

Received June 29, 2021, accepted July 9, 2021, date of publication July 12, 2021, date of current version July 20, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3096723

# A Novel Rolling Bearing Fault Diagnosis Method Based on Adaptive Feature Selection and Clustering

JINGBAO HOU<sup>1,2</sup>, (Member, IEEE), YUNXIN WU<sup>1,3</sup>, ABDULRAHAMAN SHUAIBU AHMAD<sup>3</sup>, HAI GONG<sup>1,3</sup>, AND LEI LIU<sup>1</sup>

<sup>1</sup>Light Alloy Research Institute, Central South University, Changsha 410083, China

<sup>2</sup>Changsha Institute of Mining Research Company Ltd., Changsha 410012, China

<sup>3</sup>State Key Laboratory of High Performance Complex Manufacturing, Central South University, Changsha 410083, China

Corresponding author: Yunxin Wu (wuyunxin@csu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51327902 and Grant 51975596.

**ABSTRACT** Rolling bearing is an important part of mechanical equipment. Timely detection of rolling bearing fault is one of the important factors to ensure the safe operation of equipment. In order to diagnose rolling bearing fault accurately, a novel rolling bearing fault diagnosis method based on adaptive feature selection and clustering is proposed. Firstly, the vibration signal obtained from rolling bearing is decomposed by ensemble empirical mode decomposition (EEMD) to extract as much important information as possible. Feature extraction is performed for each intrinsic mode function (IMF) component and the original signal, and finally 240 features are obtained. And the Chi-square Test algorithm, Variance-Relief-F algorithm, and hierarchical clustering algorithm are used to filter all the features in layers to obtain the optimal features. Then the optimal features are input into fuzzy c-means (FCM) clustering to complete fault diagnosis. After the fault diagnosis analysis of four groups of vibration signal data, it is found that whether the characteristic number parameters are set based on engineering experience or adaptive feature selection, good fault diagnosis results are obtained. Furthermore, through comparative experiments, the fault diagnosis effect of the method based on adaptive parameter setting is better. The results indicate that the proposed adaptive parameter fault diagnosis method is feasible and effective for rolling bearing fault diagnosis.

**INDEX TERMS** Fault diagnosis, rolling bearing, adaptive feature selection, Relief-F algorithm, clustering.

## I. INTRODUCTION

Rolling bearing is one of the most widely used mechanical components in all kinds of rotational machinery. Its running state often directly affects the performance of the whole machine and plays an essential role in mechanical equipment. But the working condition of rolling bearing requires prompt attention as it is easy to produce local damage [1], [2]. Once the fault occurs, it will adversely affect the direct operation and safety of the whole equipment [3]. The traditional bearing fault diagnosis is mainly performed based on two approaches: one is based on model analysis, while the other is based on signal processing [4]. However, these two kinds of methods often need the engineers with high proficiency and experience, as such other fault diagnosis approaches

may be difficult. Hence, with the development of artificial intelligence technology, traditional fault diagnosis methods based on probabilistic models have made new advances [5], [6]. At the same time, more and more data-driven intelligent fault diagnosis methods are flourishing [7].

Cluster analysis is one of the important research areas for pattern recognition [8], [9]. In Cluster analysis, set of samples can be divided into several classes, and the samples in each class are similar. The difference between cluster analysis and other pattern recognition technologies is that the required classification of clustering is unknown, and no training samples are needed before cluster analysis [10]. Cluster analysis is classified on the basis of similarity. The samples in the same category have greater similarity, and the samples between different classes have greater dissimilarity.

The widely used clustering methods in pattern recognition are the spectral clustering, K-means clustering, fuzzy

The associate editor coordinating the review of this manuscript and approving it for publication was Baoping Cai<sup>1</sup>.

center means(FCM) clustering, and Gustafaon-Kessel (GK) clustering algorithms. M. Zarinbal *et al.* [11] proposed relative entropy Fuzzy C-means(FCM) clustering method, and Serir *et al.* [12] proposed online clustering method of evidence evolution GK clustering.

West *et al.* [13], scholars in the University of Strathclyde, United Kingdom, used hierarchical clustering algorithm to intelligently identify the fault of nuclear power plant equipment, and the outcome is remarkable. Santis *et al.* [14], an Italian scholar, established a smart grid modeling and fault identification system with the aid of cluster analysis, and used fuzzy set decision rules to comprehensively analyze results of the identified faults. Finally, the system successfully served the power grid in Rome, Italy.

Each clustering approach has its own advantages and disadvantages [15], [16]. In the process of fault diagnosis, it is often difficult to obtain ideal results if only a single clustering method is used to classify different fault categories. Especially when the feature vector dimension is relatively large, this problem becomes much significant [17]. Therefore, if we select a clustering method based on the features of input characteristics to filter the redundant features, and then select another clustering method to classify the fault, better fault diagnosis results can be achieved.

With the rapid development of signal processing technology, more and more features are used in condition monitoring and fault diagnosis [18], [19]. Therefore, in order to improve the diagnosis accuracy and reduce the computational complexity and duration, some sensitive features should be selected from the original feature set for pattern recognition [20], [21]. At present, many researches have been carried out in feature selection [22], such as variance in statistics [23], distance evaluation technology [24], decision tree attribute weighted filtering [25] and Relief-F attribute estimation method [26]. Each feature selection technique has a different focus. When performing feature selection, by applying only one feature selection technique, some important features may be ignored. Therefore, when using these technologies, it is necessary to add certain prior knowledge to aid in the calculation of attribute weight values, and also evaluate different attributes. In order to increase the degree of intelligence of the model, an optimized feature selection process is required.

To efficiently and accurately diagnosed bearing faults, a bearing intelligent fault diagnosis method based on a new feature selection technique and clustering is proposed in this work. What this paper is to solve is to find a feature selection technique which can be used in engineering practice to improve the efficiency and accuracy of fault diagnosis. The advantage of the method proposed in this paper is that the method can adaptively select the type and quantity of features. The number of selected features will be less than or equal to this threshold. The use of setting thresholds instead of directly setting the number of features to be selected can well avoid some non-sensitive features from being selected. Firstly, the information on the status of the bearing needs to be obtained from the mechanical equipment. This method

uses the vibration generated during the operation of the machine to diagnose faults, and performs EEMD decomposition of the original signal. Secondly, the fault characteristics of the equipment are reflected in different degrees in the time domain, frequency domain, and time-frequency domain. Therefore, in this approach, the feature extraction will be performed simultaneously in the time domain, frequency domain. Furthermore, some relevant features are to be selected from the original feature set for pattern recognition. The proposed feature selection approach is used to select the primary and sensitive features. Then, the hierarchical clustering was used to perform redundant removal of sensitive features to obtain optimal features. Finally, the optimal features are implemented in the Fuzzy C-means(FCM) clustering algorithm for fault clustering, and the fault diagnosis is completed.

The remaining sections of this paper are organized as follows. The basic theory of EEMD and clustering is briefly reviewed in Section II. The proposed approach applied for fault diagnosis in this paper is extensively described in Section III. The experiments carried out and the data obtained for evaluation of the proposed method are presented in section IV. The effectiveness of the proposed approach is discussed in Section V. Finally, the summarized conclusions are presented in section VI.

## II. METHODS

### A. ENSEMBLE EMPIRICAL MODE DECOMPOSITION

When the data consists of white noise, and its scale is evenly distributed over the entire time or time and frequency, the Empirical Mode Decomposition (EMD) can be considered as a binary filter bank [27]. When the data is not purely white noise, some scales will be lost, which will cause mode aliasing. However, all data in reality are fused with signal and noise. This situation is common, so the mode aliasing phenomenon of Empirical Mode Decomposition (EMD) is inevitable.

Ensemble Empirical Mode Decomposition (EEMD) defines the overall test means including signal and limited amplitude white noise as real intrinsic mode function (IMF) components [28]. Its essence is a multiple Empirical Mode Decomposition (EMD) decomposition with superimposed Gaussian white noise. It makes use of the statistical characteristics of Gaussian white noise with uniform frequency distribution to make the signal with noise continuous on different scales. Although the additional noise may lead to a smaller signal-to-noise ratio, it can provide a uniform correlation scale distribution to overcome mode aliasing.

