Research Article

The Fault Diagnosis of Rolling Bearing Based on Variational Mode Decomposition and Iterative Random Forest

Xiwen Qin ^{1,2}, Jiajing Guo ¹, Xiaogang Dong ¹ and Yu Guo¹

¹School of Mathematics and Statistics, Changchun University of Technology, Changchun 130012, China

²Graduate School, Changchun University of Technology, Changchun 130012, China

Correspondence should be addressed to Xiaogang Dong; dongxiaogang@ccut.edu.cn

Received 10 July 2019; Accepted 5 October 2019; Published 12 February 2020

Guest Editor: Franco Concli

Copyright © 2020 Xiwen Qin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rolling bearing is a critical part of machinery, whose failure will lead to considerable losses and disastrous consequences. Aiming at the research of rotating mechanical bearing data, a fault identification method based on Variational Mode Decomposition (VMD) and Iterative Random Forest (iRF) classifier is proposed. Furthermore, EMD and EEMD are used to decompose the data. At the same time, three mainstream classifiers were selected as the benchmark model. The results show that the proposed model has the highest recognition accuracy.

1. Introduction

Rolling bearing is a kind of precise mechanical component which can change the sliding friction between the running shaft and the shaft seat into rolling friction so as to reduce the friction loss [1]. It is one of the most widely used and easily damaged parts in mechanical equipment. Whether the rolling bearings work properly or not is directly related to the safety and efficiency of equipment production [2].

However, the vibration signals of rolling bearings collected from the field of mechanical equipment are disturbed by serious background noise and other random signals, resulting in the obtained signals having the characteristics of nonlinear, nonstationary, non-Gaussian, and low signal-to-noise ratio [3]. Time-frequency analysis is a very effective method for processing nonlinear and nonstationary signals. Wavelet analysis is one of the time-frequency analysis methods [4] and has been used in fault diagnosis many times [5, 6]. But, the wavelet transform analysis method has its own defects in theory, and it lacks of adaptability. Huang et al. have proposed the Empirical Mode Decomposition (EMD) method, which is an adaptive decomposition method [7]. Since its introduction, it has been widely used in various fields, including fault diagnosis [8, 9]. Inevitably, the EMD method itself has problems such as modal aliasing and endpoint

effects, sensitivity to noise, and selection of interpolation methods. Smith proposed a local mean decomposition (LMD) method for nonlinear and nonstationary signals, which has the same adaptability as EMD [10]. In some respects, the performance of LMD is better than that of EMD. For example, the endpoint effect and modal aliasing are slightly weaker than EMD [11, 12], but it also has some drawbacks such as the lack of fast calculation when the smooth step size cannot be determined to be optimal. On the basis of EMD, Wu and Huang proposed the Ensemble Empirical Mode Decomposition (EEMD) method [13]. This method applies noise-assisted analysis to empirical mode decomposition, promotes antialiasing decomposition, and solves an important defect mode aliasing phenomenon in EMD, reflecting the superiority of EEMD [14, 15]. Recently, a new signal adaptive decomposition method called Variational Mode Decomposition (VMD) was proposed by Dragomiretskiy and Zosso [16]. The biggest advantage of this method is that it overcomes the lack of mathematical theory support for EMD and is not particularly sensitive to noise as EMD. At the same time, VMD avoids the cumulative error of envelope estimation caused by recursive mode decomposition and overcomes the endpoint effect. Since it was proposed, it has attracted much attention and has been proved that the performance of data decomposition is better than other

adaptive data decomposition methods [17]. Since VMD decomposition method was proposed, it has been widely used in various research fields, including fault diagnosis. It has been proved that the performance of VMD is better than other adaptive decomposition methods [18, 19].

As we all know, it is the data-driven era nowadays, so many machine learning methods have been applied to fault recognition. These methods include decision tree [20], support vector machine (SVM) [21], random forest (RF) [22], and so on. Among the existing classification methods, there are different performances in fault diagnosis. Random forest is constructed on the basis of decision tree, so the classification accuracy of the former is better than that of the latter [23, 24]. As far as support vector machine is concerned, the classification accuracy of random forest is usually higher than that of it, but RF is more time consuming [25]. Unfortunately, in the operation of RF algorithm there may be many similar decision trees, which hide the real results. Therefore, it is easy to cause underfitting or overfitting. At present, on the basis of RF, Basu et al. reweighted the selected features and then proposed the iRF method, which made the classification results more reasonable and had higher recognition and prediction ability [26].

