also complete the fault classification of the extracted features, and its performance is better than that of XGboost. Through the classification of two types of four different types of fault data, it is found that the two parameters of AFCC method, the number of filters and the order of discrete cosine transform, easily affect the result of feature extraction when extracting features from different data sets. The number of filters affects the accuracy of feature extraction. For vibration signals with complex spectrum components, the number of filters should be increased, and the order obtained by discrete cosine transform is used to express the distribution of features. If the distance between features is close, the order can be increased to show the difference between features. Therefore, in the future research, we can improve the acquisition of these two parameters, so as to achieve better feature extraction effect.

REFERENCES

[1] A. Ragab, S. Yacout, M.-S. Ouali, and H. Osman, "Prognostics of multiple failure modes in rotating machinery using a pattern-based classifier and cumulative incidence functions," *J. Intell. Manuf.*, vol. 30, no. 1, pp. 255–274, Jan. 2019.

[2] J. Yang, Y. Chang, T. Gao, and J. Wang, "Failure prediction of the rotating machinery based on CEEMDAN-ApEn feature and AR-UKF model," *Appl. Sci.*, vol. 10, no. 6, p. 2056, Mar. 2020.

[3] Z. J. Wang and N. N. Yang, "A new fault diagnosis method based on adaptive spectrum mode extraction," *Struct. Health Monit.*, 2021, doi: 10.1177/1475921720986945.

[4] H. Yu, Y. Ran, G. Zhang, X. Li, and B. Li, "A time-varying comprehensive dynamic model for the rotor system with multiple bearing faults," *J. Sound Vib.*, vol. 488, Dec. 2020, Art. no. 115650.

[5] X. Bi, S. Cao, and D. Zhang, "A variety of engine faults detection based on optimized variational mode decomposition-robust independent component analysis and fuzzy C-mean clustering," *IEEE Access*, vol. 7, pp. 27756–27768, 2019.

[6] Z. Chen, M. Wu, R. Zhao, F. Guretno, R. Yan, and X. Li, "Machine remaining useful life prediction via an attention-based deep learning approach," *IEEE Trans. Ind. Electron.*, vol. 68, no. 3, pp. 2521–2531, Mar. 2021.

[7] K. Zhang, Y. Dong, and A. Ball, "Feature selection by merging sequential bidirectional search into relevance vector machine in condition monitoring," *Chin. J. Mech. Eng.*, vol. 28, no. 6, pp. 1248–1253, Nov. 2015.

[8] S. Fatima, B. Guduri, A. R. Mohanty, and V. N. A. Naikan, "Transducer invariant multi-class fault classification in a rotor-bearing system using support vector machines," *Measurement*, vol. 58, pp. 363–374, Dec. 2014.

[9] Z. Wang, J. Zhou, Y. Lei, and W. Du, "Bearing fault diagnosis method based on adaptive maximum cyclostationarity blind deconvolution," *Mech. Syst. Signal Process.*, 2021, doi: 10.1016/j.ymssp.2021.108018.

[10] M. Deng, T. Meng, J. Cao, S. Wang, J. Zhang, and H. Fan, "Heart sound classification based on improved MFCC features and convolutional recurrent neural networks," *Neural Netw.*, vol. 130, pp. 22–32, Oct. 2020.

[11] M. S. Fahad, A. Deepak, G. Pradhan, and J. Yadav, "DNN-HMM-based speaker-adaptive emotion recognition using MFCC and epoch-based features," *Circuits, Syst., Signal Process.*, vol. 40, no. 1, pp. 466–489, Jul. 2020.

[12] C.-S. Alex Gong, C.-H. Simon Su, and K.-H. Tseng, "Implementation of machine learning for fault classification on vehicle power transmission system," *IEEE Sensors J.*, vol. 20, no. 24, pp. 15163–15176, Dec. 2020.

[13] Q. Jiang, F. Chang, and B. Sheng, "Bearing fault classification based on convolutional neural network in noise environment," *IEEE Access*, vol. 7, pp. 69795–69807, 2019.

[14] B. Erol, M. G. Amin, and S. Z. Gurbuz, "Automatic data-driven frequency-warped cepstral feature design for micro-Doppler classification," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 4, pp. 1724–1738, Aug. 2018.

[15] H. Liu, W. Song, Y. Niu, and E. Zio, "A generalized cauchy method for remaining useful life prediction of wind turbine gearboxes," *Mech. Syst. Signal Process.*, vol. 153, May 2021, Art. no. 107471.

[16] K. Wang, N. An, B. Nan Li, Y. Zhang, and L. Li, "Speech emotion recognition using Fourier parameters," *IEEE Trans. Affect. Comput.*, vol. 6, no. 1, pp. 69–75, Jan. 2015.