The algorithm process of EEMD is as follows [29]:

*Step 1:* Initialize the total number of times to perform EMD decomposition  $M$  and the amplitude coefficient  $k$ . Let  $m = 1$ .

*Step 2:* Perform the  $m - th$  EMD decomposition.

a. Add a random Gaussian white noise sequence  $n_m(t)$  to the input signal  $x(t)$  to get the noisy signal to be processed.

b. The EMD is used to Decompose the signal  $x_m(t)$  to get

$$\text{IMF } c_{j,m} (j = 1, 2, \dots, I) \quad (1)$$

IMF  $c_{j,m}$  denotes the  $m$ -th intrinsic mode function (IMF) obtained from the  $j$ -th decomposition test;

c. If  $m < M$ , return to Step 2, let  $m = m + 1$ .

Step 3: Calculate the mean for each intrinsic mode function(IMF) of  $M$  decomposition trials.

$$\bar{c}_j = \sum_{m=1}^M c_{j,m}/M \quad (j = 1, 2, \dots, I; m = 1, 2, \dots, M) \quad (2)$$

Step 4: Output  $\bar{c}_j$  as the  $j$ -th intrinsic mode function (IMF) obtained by EEMD decomposition.

### B. FUZZY C-MEANS CLUSTERING

Generally, in the Fuzzy C-means(FCM) clustering algorithm, the degree of similarity between sample points is defined as membership. Suppose that the data set is  $x_1, x_2, \dots, x_n$ , which is divided into  $i$  classes. And the cluster center point is  $c_i$ . Let the membership degree between any sample point  $j$  and a class center point  $c_i$  be  $u_{ij}$ . Then the objective function and constraint condition can be expressed as follows:

$$J = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|x_j - c_i\|^2 \quad (3)$$

$$\sum_{i=1}^c u_{ij} = 1, j = 1, 2, \dots, n \quad (4)$$

where  $m$  is the exponential weight factor of the membership degree  $u_{ij}$ , and  $n$  is the total number of samples in the dataset. Using Lagrange multiplier method and other mathematical operation methods, we can get the value of  $u_{ij}$  and  $c_i$ , which are shown in equation (5) and (6), under the condition that the objective function is the smallest.

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left( \frac{\|x_j - c_i\|^2}{\|x_j - c_k\|^2} \right)^{\frac{2}{m-1}}} \quad (5)$$

$$c_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m} \quad (6)$$

The steps of Fuzzy C-means(FCM) algorithm are as follows: Firstly, randomly initialize  $u_{ij}$  satisfying the constraint condition, and then the value of  $c_i$  is calculated according to formula (6). Secondly, the obtained  $c_i$  is used as input, and the new  $u_{ij}$  value is calculated according to formula (5). At this time, the value of objective function  $J$  is calculated according to equation (3). Thirdly, according to the formula (6), (5) and (3), the values of  $c_i$ ,  $u_{ij}$  and  $J$  are calculated iteratively. According to this method, calculate circularly. When  $J$  reaches the minimum value, stop the calculation and the values of  $c_i$  and  $u_{ij}$  are output to complete the clustering [30]–[32].

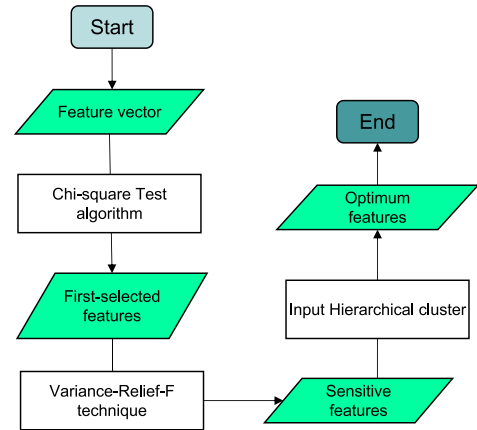


FIGURE 1. Feature selection flowchart.

### III. INTELLIGENT FAULT DIAGNOSIS

#### A. THE NOVEL ADAPTIVE FEATURE SELECTION TECHNIQUE

When performing bearing fault diagnosis, in order to improve the accuracy of fault diagnosis, we need to extract as many features as possible from the vibration signal that can reflect the running state of the bearing to form a high-dimensional feature vector. But this may also cause some problems. On the one hand, the high-dimensional feature vector is not conducive to further processing of the data. On the other hand, for a specific fault diagnosis process, among the many features, the importance of each feature to fault diagnosis is different. Such a high-dimensional feature vector may cause the accuracy of fault diagnosis to decrease due to the existence of non-important features. Therefore, we need to perform targeted screening of features according to their importance, which not only reduces the dimensionality of data, but also improves the quality of information data. To achieve this goal, this paper presents an adaptive feature selection method. This method is divided into three stages: Chi-square test, Variance-Relief-F feature selection, and removing redundant feature based on Hierarchical clustering. The feature selection flowchart is shown in Figure 1. The detailed steps are as follows:

Step 1: Select the first selected feature from the original features based on Chi-square test algorithm to achieve the initial cleaning of the data.

Step 2: Then, selected the sensitive features from the primary features according to the variance value and Relief-F value obtained by the Variance-Relief-F technique.

Step 3: Finally, input the sensitive features into hierarchical cluster, removing the redundant features among the sensitive features to obtain the optimum features.

#### 1) THE CHI-SQUARE TEST

The idea of Chi-square test is to judge the correct rate of the theoretical value by comparing the deviation of the observed value to the theoretical value, and to find the correlation of the feature to the category through statistical theory [33], [34].

The larger the Chi-square value, the higher the correlation of the feature to the category, that is, the higher the distinction of the feature to the category. Therefore, when performing feature classification, the feature with higher chi-square value should be selected first.

For a feature  $f$ , the formula for calculating the Chi-square value of the feature  $f$  to the category is as follows:

$$\text{chi}(f, c) = \sum_{i=1}^n \frac{(f_{obs} - f_{exp})^2}{f_{exp}} \quad (7)$$

where  $f_{obs}$  is the observed value of the feature and  $f_{exp}$  is the expected value of the feature.  $n$  is the number of samples.  $\text{chi}(f, c)$  indicates the correlation of features to categories.  $f_{exp}$  is the expected value not the mean value. In the subsequent experimental process we are using the chi-square test theory to get the chi-square value for feature selection and select the first-selected features, rather than using the expectation value to judge the features.

For multi class data, after calculating the Chi-square value of a feature to each category, the maximum value of the Chi-square value is obtained by the following formula.

$$\text{chi}(f) = \max \{ \text{chi}(f, c_i), (1 \leq i \leq p) \} \quad (8)$$

where  $p$  is the number of categories. In order to eliminate the mutual influence caused by the large difference between the two scores, we need to normalize  $\text{chi}(f)$ . The formula is as follows:

$$x^* = \frac{x - \min(X)}{\max(X) - \min(X)} \quad (9)$$

The adaptive algorithm select the first-selected features from the original features using a card-squared test algorithm to achieve the initial cleaning of the data.

## 2) THE VARIANCE-RELIEF-F TECHNIQUE

Relief-F feature selection algorithm selects features based on the correlation of features to categories [35]. The algorithm sets different weights for features according to the correlation of features to categories, and then remove the features whose weights are less than the threshold to realize feature selection. But Relief-F can only deal with the binary classification. In order to deal with multi-classification problems, the Relief-F algorithm are improved, eventually forming the Relief-F algorithm [36].

For a data set  $D$ , suppose the number of samplings is  $m$ , the number of nearest neighbor samples is  $k$ , and the threshold of feature weight is  $T$ . The Relief-F algorithm randomly takes a sample  $R$  from the data set each time, and then finds  $k$  neighboring samples from the sample set that belongs to the same type as  $R$ , and finds  $k$  neighboring samples from the sample set that does not belong to the same type as  $R$ . Then update the weight of each feature. Finally, it will calculate the feature weight of each feature. The significance of weight is to subtract the difference of the feature of the same category and add the difference of the feature of different categories. If the feature is related to classification, the value of the

feature of the same category should be similar, while the value of different categories should not be similar. Finally, according to the different ranking of the weights, a suitable feature subset is selected.