For rotating machinery fault data, a diagnosis method based on variational mode decomposition and iterative random forest is proposed. Firstly, the vibration signals of rotating machinery are decomposed into different modal functions by empirical mode decomposition, global empirical mode decomposition, and variational mode decomposition, so that the time domain eigenvalues of each component can be extracted. Then C5.0, SVM, RF, and iRF are used to identify the fault state. Finally, the recognition accuracy of different models is compared.

The paper is organized as follows: Section 2 is dedicated to introduce the VMD method. In the Section 3, we are going to explain the iRF. While in Section 4, empirical analysis and comparison are carried out. Finally, the conclusion is drawn in Section 5.

2. VMD and Characteristic Value

2.1. VMD. VMD is a new signal adaptive decomposition method that can realize nonlinear and nonstationary signal analysis. At the same time, it solves the problems of modal aliasing, endpoint effects, and sensitivity to noise caused by EMD decomposition. The essence of the VMD method is a plurality of adaptive Wiener filter sets and exhibits better noise robustness. In terms of modal separation, VMD determines the frequency center and bandwidth of each component by iteratively searching for the optimal solution of the variational model, adaptively realizing the frequency domain splitting of the signal and the effective separation of the VMD of each component.

The VMD algorithm decomposes the signal into a series of Intrinsic Modal Functions (IMF), and each IMF can be represented as an AM modulated signal.

$$u_k(t) = A_k(t) \cos[\phi_k(t)], \quad (1)$$

where $A_k(t)$ is the instantaneous amplitude, $A_k(t) \geq 0$; $\cos[\phi_k(t)]$ is the instantaneous frequency; and $\phi'_k(t) = \omega_k(t)$ is the instantaneous angular frequency, $\omega_k(t) \geq 0$.

For an input signal f , the ultimate goal of the VMD decomposition method is to decompose this input signal into a number of subsignals u_k , that not only reproduce the input, but also ensure sparsity.

Steps to construct a variational model:

- (1) Obtaining the analytical signal of each modal function u_k by Hilbert transform
- (2) Correcting the respective estimated center frequency ω_k exponentially
- (3) Perform Gaussian smoothing on the demodulated signal to obtain the bandwidth of each segment. Assuming that the signal is decomposed into K IMFs by VMD, the constrained variational model is constructed as

$$\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k(t)} \right\|_2^2 \right\} \quad (2)$$

$$\text{s.t.} \quad \sum_k u_k = f,$$

where $u_k = \{u_1, \dots, u_K\}$ is a set of modal components, and $\omega_k = \{\omega_1, \dots, \omega_K\}$ is a set of central frequencies.

In order to obtain the optimal solution of the variational model, namely the modal function (IMF), the penalty factor α is introduced, and the augmented Lagrange function is constructed:

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) = & \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k(t)} \right\|_2^2 \\ & + \left\| f(t) - \sum_k u_k(t) \right\| + \langle \lambda(t), f(t) - \sum_k u_k(t) \rangle. \end{aligned} \quad (3)$$

The Lagrange function is transformed from the time domain to the frequency domain, and the extremum is obtained to get the frequency domain expression of the modal component u_k and the center frequency ω_k , respectively.

$$\begin{aligned} \hat{u}_k^{n+1}(\omega) = & \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i^{n+1}(\omega) + (\hat{\lambda}(\omega)/2)}{1 + 2\alpha(\omega - \omega_k)^2}, \\ \omega_k^{n+1} = & \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega}. \end{aligned} \quad (4)$$

Then, the alternating direction multiplier algorithm (ADMM) is used to find the optimal solution of the constrained variational model, so that the original signal is decomposed into K narrowband modal components.

2.2. Number of Decomposition. In the implementation of VMD decomposition, the determination of decomposition number is very important. On the basis of many experiments, it is found that the iteration termination condition ε of VMD decomposition increases suddenly and then decreases with the increase in decomposition number, while the sum of squares of errors (ERROR) decreases gradually.