[17] A. Guerid and A. Houacine, "Recognition of isolated digits using DNN-HMM and harmonic noise model," *IET Signal Process.*, vol. 13, no. 2, pp. 207–214, Apr. 2019.

[18] K. Paliwal, B. Shannon, J. Lyons, and K. Wojcicki, "Speech-signal-based frequency warping," *IEEE Signal Process. Lett.*, vol. 16, no. 4, pp. 319–322, Apr. 2009.

[19] A. Bakhshi, S. Chalup, A. Harimi, and S. M. Mirhassani, "Recognition of emotion from speech using evolutionary cepstral coefficients," *Multimedia Tools Appl.*, vol. 79, nos. 47–48, pp. 35739–35759, Dec. 2020.

[20] M. S. Ahmad, J. Mir, M. O. Ullah, M. L. U. R. Shahid, and M. A. Syed, "An efficient heart murmur recognition and cardiovascular disorders classification system," *Australas. Phys. Eng. Sci. Med.*, vol. 42, no. 3, pp. 733–743, Sep. 2019.

[21] H. Liu, W. Song, and E. Zio, "Generalized cauchy difference iterative forecasting model for wind speed based on fractal time series," *Nonlinear Dyn.*, vol. 103, no. 1, pp. 759–773, Jan. 2021.

[22] Z. K. Abdul, A. K. Al-Talabani, and D. O. Ramadan, "A hybrid temporal feature for gear fault diagnosis using the long short term memory," *IEEE Sensors J.*, vol. 20, no. 23, pp. 14444–14452, Dec. 2020.

[23] T. H. Chen *et al.*, "T4SE-XGB: Interpretable sequence-based prediction of type IV secreted effectors using extreme gradient boosting algorithm," *Front Microbiol.*, vol. 11, 2020, doi: 10.3389/fmicb.2020.580382.

[24] Z. Wang, W. Zhao, W. Du, N. Li, and J. Wang, "Data-driven fault diagnosis method based on the conversion of erosion operation signals into images and convolutional neural network," *Process Saf. Environ. Protection*, 2021, doi: 10.1016/j.psep.2021.03.016.

[25] P. Harar, Z. Galaz, J. B. Alonso-Hernandez, J. Mekyska, R. Burget, and Z. Smekal, "Towards robust voice pathology detection: Investigation of supervised deep learning, gradient boosting, and anomaly detection approaches across four databases," *Neural Comput. Appl.*, vol. 32, no. 20, pp. 15747–15757, Oct. 2020.

[26] J. Zhang, Y. Sun, L. Shang, Q. Feng, L. Gong, and K. Wu, "A unified intelligent model for estimating the (gas + n-alkane) interfacial tension based on the extreme gradient boosting (XGBoost) trees," *Fuel*, vol. 282, Dec. 2020, Art. no. 118783.

[27] Y. Lei, Z. He, Y. Zi, and X. Chen, "New clustering algorithm-based fault diagnosis using compensation distance evaluation technique," *Mech. Syst. Signal Process.*, vol. 22, no. 2, pp. 419–435, Feb. 2008.

[28] Y. Tan, L. Guo, H. Gao, and L. Zhang, "Deep coupled joint distribution adaptation network: A method for intelligent fault diagnosis between artificial and real damages," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021, doi: 10.1109/TIM.2020.3043510.

[29] *The Data is From the Bearing Data Center of Case Western Reserve University, USA*. Accessed: Dec. 2020. [Online]. Available: https://csegroups.case.edu/

MINGSI QI was born in Henan, China, in 1979. He received the Ph.D. degree from the North University of China, Taiyuan, China. He is currently an Associate Professor with the North University of China. He is also the Founder of the intangible cultural heritage Qi Ziwen Needle. His research interests include buffer protection technology, intelligent medical apparatus, and intelligent medical robot. He has published more than 12 articles in this field. He is a member of the Chinese Society of Packaging Engineering.

RUI ZHOU was born in Yuncheng, Shanxi, in 1997. He received the bachelor's degree in mechanical engineering from the North University of China, in 2020, where he is currently pursuing the degree with the School of Mechanical Engineering. His main research interest includes mechanical vibration signal processing.

QIANQIAN ZHANG was born in 1987. She received the Ph.D. degree from the Taiyuan University of Technology, Taiyuan, China. She is currently a Lecturer with Shanxi University. Her research interests include optimal design of cutting head, fault diagnosis, and computational mechanics.

YONGSHENG YANG was born in Neimenggu, China, in 1996. He is currently pursuing the master's degree in mechanical engineering with the North University of China. His research interests include buffer protection technology, intelligent medical apparatus, and intelligent medical robot.