The Variance-Relief-F selection method first calculates the variance value of each first-selected feature in all the samples and then ranks all the features using the variance values in order from largest to smallest to obtain the sequence  $V(i)$ . At the same time, using the Relief-F algorithm, the weight value of each feature is calculated and the features are sorted from largest to smallest using the weight value to obtain the sequence  $R(i)$ . We need to set a threshold value to obtain the number of sensitive features, assuming that this threshold value is  $T$  features, we need to extract from the sequence  $V(i)$  the feature set  $A$  consisting of  $T$  features sorted in the previous  $T$ , and we also need to extract from the sequence  $R(i)$  the feature set  $B$  consisting of  $T$  features sorted in the previous  $T$ . If a feature is both in set  $A$  and in set  $B$ , then this feature will be selected for the sensitive features. The selected sensitive features form the set  $C$  of sensitive features.

## 3) THE HIERARCHICAL CLUSTERING

Generally, there are two types of hierarchical clustering methods. One is agglomerative hierarchical clustering, which adopts bottom-up strategy. First, each object is regarded as a cluster, and then these atomic clusters are merged into larger clusters until all the objects are in a cluster or some termination condition is satisfied. The other is split hierarchical clustering, which adopts the top-down strategy. In contrast to agglomerative hierarchical clustering, it first puts all objects in a cluster, and then gradually subdivides them into smaller clusters until some objects form a cluster by themselves, or the calculation results reaches some end condition, such as reaching a desired number of clusters, or the distance between two nearest clusters exceeding a set value. This paper adopts the first one. This method is described as follows:

Suppose that the  $N \times N$  sensitive feature matrix is  $D = [d(i, j)]$ . The clustering results are denoted as serial numbers  $0, 1, \dots, (n-1)$ , and  $L(m)$  is the level of the  $m$ -th clustering. The number of clusters is expressed by  $(m)$ , and the similarity coefficients of clusters  $(r)$  and  $(s)$  are expressed by  $d[(r), (s)]$ . The specific steps are as follows:

*Step 1:*  $L(0) = 0, m = 0$ ;

*Step 2:* From all current cluster pairs, find the two closest clusters  $(r)$  and  $(s)$  according to  $d[(r), (s)] = \min d[(i), (j)]$ ;

*Step 3:* Add 1 to the sequence number of the cluster, that is,  $m = m + 1$ , then merge the clusters  $(r)$  and  $(s)$ , and let  $L(m) = d[(r), (s)]$ ;

*Step 4:* Update the similarity matrix  $D$ . Delete the corresponding rows and columns of clusters  $(r)$  and  $(s)$ , and add the corresponding rows and columns of the newly generated clusters to the matrix. The similarity between the newly generated cluster  $(r, s)$  and the original cluster  $(k)$  in the similarity matrix is defined by the following formula:  $d[(k), (r, s)] = \min d[(k), (r)], d[(k), (s)]$ ;

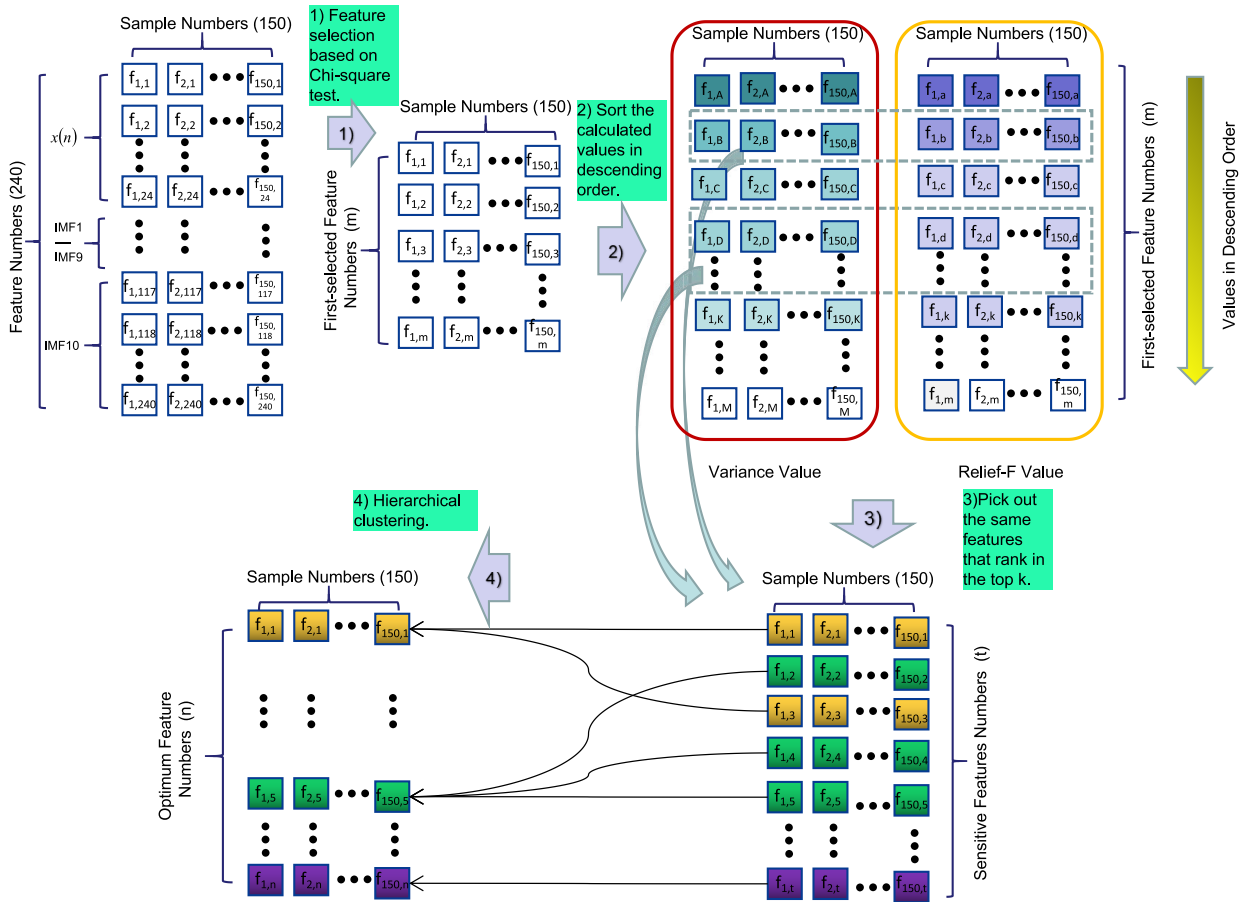


FIGURE 2. Detailed view of the adaptive feature selection process.

Step 5: Repeat Step 2 to 4 until all objects are clustered into one cluster.

The feature selection technique uses hierarchical clustering to extract redundant features from sensitive features to obtain the optimal features.

The decomposition diagram of the whole adaptive feature selection technology is shown in Figure 2. As can be seen from Figure 2, the whole feature selection process consists of four parts: Feature selection based on Chi-square test; Sort the calculated values in descending order; Pick out the same features that rank in the top  $k$ ; Hierarchical clustering. The original data  $[f_{1,1} \cdots f_{150,240}]$  has 150 samples, and each sample has 240 features. The first step is to select  $m$  First-selected features  $[f_{1,1} \cdots f_{150,m}]$  from these 240 features using the Chi-square test to complete the first feature screening. Then the features are ranked using variance and Relief-F analysis to select features that are within the threshold setting. For these features, Hierarchical clustering is used to remove redundant features to obtain  $n$  optimal features  $[f_{1,1} \cdots f_{150,n}]$ .