Therefore, a threshold is given to select the minimum K that can make both of them reach at the same time, which will make the best effect of data decomposition. So, this paper presents a threshold method for determining the K value, and the specific steps are as follows:

- (1) Decompose the data set n times, respectively, where $K \in (5, 6, \dots, 15)$
- (2) Record the iteration termination condition ε of each decomposition
- (3) Calculate $ERROR_n$ after each decomposition, $ERROR_n = (f - \sum_{i=1}^K IMF_i)^2$, where n is the number of attempts to decompose
- (4) Given the threshold condition $\gamma = 0.1$, the K value satisfying $ERROR < \gamma$ is chosen as K'
- (5) Finally, choose the K that minimizes ε from K'

2.3. Characteristic Value. When the fault location of rolling bearing is different, the frequency distribution of vibration signal will change, and the time domain characteristics of the fault vibration signal will also change [14]. Therefore, feature extraction of the vibration signal is very important.

Part of the eigenvalues in this paper is obtained by calculating the simple statistics of each IMFs. It includes mean, standard deviation (Std), median, trimmed mean, median absolute deviation (Mad), maximum, minimum, range, kurtosis, skewness, the standard errors of mean (Se), and corrected sum of squares (CSS).

The above statistics can be directly calculated by software. In addition, it is necessary to calculate the other four eigenvalues which includes coefficient of variation, fluctuation index, energy entropy, and information entropy.

2.3.1. Coefficient of Variation. The coefficient of variation will measure the change in degree of the vibration signal amplitude of rolling bearings. For those signals whose amplitude changes regularly, the variation coefficient is a relatively smaller value.

$$CV = \frac{\sigma^2}{\mu^2}, \quad (5)$$

where $\mu = 1/l \sum_{j=1}^l |IMF_{(j)}|$ and $\sigma = \sqrt{1/l \sum_{j=1}^l (IMF_{(j)} - 1/l \sum_{j=1}^l IMF_{(j)})^2}$, l is the length of the intrinsic mode function.

2.3.2. Volatility Index. The volatility index can measure the intensity of signal change. Generally, when the rolling

bearing fails, signal fluctuations will be more intense than normal fluctuations. It is defined as

$$F_i = \frac{1}{l} \sum_{jj=1}^l |IMF_{(j+1)} - IMF_{(j)}|. \quad (6)$$

2.3.3. Energy Entropy. In order to extract the feature, the energy entropy is selected to represent the difference of the characteristics of different types of intrinsic modal functions. According to the definition of Shannon Entropy in information theory, the energy of intrinsic mode function IMF_i can be calculated, and t_1 and t_2 are signal start time and signal end time, respectively.

$$E_1 = \int_{t_1}^{t_2} IMF_i^2(t) dt. \quad (7)$$

The energy entropy can be defined as

$$H = - \sum_{i=1}^n P_i \log P_i. \quad (8)$$

Among them, $P_i = E_i/E$ is the proportion of the energy of the j intrinsic mode function to the whole signal energy, and E is the energy of the whole rolling bearing vibration signal.

2.3.4. Information Entropy. The average amount of information excluding redundancy in information is called "information entropy," which can reflect the uncertainty of the system. Faults of rolling bearings will lead to changes in information entropy. Hypothesis X is a random variable, and information entropy can be defined as

$$\begin{aligned} H(x) &= E[I(x_i)] = E\left[\log\left(2, \frac{1}{p(x_i)}\right)\right] \\ &= - \sum p(x_i) \log(2, p(x_i)), \quad i = 1, 2, \dots, n. \end{aligned} \quad (9)$$

3. Iterative Random Forest

Iterative Random Forest was proposed by Basu et al. [26]. The basic idea of iRF is to obtain random forests with characteristic weights by "iteration reweighting" on the basis of random forests. Then, generalized random intersection trees (RITs) is used to the random forest (RF) with feature weight, and the high-order interaction of features is identified. At the same time, it can ensure that the iRF has high prediction ability.

The implementation of the iRF algorithm consists of three parts, and the detailed steps are as follows:

- (1) Iteratively Reweighted RF: given the number of iterations M , M feature-weighted RFs are generated iteratively on the data D , which is denoted as $RF(w^m)$, $m = 1, \dots, M$. For iterations $m = 1$, set $w^{(1)} = (1/p, \dots, 1/p)$ and store the importance (mean decrease in Gini impurity) of the p features as

$v^{(1)} = (v_1^{(1)}, \dots, v_p^{(1)})$. For iterations $k = 2, \dots, K$, set $w^{(m)} = v^{(m-1)}$. In other words, the setting of the weight is equivalent to the feature importance of the RF in the last iteration.