### B. THE PROPOSED INTELLIGENT FAULT DIAGNOSIS METHOD

During the operation of the equipment, if the equipment fails, the amplitude and frequency of the vibration signal from the equipment will change. Therefore, by analyzing

the time-domain waveform and frequency components in the vibration signal, finding out the information contained in the vibration signal can provide a powerful basis for the fault diagnosis of the equipment. The common features used to analyze vibration signals include time domain statistical features, frequency domain statistical features, and entropy features, etc. In this paper, 24 features are mainly used for fault diagnosis, including 16 time domain features and 5 frequency domain features, as shown in Table 1. In addition, three entropy value features are also used: permutation entropy, sample entropy and fuzzy entropy.

In order to dig out more information contained in the vibration signal, we usually do not directly extract these 24 eigenvalues from the original signal, but first perform further processing on the original signal. Before feature extraction, we first perform EEMD decomposition on the original signal and select the first 9 intrinsic mode function (IMF) components. Then extract the 24 features mentioned above from each original signal and its components. In this way, from a signal sample, we can obtain 240 features that can characterize the characteristics of the vibration signal, thus forming a feature vector. This feature vector will be used as the input of the adaptive feature selection technology. The process of feature extraction is shown in Figure 3.

TABLE 1. Statistical features used in this paper.

Serial Number	Feature Names	Calculating Formulas
1	Mean	$\bar{X} = \frac{1}{N} \sum_{n=1}^N x(n)$
2	Root Mean Square	$X_{rms} = \sqrt{\frac{1}{N} \sum_{n=1}^N x^2(n)}$
3	Square Root Amplitude	$X_r = \left( \frac{1}{N} \sum_{n=1}^N \sqrt{x(n)} \right)^2$
4	Average Amplitude	$ \bar{X}  = \frac{1}{N} \sum_{n=1}^N  x(n) $
5	Slope	$\alpha = \frac{1}{N} \sum_{n=1}^N x^3(n)$
6	Kurtosis	$\beta = \frac{1}{N} \sum_{n=1}^N x^4(n)$
7	Variance	$\sigma^2 = \frac{1}{N} \sum_{n=1}^N (x(n) - \bar{X})^2$
8	Maximum	$X_{max} = \max\{x(n)\}$
9	Minimum	$X_{min} = \min\{x(n)\}$
10	Peak to Peak Value	$X_k = X_{max} - X_{min}$
11	Shape Factor	$X_{sf} = \frac{X_{rms}}{ \bar{X} }$
12	Crest Factor	$X_{cf} = \frac{X_{max}}{X_{rms}}$
13	Impulse Factor	$X_{if} = \frac{X_{max}}{ \bar{X} }$
14	Clearance Factor	$X_{clf} = \frac{X_{max}}{X_r}$
15	Skewness	$X_s = \frac{\alpha}{(\sqrt{\sigma^2})^3}$
16	Kurtosis Factor	$X_{kf} = \frac{\beta}{\sigma^2}$
17	Center frequency	$F_{cf} = \frac{\sum_{k=1}^K f_k \cdot s(k)}{\sum_{k=1}^K s(k)}$
18	Mean Square Frequency	$F_{mfs} = \frac{\sum_{k=1}^K f_k \cdot s(k)}{\sum_{k=1}^K s(k)}$
19	Root Mean Square Frequency	$F_{rmfs} = \sqrt{\frac{\sum_{k=1}^K f_k \cdot s(k)}{\sum_{k=1}^K s(k)}}$
20	Deviation Frequency	$F_{td} = \frac{\sum_{k=1}^K (f_k - F_{cf})^2 \cdot s(k)}{\sum_{k=1}^K s(k)}$
21	Standard Deviation Frequency	$F_{std} = \sqrt{\frac{\sum_{k=1}^K (f_k - F_{cf})^2 \cdot s(k)}{\sum_{k=1}^K s(k)}}$

The symbols on the left of the equation of the calculation formula in the third column in the table represent the meaning of the corresponding feature name. For example, The seventeenth feature  $F_{cf}$  represents center frequency  $x(n)$  denotes a time-domain signal.  $N$  represents the number of sample points;  $s(k)$  represents the spectrum of signal.  $K$  is the number of spectrum lines;  $f_k$  represents the frequency of the  $k$ -th spectrum line.

Using the adaptive feature selection technique described above, the 240 features are screened to obtain the optimal features, which are used as the input of the Fuzzy C-means(FCM) clustering algorithm to perform fault diagnosis, and finally the final fault diagnosis results are obtained. The specific flowchart of the fault diagnosis method is shown in Figure 4. Rolling bearings working under different operating conditions present different fault characteristics. In order to obtain the optimal fault diagnosis, it is necessary to set the parameters required in the feature selection process for different operating conditions. Therefore, in Figure 4, the flow

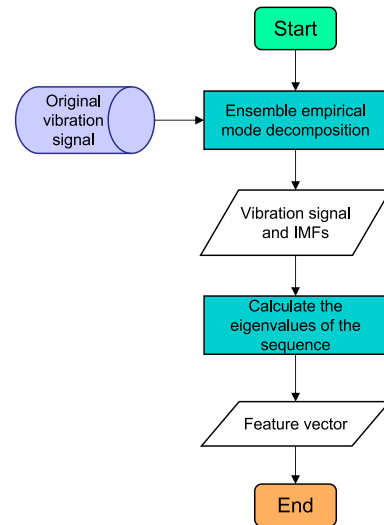


FIGURE 3. Feature extraction flowchart.

chart contains the content about the feature parameters selection, which will be elaborated in the next section.

IV. EXPERIMENTAL EVALUATIONS

The experimental data of rolling bearings used in this paper comes from the Bearing Data Center of Case Western Reserve University. The bearings used in the experiment were manufactured by Svenska Kullager Fabriken (SKF). The motor power is 2HP. Partial bearing failure is a single point failure caused by electrical discharge machining (EDM). The fault diameter is 0.1778mm, and the experimental process is shown in Figure 2. The failure types of vibration signal experimental data can be divided into three categories: outer ring failure (OF), spherical failure (B), and inner ring failure (IF). A group of normal signals (N) were set as experimental control. The sampling frequency of the vibration signal is 12khz.

Fifty samples were extracted from each signal type, and the length of each sample was set to 2048 samples. Thus a total of 200 samples were obtained from the four fault signals. The original vibration signals and their spectra are shown in Figure 5. To obtain finer fault information, the original vibration signals are decomposed by EEMD. A series of intrinsic mode function (IMF) components are obtained as shown in Figure 6. The decomposed intrinsic mode function (IMF) components are then subjected to feature extraction together with the original signal. From each component, 21 frequency and time domain features as shown in Table 1, and 3 entropy features are reextracted. In this way, a 240-dimensional feature vector is obtained for each sample, and the feature extraction of the signal is completed.

The next step is to select features. After the above calculation, we have got as many as 240 features in total. A large number of features will not only reduce computational efficiency, but also cause dimensionality disasters. So next, we need to perform feature selection on the 240 features obtained, and select the best feature. In the

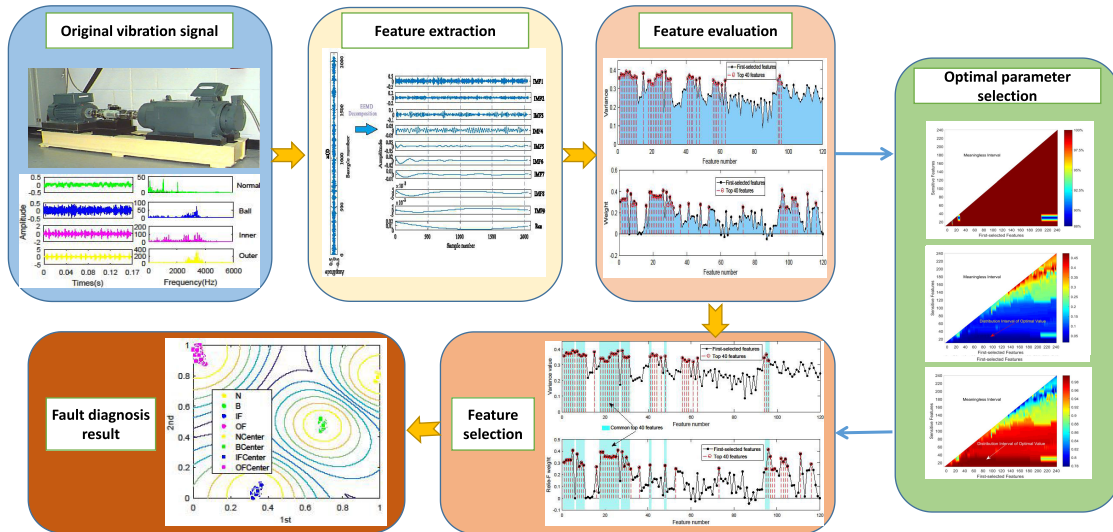


FIGURE 4. The specific flowchart of the fault diagnosis method.

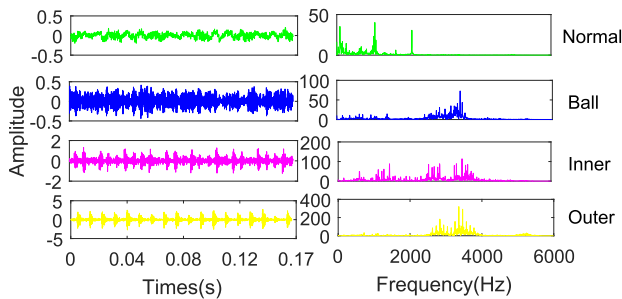


FIGURE 5. The original vibration signals and their spectra.

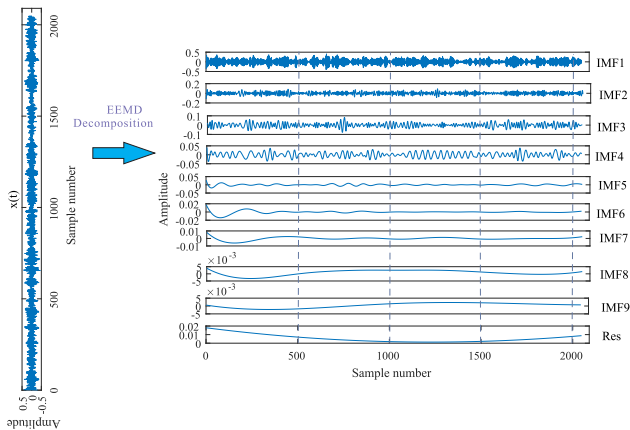


FIGURE 6. The intrinsic mode function components (IMFs) of the vibration signals.

feature selection step, we first use the selected algorithm to select the first-selected features, and then use the Variance-Relief-F technology to select the sensitive features. Perform hierarchical clustering on the obtained sensitive features to obtain the optimal features. Here we set the number of first-selected features to 120, and the feature threshold of the

Variance-Relief-F technology is 40 features (the determination of the number of features will be discussed in detail in the next subsection).

When using the Variance-Relief-F technology for feature selection, it is performed in two steps. First, calculate the Relief-F value of 120 first-selected features, and select the top 40 features from large to small, as shown in Figure 7. At the same time, calculate the variance values of these 120 first-selected features, sort them from large to small, and select the top 40 features with variance values, as shown in Figure 8. Then find out the features that are ranked in the top 40 using Relief-F value and variance value as the sensitive features obtained using the Variance-Relief-F technology, as shown in Figure 9. In the figure, the “40” in “common top 40 features” is just a threshold set for the algorithm. The number of selected features will be less than or equal to this threshold. Finally, hierarchical clustering is used to perform hierarchical clustering on sensitive features, remove redundant features to obtain optimal features, and realize adaptive feature selection. Finally, Fuzzy C-means(FCM) clustering analysis is carried out on the optimal features obtained by feature selection, and the fault diagnosis results are obtained. As shown in the Figure 10.

**A. FEATURE PARAMETER SELECTION**

In the previous section, we set the number of first-selected features to 120 and the selection threshold of sensitive features to 40. This is based on the existing engineering practice experience to make the choice. This not only makes it difficult to ensure that the parameters we set are optimal, but also makes this set of fault diagnosis methods rely too much on the experience of existing experts in related fields, which greatly reduces the adaptive ability of the fault diagnosis method. In order to improve the intelligence level of the fault diagnosis method, when facing specific engineering problems,

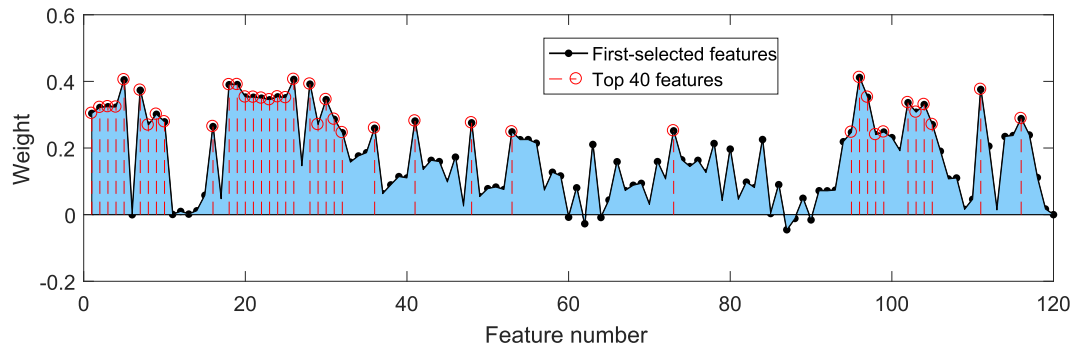


FIGURE 7. Top 40 features based on relief-F value.

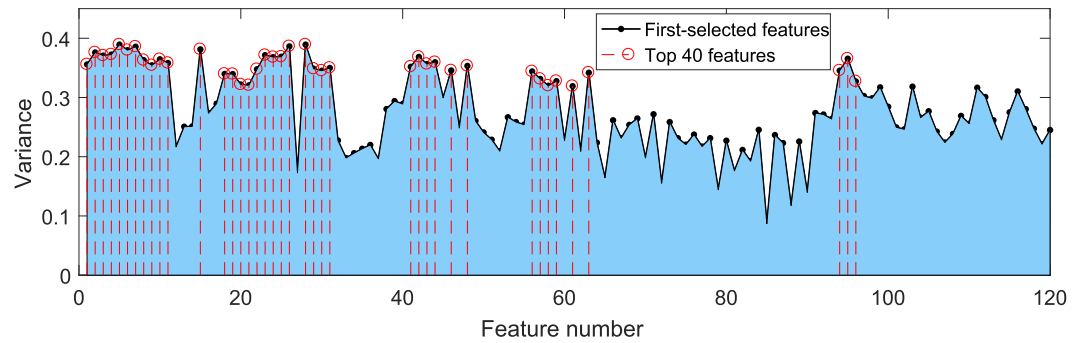


FIGURE 8. Top 40 features based on variance value.