- (2) Generalized RIT Act on $RF(w^m)$: applying the generalized RIT to the RF with feature-weighted generated by the K th iteration produces a collection of interactions S .
- (3) Bagged Stability Scores: use an “outer layer” of bootstrapping to evaluate the stability of recovered interactions, specifically referred to as generating bootstrap samples data $D_{(b)}$, $b = 1, \dots, B$. Fit $RF(w^m)$ on each bootstrap sample data $D_{(b)}$, and use generalized RIT to identify interactions $S_{(b)}$ across each bootstrap sample. The stability score of an interaction S can be defined as

$$\text{sta}(S) = \frac{1}{B} \sum_{b=1}^B I\{S \in S_{(b)}\}, \quad (10)$$

where $S \in \cup_{b=1}^B S_{(b)}$.

4. Application

The structure framework of the fault diagnosis process is shown in Figure 1. In this flow chart, IE is the abbreviation of important eigenvalues, and classifiers include C5.0, SVM, RF, and iRF. In the experimental part, two data sets with different bearing fault locations will be used to verify the significance of the proposed method.

4.1. Empirical Data. The experimental data of rolling bearings studied in this paper are from the website of Case Western Reserve University. Based on motor experimental platform, torque sensor, power meter, and electronic control equipment, it provides experimental data of normal and fault bearings. The data we use are generated by motors with a rotational speed of about 1797 rpm and fault diameter of 0.014 inches. They contain data of normal bearings, single-point drive end (DE) and fan end (FE) defects, and the sampling frequency is set to 12 khz. There are 168 data samples in this study, including 7 bearing states, and each sample has 5000 data points. Partial data for different fault states are shown in Figure 2.

Case 1. Models are built based on normal bearing data and three kinds of fault state data of single-point DE, which includes normal baseline (NOR), ball fault (BF), inner race fault (IRF), and outer race fault (ORF). **Case 2.** The experimental data consist of NOR and three kinds of FE data (BALL, IR, and OUTF).

Based on Case 1, draw the first sample sequence diagram in each state as shown in Figure 2. It can be clearly seen from Figure 2 that the signal fluctuation trend and amplitude in the fault state are different from those in the normal state.

Therefore, it is necessary to extract features from different state samples for effective fault diagnosis.

4.2. Parameter Setting. It is well known that the success of a model depends largely on the setting of the parameters. In this paper, data decomposition is realized by MATLAB 2016, while eigenvalue calculation and the classification model are implemented by R 3.6.1. Therefore, after repeated experiments and multiple optimizations, the result of parameter setting that maximizes the accuracy of each classification model is obtained.

- (1) EEMD: Nstd = 0.01 and NE = 100
- (2) VMD: alpha = 1000; tau = 0; K = 12; DC = 0; init = 1; and tol = $1e-7$
- (3) C5.0: trials = 7 and CF = 0.25, control = C5.0Control (winnow = F, no GlobalPruning = T, and minCases = 20)
- (4) SVM: kernel = “radial,” type = “Cclassification,” gamma = 0.5, and cost = 25
- (5) RF: ntree = 500, mtry = 1, replace = TRUE, proximity = TRUE, and importance = TRUE
- (6) iRF: n.iter = 7, ntree = 1000, n.core = 1, depth = 5, and n.bootstrap = 50

4.3. Case 1. Fault diagnosis and recognition are carried out based on normal state data and three kinds of fault state data of DE.

4.3.1. Data Decomposition. In the process of data decomposition, three different adaptive decomposition methods are adopted. They are EMD, EEMD, and VMD. For EMD and EEMD, the program automatically gives the number of data decomposition. When it comes to VMD, the determination of K value is a key step. Too little or too much decomposition will affect the establishment of the model and the accuracy of subsequent fault diagnosis. Therefore, the number of K is determined according to the threshold method mentioned in Section 2.2, and $K = 12$. Finally, for Case 1, the EMD was decomposed into 1092 IMF, while the EEMD and VMD were, respectively, 1152 IMFs.

Taking the first sample of spherical faults as an example, some IMFs of three different decomposition methods are shown in Figures 3–5. In Figures 3 and 4, the frequency of each IMF component obtained by decomposing the signal through EMD and EEMD is arranged from high to low, and it can be seen from the figures that both IMF8 components have obvious end effect. But, when it comes to VMD, each component is sequentially distributed from low frequency to high frequency. IMF1 is a low frequency component, and the remaining 3 components exhibit a more regular repetitive impact characteristic. In a word, VMD can clearly separate high frequency and low frequency components and obtain meaningful signals, thus avoiding mode mixing and end-point effect. At the same time, the fact that VMD has strong

FIGURE 1: Flow chart of fault diagnosis experiment based on VMD-iRF.