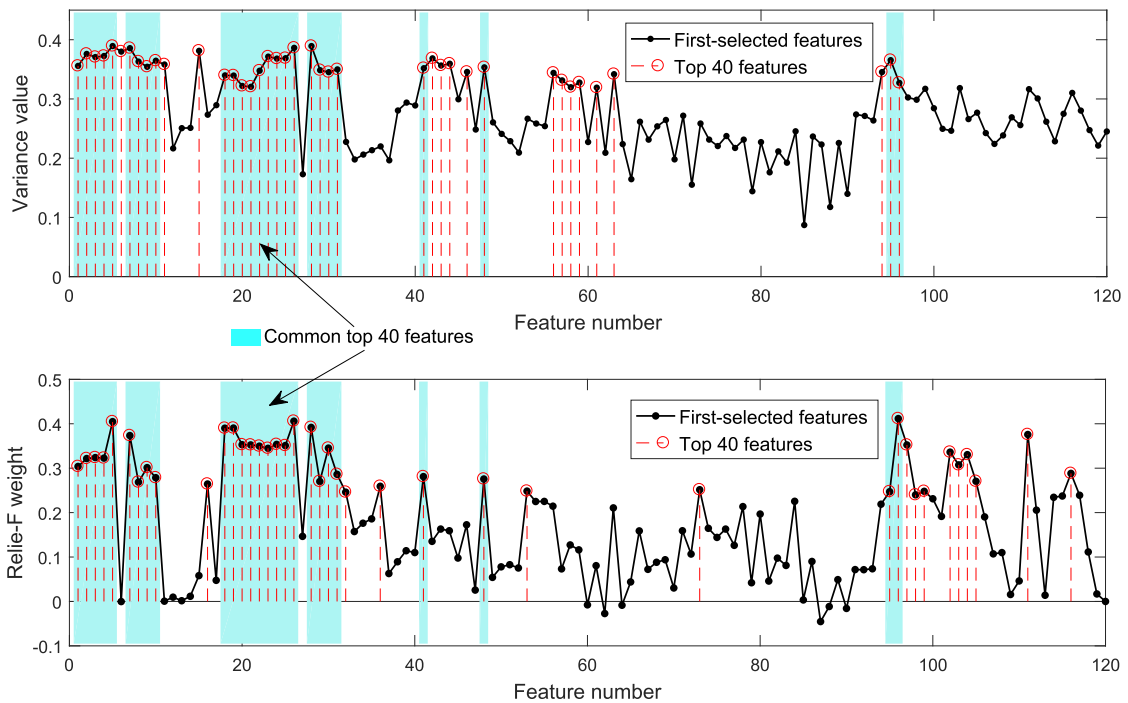


FIGURE 9. Common top 40 features.

it is necessary to determine the threshold of the number of features that need to be set in the fault diagnosis process

through two clustering indicators and a fault diagnosis accuracy indicator in advance. The calculation methods of the two



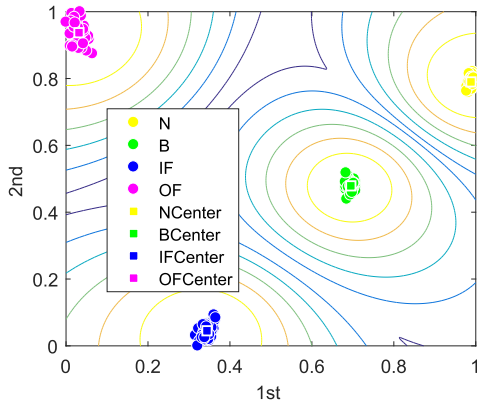


FIGURE 10. The fuzzy c-means (FCM) clustering contour.

clustering indexes (Partition Coefficients and Classification Entropy) are as follows:

$$PC = \frac{1}{n} \sum_{i=1}^c \sum_{k=1}^n u_{ik}^2 \quad (10)$$

$$CE = -\frac{1}{N} \sum_{i=1}^c \sum_{k=1}^n u_{ik} \log(u_{ik}) \quad (11)$$

From the calculation methods of the two clustering indexes, it can be seen that when the Partition Coefficients(PC) index is closer to 1, and the Classification Entropy(CE) index is closer to 0, the better the clustering effect is, and the effect of fault diagnosis will be improved accordingly.

Perform data experiments on the selected signals according to the experiment method in the previous section. When the number of first-selected features is set to 1, and the number of sensitive features is set to 1, calculate the clustering index and accuracy index. Then, when the number of first-selected features is set to 2, and the number of sensitive features is set to 1 and 2, calculate the corresponding clustering index and accuracy index. By analogy, when the number of first-selected features is set from 1 to 240, and the number of sensitive features changes from 1 to 240 accordingly, all the corresponding clustering indicators and accuracy indicators are obtained. Accuracy is the ratio of the number of samples that are accurately diagnosed in the whole fault diagnosis to the whole number of samples. The heatmaps of Partition Coefficients, Classification Entropy and Accuracy indicator are shown in Figure 11, Figure 12 and Figure 13, respectively. A heat map like Figure 13 is drawn by calculating the corresponding accuracy for the cases with different number of first-selected features and sensitive features.

It can be seen from the above index heat map that the optimal clustering effect and fault diagnosis accuracy appear when the number of first-selected features is 30-180, and the number of sensitive features is 10-40. Through the calculation results, it is found that when the number of first-selected features is set to 30, and the number of sensitive features is set to 11, the best evaluation index can be obtained. The

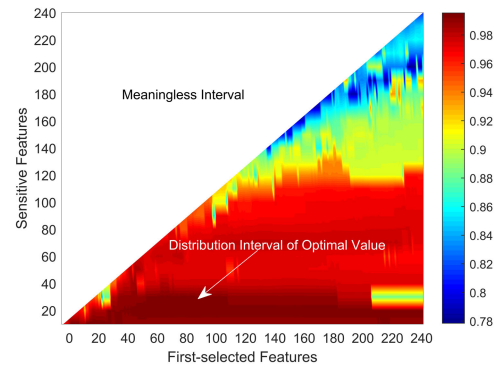


FIGURE 11. The heatmap of partition coefficients.

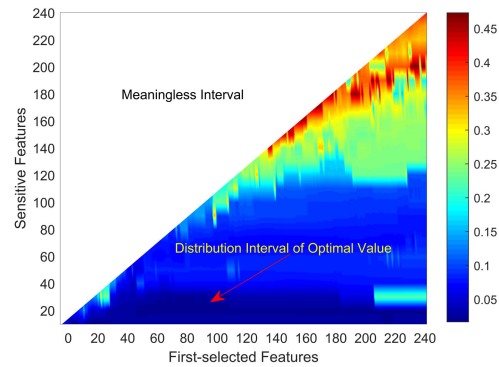


FIGURE 12. The heatmap of classification entropy.

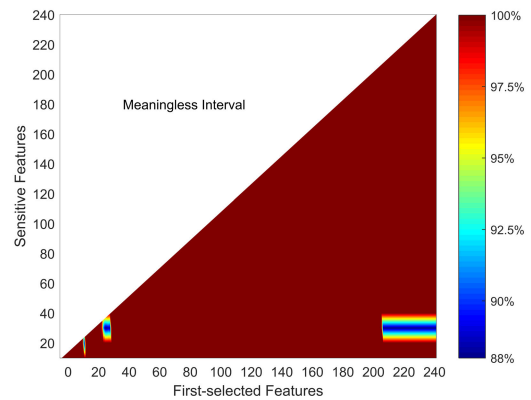


FIGURE 13. The heatmap of accuracy indicator.

corresponding partition coefficients(PC) index is 0.998, classification entropy(CE) is 0.018, and the accuracy is 100%.

### B. COMPARISON STUDY

In order to further investigate the effectiveness of the proposed fault diagnosis method, further data experiments are needed for comparison with other similar fault diagnosis methods. The first fault diagnosis method operates with no feature selection and performs fault diagnosis directly with the Fuzzy C-means(FCM) clustering algorithm. The second fault diagnosis method uses only the variance value to select

**TABLE 2. Comparisons with different fault diagnosis methods.**

Feature selection	Clustering algorithm	Rate of accuracy	Partition Coefficients	Classification Entropy
none	Fuzzy C-means(FCM)	92%	0.825	0.262
variance	Fuzzy C-means(FCM)	97%	0.856	0.092
Relief-F	Fuzzy C-means(FCM)	96%	0.833	0.102
variance + Relief-F	Gath–Geva(GG)	98%	0.901	0.025
variance + Relief-F	Fuzzy C-means(FCM)	100%	0.998	0.018

**TABLE 3. Comparisons with different fault diagnosis methods on bearings with different degrees of damage.**

Feature selection	Clustering algorithm	Rate of accuracy	Partition Coefficients	Classification Entropy
none	Fuzzy C-means(FCM)	91%	0.785	0.276
variance	Fuzzy C-means(FCM)	97%	0.867	0.106
Relief-F	Fuzzy C-means(FCM)	95%	0.833	0.112
variance + Relief-F	Gath–Geva(GG)	98%	0.899	0.036
variance + Relief-F	Fuzzy C-means(FCM)	99%	0.918	0.026

sensitive features and then performs fault diagnosis. The third fault diagnosis method uses only Relief-F algorithm to select sensitive features, and then performs fault diagnosis. The fourth fault diagnosis method operates by replacing the Fuzzy C-means(FCM) algorithm with the Gath–Geva(GG) algorithm for fault clustering. The fault diagnosis results of various methods are shown in Table 2.

As can be seen from Table 2, diagnosing bearing faults based on EEMD and clustering can obtain accuracy rate of more than 90%, which shows that it is feasible and effective to diagnose the bearing fault by clustering. At the same time, it can be seen from the comparative analysis that the fault diagnosis method proposed in this paper can detect faults with better clustering effect and a higher accuracy rate.

The above experiments were performed to diagnose the fault of bearing data with different damage types. In order to further illustrate the effectiveness of this method, we perform experimental analysis on bearing data with different damage levels. We still extracted the data from the Bearing Data Center of Case Western Reserve University. The final experimental results are shown in Table 3. From Table 3, we can see that the proposed method can still achieve good fault diagnosis results for bearings with different degrees of damage. It proves that this method has good adaptability to different working conditions data.

**V. DISCUSSION**

In order to efficiently perform fault diagnosis on rolling bearings, a fault diagnosis method based on clustering and parameter adaptive selection is proposed in this paper. From the data experimental results, it is easy to conclude that the EEMD decomposition of the original signal and then feature extraction and screening can obtain a good clustering effect and fault diagnosis effect. The features in Table 1 are some common features, and the fault diagnosis method proposed in this paper does not need to input all these features into the clustering for fault diagnosis, but to make adaptive selection of these features and select the sensitive features in fault diagnosis. Different fault diagnosis objects have their unique most sensitive features, and the number of sensitive

features unique to different fault diagnosis objects is also different. The adaptive feature selection method proposed in this paper can adaptively select the types and numbers of the corresponding sensitive features according to different fault diagnosis objects. The adaptiveness of this feature selection method is reflected in the fact that the feature selection is always based on the relationship between features and classes. Therefore this method does not differ depending on whether the feature is commonly used or not. Therefore, other features that are not used in this paper can be added when using the method proposed in this paper for fault diagnosis.

In the experimental process, the setting of the feature parameters has a great influence on the fault diagnosis results. By setting the feature parameters in advance, a better diagnosis effect can be obtained. Therefore, in engineering practice, it is very necessary to extract a certain amount of vibration information in advance to obtain the required set of feature parameter values for fault diagnosis of different equipment or its components. In addition, the fault diagnosis method in this paper is based on rolling bearings, and there may be some disadvantages for the fault diagnosis of other mechanical mechanisms, and it is often necessary to modify some processes in the method when using the proposed method for fault diagnosis, in order to obtain the required fault diagnosis effect.

**VI. CONCLUSION**

This paper proposes an adaptive fault diagnosis method for bearings based on clustering. The main idea is to extract features from the original vibration signal, then filter the features layer by layer, and finally use clustering for fault classification and diagnosis. Using the fault diagnosis method proposed in this paper to diagnose the fault of the rolling bearing, a better fault diagnosis effect can be obtained.

In order to obtain as much valuable information as possible from the original signal, it is necessary to perform EEMD decomposition of the signal before performing feature selection. Then the Chi-square Test algorithm is applied to obtain the first-selected features to achieve the initial selection of features. Then Variance-Relief-F Technique is used

to obtain sensitive features. Finally, the optimal features are obtained by removing redundant features using Hierarchical Clustering. For rolling bearing fault diagnosis, the number of first-selected features should be set to 30, sensitive features to about 11, and optimal features to 5.

When performing the selection of sensitive features, the combined use of the two methods, Variance and Relief-F, is more effective for fault diagnosis than when one of these methods is used alone. For the fault diagnosis method proposed in this paper, Fuzzy C-means (FCM) clustering is more effective than Gath–Geva (GG) clustering in classifying fault classes using clustering algorithms.

Although the fault diagnosis method proposed in this paper is more suitable for the fault diagnosis of rolling bearing and has better fault diagnosis effect, it can also be applied to the equipment with rolling bearing. For example, this method can diagnose the faults of rolling bearings on hydraulic pump, fan and generator. Because some non-rolling bearing parts also vibrate when they are running, the method proposed in this paper can also be used for fault diagnosis of other non-rolling bearing parts. For the mechanical equipment system, because there are various types of vibration signals inside, the signals are often intertwined with each other, which brings great challenges to the fault diagnosis. To solve the problem of equipment system fault diagnosis will be the focus of future research. In order to diagnose the fault of equipment system, we need to find a method to separate all kinds of signals, and then analyze the signals to get the fault diagnosis results of each part. After that, we use the knowledge database established in advance to distinguish the fault diagnosis results, and finally form the fault diagnosis results of the system.