FIGURE 2: Time series diagram of the sample set under four fault conditions in Case 1. (a) Normal base line. (b) Ball fault. (c) Inner race fault. (d) Outer race fault.

adaptability in signal decomposition is verified. The next step is feature extraction.

4.3.2. Effective Feature Extraction. For various classifiers, to get accurate classification results, the setting of model parameters is very important, and the effective selection of feature values is also the most important. Bearing fault diagnosis is not an exception. Extracting more essential variables as input features can greatly improve the accuracy and efficiency of fault diagnosis.

Firstly, the eigenvalues of IMF are calculated according to the contents in Section 2.3. Then, the importance function of variables in the RF package is used to select effective features. The principle of selecting important variables in this paper is to maximize the fault identification accuracy of

each decomposition method. Therefore, the number of eigenvalues varies with the decomposition method. Table 1 summarizes the importance of eigenvalue variables generated based on the VMD method.

With mean decrease gini as the criterion, given a threshold of 35, the eigenvalues greater than 35 are selected as input variables. Through the selection of important eigenvalue variables corresponding to EMD, EEMD, and VMD, three eigenvector matrices with dimensions of 1092×10 , 1152×10 , and 1152×12 are finally obtained.

4.3.3. Classification. In the part of fault diagnosis, the eigenvector is divided into a training set and test set according to 9:1 to construct the model, and the validity of different classification models is compared. The

FIGURE 3: Partial IMFs of BF1 decomposed by EMD.

FIGURE 4: Partial IMFs of BF1 decomposed by EEMD.

randomly selected training sets were successively input into C5.0, SVM, RF, and iRF, and then the model was detected with the test set. The accuracy of these models for test sets is as shown in Table 2.

Table 2 shows the classification accuracy of 12 different models. Obviously, the recognition accuracy of the model VMD-iRF proposed in this paper is the highest. On the premise of the same classifier, the classification accuracy of VMD decomposition is higher than that of EMD and EEMD. This just right proves that the method proposed in this paper is effective. In this paper, the CPU runtime of each

FIGURE 5: Partial IMFs of BF1 decomposed by VMD.

TABLE 1: The importance of eigenvalues under the VMD.

Characteristic variables	Mean decrease accuracy	Mean decrease gini
Mean	36.1008	69.7278
Std	25.8246	53.9403
Median	19.7956	11.3099
Trimmed	23.7853	33.9644
Mad	23.4675	47.0163
Min	20.5300	52.2737
Max	21.9004	46.3639
Range	20.2021	50.0970
Skew	20.9814	20.4321
Kurtosis	40.1319	81.4969
Se	27.2219	55.9879
CV	36.4835	69.7447
CSS	24.2071	51.2260
Volatility index	9.7392	7.0719
Energy entropy	50.6705	115.4719
Information entropy	25.2835	48.0819

TABLE 2: Accuracy of fault diagnosis for various models.

Decomposition method	Classification accuracy			
	C5.0	SVM	RF	iRF
EMD	0.8087	0.7217	0.8000	0.8348
EEMD	0.7864	0.8058	0.9709	0.9806
VMD	0.7921	0.8119	0.9802	1.0000

model is recorded during the experiment. From the perspective of runtime, due to the complexity of the algorithm, the time of the iRF is higher than others in various decomposition methods. However, its time did not exceed 0.2 seconds.

FIGURE 6: Confusion matrix comparisons of EMD: (a) the result of C5.0 classifier, (b) the result of SVM classifier, (c) the result of RF classifier, and (d) the result of iRF classifier.

The feature set based on EMD is randomly divided into the training set and test set according to the ratio of 9:1. The test sample consists of 115 samples (NOR: 25 samples, BF: 24 samples, IRF: 31 samples, and ORF: 35 samples). The sampled training set is input into three benchmark classifiers (C5.0, SVM, and RF) and iRF, respectively, and then the model is validated with the test set. The confusion matrix of the four classifier test sets is shown in Figure 6. As can be seen from Figure 6, the total number of iRF misjudgments is the least. The effect of other classifiers is relatively poor.