## REFERENCES

- [1] Z. Chen, A. Mauricio, W. Li, and K. Gryllias, "A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks," *Mech. Syst. Signal Process.*, vol. 140, Jun. 2020, Art. no. 106683.
- [2] W. Mao, Y. Liu, L. Ding, and Y. Li, "Imbalanced fault diagnosis of rolling bearing based on generative adversarial network: A comparative study," *IEEE Access*, vol. 7, pp. 9515–9530, 2019.
- [3] F. Zou, H. Zhang, S. Sang, X. Li, and X. Liu, "Bearing fault diagnosis based on combined multi-scale weighted entropy morphological filtering and bi-LSTM," *Appl. Intell.*, vol. 51, no. 6, pp. 1–18, 2021.
- [4] J. Gu and Y. Peng, "An improved complementary ensemble empirical mode decomposition method and its application in rolling bearing fault diagnosis," *Digit. Signal Process.*, vol. 113, Jun. 2021, Art. no. 103050.
- [5] B. Cai, H. Liu, and M. Xie, "A real-time fault diagnosis methodology of complex systems using object-oriented Bayesian networks," *Mech. Syst. Signal Process.*, vol. 80, pp. 31–44, Dec. 2016.
- [6] B. Cai, Y. Liu, and M. Xie, "A dynamic-Bayesian-network-based fault diagnosis methodology considering transient and intermittent faults," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 1, pp. 276–285, Jan. 2017.
- [7] 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*, vol. 149, pp. 591–601, May 2021.
- [8] A. Naveen and T. Velmurugan, "Identification of calcification in MRI brain images by k-means algorithm," *Indian J. Sci. Technol.*, vol. 8, no. 29, p. 1, Nov. 2015.
- [9] T. Kawai and S. Akira, "The role of pattern-recognition receptors in innate immunity: Update on toll-like receptors," *Nature Immunol.*, vol. 11, no. 5, p. 373, 2010.
- [10] R. A. Harris, A. Yang, R. C. Stein, K. Lucy, L. Brusten, A. Herath, R. Parekh, M. D. Waterfield, M. J. O'Hare, M. A. Neville, M. J. Page, and M. J. Zvelebil, "Cluster analysis of an extensive human breast cancer cell line protein expression map database," *Proteomics*, vol. 2, no. 2, pp. 212–223, Feb. 2002.
- [11] M. Zarinbal, M. H. F. Zarandi, and I. B. Turksen, "Relative entropy fuzzy C-means clustering," *Inf. Sci.*, vol. 260, pp. 74–97, Mar. 2014.
- [12] L. Serir, E. Ramasso, and N. Zerhouni, "Evidential evolving Gustafson–Kessel algorithm for online data streams partitioning using belief function theory," *Int. J. Approx. Reasoning*, vol. 53, no. 5, pp. 747–768, Jul. 2012.
- [13] G. M. West, S. D. J. McArthur, and D. Towle, "Industrial implementation of intelligent system techniques for nuclear power plant condition monitoring," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7432–7440, Jun. 2012.
- [14] E. De Santis, L. Livi, A. Sadeghian, and A. Rizzi, "Modeling and recognition of smart grid faults by a combined approach of dissimilarity learning and one-class classification," *Neurocomputing*, vol. 170, pp. 368–383, Dec. 2015.
- [15] S. Z. Selim and K. Alsultan, "A simulated annealing algorithm for the clustering problem," *Pattern Recognit.*, vol. 24, no. 10, pp. 1003–1008, 1991.
- [16] B. Sowmya and B. S. Rani, "Colour image segmentation using fuzzy clustering techniques and competitive neural network," *Appl. Soft Comput.*, vol. 11, no. 3, pp. 3170–3178, Apr. 2011.
- [17] H. Guan, B. Xiao, J. Zhou, and M. Guo, "Fast dimension reduction for document classification," *Inf. Sci.*, vol. 222, no. 3, pp. 147–162, 2010.
- [18] J. H. Suh, S. R. T. Kumara, and S. P. Mysore, "Machinery fault diagnosis and prognosis: Application of advanced signal processing techniques," *CIRP Ann.-Manuf. Technol.*, vol. 48, no. 1, pp. 317–320, 1999.
- [19] J. A. A. Daviu and R. A. O. Rios, "Signal processing techniques for transient fault diagnosis," in *Electrical Systems 2*. Hoboken, NJ, USA: Wiley, 2020, ch. 2, pp. 37–75. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119720584.ch2>, doi: [10.1002/9781119720584.ch2](https://doi.org/10.1002/9781119720584.ch2).
- [20] L. Qiao, A. Esmaily, and H. G. Melhem, "Structural damage detection using signal pattern-recognition," *Key Eng. Mater.*, vols. 400–402, pp. 465–470, 2009.
- [21] C. Bo and C. Zang, "Unsupervised structure damage classification based on the data clustering and artificial immune pattern recognition," in *Proc. 8th Int. Conf. Artif. Immune Syst. (ICARIS)*, York, U.K., Aug. 2009, pp. 206–219.
- [22] T. Behrens, A.-X. Zhu, K. Schmidt, and T. Scholten, "Multi-scale digital terrain analysis and feature selection for digital soil mapping," *Geoderma*, vol. 155, nos. 3–4, pp. 175–185, Mar. 2010.
- [23] X. He, M. Ji, C. Zhang, and H. Bao, "A variance minimization criterion to feature selection using Laplacian regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 2013–2025, Oct. 2011.
- [24] M. Van and H. Kang, "Bearing-fault diagnosis using non-local means algorithm and empirical mode decomposition-based feature extraction and two-stage feature selection," *IET Sci., Meas. Technol.*, vol. 9, no. 6, pp. 671–680, Sep. 2015.
- [25] V. Sugumar, V. Muralidharan, and K. I. Ramachandran, "Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing," *Mech. Syst. Signal Process.*, vol. 21, no. 2, pp. 930–942, Feb. 2007.
- [26] I. Kononenko, "Estimating attributes: Analysis and extensions of RELIEF," in *Proc. Eur. Conf. Mach. Learn.*, 1994, pp. 171–182.
- [27] G. Huang, X. Li, B. Zhang, and J. Ren, "PM2.5 concentration forecasting at surface monitoring sites using GRU neural network based on empirical mode decomposition," *Sci. Total Environ.*, vol. 768, May 2021, Art. no. 144516.
- [28] H. Liu, Q. Zhan, C. Yang, and J. Wang, "The multi-timescale temporal patterns and dynamics of land surface temperature using ensemble empirical mode decomposition," *Sci. Total Environ.*, vol. 652, pp. 243–255, Feb. 2019.
- [29] T. Wang, M. Zhang, Q. Yu, and H. Zhang, "Comparing the applications of EMD and EEMD on time–frequency analysis of seismic signal," *J. Appl. Geophys.*, vol. 83, pp. 29–34, Aug. 2012.
- [30] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy C-means clustering algorithm," *Comput. Geosci.*, vol. 10, no. 2–3, pp. 191–203, 1984.
- [31] T. C. Havens, J. C. Bezdek, C. Leckie, L. O. Hall, and M. Palaniswami, "Fuzzy C-means algorithms for very large data," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 6, pp. 1130–1146, Dec. 2012.
- [32] K. Xu, W. Pedrycz, and Z. Li, "Augmentation of the reconstruction performance of fuzzy C-means with an optimized fuzzification factor vector," *Knowl.-Based Syst.*, vol. 222, Jun. 2021, Art. no. 106951.

[33] H. Qin, Q. Fei, X. Ma, and W. Chen, "A new parameter reduction algorithm for soft sets based on chi-square test," *Appl. Intell.*, vol. 51, nos. 4–5, pp. 1–13, 2021.

[34] J. D. Flora, Jr., "Chi-square test," *Ann. Thoracic Surg.*, vol. 33, no. 6, p. 642, 1982. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S000349751060834X>, doi: 10.1016/S0003-4975(10)60834-X.

[35] L.-X. Zhang, J.-X. Wang, Y.-N. Zhao, and Z.-H. Yang, "A novel hybrid feature selection algorithm: Using ReliefF estimation for GA-wrapper search," in *Proc. Int. Conf. Mach. Learn. Cybern.*, 2004, pp. 380–384.

[36] Z. Abbasi, M. Rahmani, and H. Ghaffarian, "IFSB-ReliefF: A new instance and feature selection algorithm based on ReliefF," *Signal Data Process.*, vol. 17, no. 4, pp. 49–66, 2021.



**ABDULRAHAMAN SHUAIBU AHMAD** received the bachelor's degree in mechanical engineering from the Kano University of Science and Technology (KUST), Wudil, Nigeria, and the master's degree in mechanical engineering from the Liaoning University of Technology (LUT), China. He is currently pursuing the Ph.D. degree in mechanical engineering with Central South University, Changsha, China.

His research interests include modeling and evaluation of welding induced residual stresses, mechanical dynamics theory and application, and post-weld shot peening.



**JINGBAO HOU** (Member, IEEE) was born in Guiyang County, Hunan, China, in 1989. He received the B.S. degree in mechanical engineering from the Hunan University of Science and Technology, Xiangtan, Hunan, in 2012, and the M.S. degree in machine design and theory from the Changsha Institute of Mining Research, Changsha, Hunan, in 2015. He is currently pursuing the Ph.D. degree in mechanical engineering with Central South University, Changsha.

Since 2019, he has been an Engineer with the Changsha Institute of Mining Research Company Ltd., where he was an Assistant Engineer, from 2015 to 2018. His research interests include intelligent fault diagnosis, artificial intelligence, and intelligent technology of deep sea equipment.



**HAI GONG** received the B.S., M.S., and Ph.D. degrees in mechanical engineering from Central South University, Changsha, China, in 2004, 2007, and 2011, respectively.

He is currently an Associate Professor with the Light Alloy Research Institute, Central South University. He has published over 30 articles in refereed journals and conferences. His interests research include mechanical fault diagnosis, mechanical dynamics theory and application, development of neutron diffraction stress spectrometer, nondestructive detection, and characterization of deep stress fields in components.



**YUNXIN WU** was born in Xinning County, Guangdong Province, China, in 1963. He received the B.S. and M.S. degrees in mechanical engineering from Central South University, Changsha, China, in 1983 and 1986, respectively, and the Ph.D. degree in mechanical engineering from the University of Mons, Belgium, in 1999.

He is currently a Professor with the Light Alloy Research Institute, Central South University. He has published over 200 articles in refereed journals and conferences. His research interests include mechanical fault diagnosis, mechanical dynamics theory and application, mechanical behavior of metallurgical machinery, residual stress detection, and reduction of aluminum alloy materials.

Dr. Wu was a recipient of Education Department's "New Century Talents Support program" and the National Science and Technology Progress Award. He was a Distinguished Professor of "Furong Scholar program" in Hunan.



**LEI LIU** was born in Rizhao, Shandong, China, in 1991. He received the B.S. degree in mechanical engineering from Central South University, Changsha, China, in 2014, where he is currently pursuing the Ph.D. degree with the Light Alloy Research Institute.

He has published over ten articles in refereed journals. His research interests include mechanical dynamics theory and application, and advanced forming technology of light alloy.

...