Similarly, the set of feature attributes is divided into training set and test set in 9:1 ratio by EEMD decomposition. Figure 7 is an obfuscation matrix diagram of the test set. The test sample consisted of 103 samples (NOR: 25 samples, BF: 29 samples, IRF: 19, and ORF: 30). As can be seen from Figure 7(d), the number of iRF misjudgments is 2, which is the least in the four models. At the same time, the diagnostic accuracy of RF classifier in Figure 7(c) is also very high, but second to that of iRF. The classification results of

C5.0 and SVM are not ideal. Overall, the accuracy of EEMD classifiers is higher than that of EMD classifiers.

Figure 8 shows the confusion matrix of the test set after VMD decomposition under different classifiers. The total number of test sets is 101. The sample sets of the four states are 27, 20, 32, and 22, respectively. Figure 8(d) is the result of iRF, all of which are correctly classified. 8(c) is the result of RF classifier, in which there are two data judgment errors. In 8(a) and 8(b), not only normal state is misjudged as fault but also fault state is misjudged as normal, especially C5.0 classifier. Every state's data have been misjudged. The classification results of RF classifier and iRF classifier are similar, but iRF classifier is better than iRF classifier. Next, the cross-validation method is used to further verify the superiority of this model.

Given the ratio of training set to test set in advance, it is easy to cause the contingency of experimental results. At the same time, a large number of literatures prove that cross-validation can better reflect the advantages of the proposed method. The essence of cross-validation is to group the

FIGURE 7: Confusion matrix comparisons of EEMD: (a) the result of C5.0 classifier, (b) the result of SVM classifier, (c) the result of RF classifier, and (d) the result of iRF classifier.

original data in a sense, one part as the training set and the other part as the validation set. Firstly, the classifier is trained with the training set, and then the training model is tested with the verification set to evaluate the performance of the classifier. So, the cross-validation is a convincing method to test the validity of the model. Based on the 12 classification models used in this experiment, multi-fold cross-validation is carried out. Repeat 10 times, and then record the average error rate and summarize it in Table 3.

The error rate of each model obtained by cross-validation is summarized in Table 3, and the minimum values corresponding to different folds are coarsened. There are five minimum values in the table, four of which are derived from the model proposed in this paper. The other minimum value is 0.0190, which is calculated by the VMD-RF model, while the value of the VMD-iRF model is 0.0191. There is little difference between them. Referring to the results of other models, it is found that their classification accuracy is far less than that of the method proposed in this paper. By comparing the results of Table 3, it is easy to get results that the error rate of the VMD-iRF fault diagnosis method which was

proposed in this paper is much lower than that of others in all cross-validations. In summary, the method for identifying fault state of rotating machinery proposed in this paper is effective, accurate, and competitive.

4.4. Case 2. This section is an empirical analysis of another data set that is based on FE fault status data and normal state data for troubleshooting. This article will not be as comprehensive as Case 1. The specific data can be found in Supplementary Materials (available here).

The second data set is also decomposed by EMD, EEMD, and VMD, and then 1100, 1152, and 1152 IMFS can be obtained, respectively. Next, through the screening of important variables, the three feature matrices with dimensions 1100×11 , 1152×12 , and 1152×14 are finally obtained. The three input matrices are randomly divided into a training set and a test set according to a ratio of 9:1. Then, C5.0, SVM, RF, and iRF models are established, respectively, and the model is detected by the verification set. The accuracy and runtime of each model are shown in Table 4.

FIGURE 8: Confusion matrix comparisons of VMD: (a) the result of C5.0 classifier, (b) the result of SVM classifier, (c) the result of RF classifier, and (d) the result of iRF classifier.

TABLE 3: Summary of cross-validation results for each classification method.

Model	Cross-validation fold				
	2 folds	4 folds	6 folds	8 folds	10 folds
EMD-C5.0	0.2802	0.2436	0.2463	0.2637	0.2418
EMD-SVM	0.1987	0.1987	0.1987	0.1987	0.1987
EMD-RF	0.1976	0.1963	0.1966	0.1971	0.1975
EMD-iRF	0.2063	0.1795	0.1719	0.1733	0.1678
EEMD-C5.0	0.1656	0.1194	0.1000	0.0844	0.0953
EEMD-SVM	0.1909	0.1909	0.1909	0.1909	0.1909
EEMD-RF	0.0875	0.0903	0.0890	0.0903	0.0889
EEMD-iRF	0.0840	0.0409	0.0307	0.0356	0.0259
VMD-C5.0	0.0755	0.0651	0.0590	0.0564	0.0590
VMD-SVM	0.1094	0.1094	0.1094	0.1094	0.1094
VMD-RF	0.0190	0.0196	0.0194	0.0195	0.0197
VMD-iRF	0.0191	0.0174	0.0156	0.0154	0.0151

It is apparent from Table 4 that the recognition accuracy of each model is high (greater than 0.85). Firstly, regardless of the decomposition method, when only considering the accuracy, we will find that the performance of iRF is always

TABLE 4: Accuracy and time for various models.

Decomposition method	Classification accuracy			
	C5.0	SVM	RF	iRF
EMD	0.8793	0.9051	0.9482	0.9741
EEMD	0.9000	0.9083	0.9500	0.9750
VMD	0.9596	0.9839	0.9919	1.0000

the best, and the order of other classifiers is RF, SVM, and C5.0. When it comes to CPU time, the iRF's is not short, but it is less than 0.3 seconds. The speed of the SVM is the fastest, and the CPU running time of C5.0 and RF is almost the same. Taking the decomposition method into consideration, it is not difficult to find that through the upgrading of decomposition algorithm, the performance of each classifier is getting better and better, among which the classification accuracy of VMD-iRF reaches 100%. The above experimental results show that the proposed VMD-iRF has good classification performance.

In order to further verify the rationality and effectiveness of the method presented in this paper, cross-validation will

TABLE 5: Summary of cross-validation results for each classification method.

Model	Cross-validation fold				
	2 folds	4 folds	6 folds	8 folds	10 folds
EMD-C5.0	0.1449	0.1064	0.1040	0.1014	0.0966
EMD-SVM	0.0827	0.0827	0.0827	0.0827	0.0827
EMD-RF	0.0687	0.0686	0.0681	0.0683	0.0685
EMD-iRF	0.0573	0.0385	0.0390	0.0329	0.0315
EEMD-C5.0	0.1256	0.0921	0.0882	0.0866	0.0839
EEMD-SVM	0.0877	0.0877	0.0877	0.0877	0.0877
EEMD-RF	0.0616	0.0615	0.0606	0.0615	0.0611
EEMD-iRF	0.0497	0.0326	0.0302	0.0297	0.0288
VMD-C5.0	0.0490	0.0407	0.0357	0.0342	0.0319
VMD-SVM	0.0130	0.0130	0.0130	0.0130	0.0130
VMD-RF	0.0101	0.0097	0.0095	0.0092	0.0095
VMD-iRF	0.0150	0.0107	0.0085	0.0071	0.0069

be carried out next. Based on the 12 fault diagnosis models established in case 2, 2, 4, 6, 8, and 10 times of folding and cross-validation were, respectively, conducted, and each cross-validation was repeated 10 times to record the average error rate. The results are shown in Table 5.

Table 5 is a summary of the average error rate of cross-validation of each model, and the bold value is the minimum value for each cross-validation. Three of the five values are derived from VMD-iRF. The others are derived from the VMD-RF model, and the two values correspond to the 2-fold and 4-fold cross-validation, respectively. It illustrates that the VMD-iRF model is more effective in the case of large samples. The error rate of the other two benchmark models is higher than that of VMD-iRF. All in all, the result of Case 2 also proves that the fault identification method proposed in this article is reasonable.

5. Conclusions and Discussion

A fault diagnosis method based on VMD and iRF is proposed. VMD is an adaptive decomposition method that can process nonstationary, nonlinear data, and its decomposition performance is better than other decomposition methods, which has been proved. VMD has been widely used in various fields since it was put forward. The iRF is a new algorithm which is an improved version of RF. Theoretically, the high classification ability of iRF is guaranteed. Besides, EMD and EEMD are also used for data processing. C5.0, SVM, and RF classifiers are used as benchmark models. After determining the decomposition number of VMD by the threshold method, the eigenvalues of each component are calculated and the important feature attributes are extracted. Then, IE is input into each classifier for fault diagnosis. In this paper, fault diagnosis is implemented for two data sets with different fault locations. The results show that the classification effect of VMD-iRF is the best whether the ratio of training set to test set is 9:1 or multi-fold cross-validation. Therefore, this paper not only proves the validity of VMD decomposition but also proves that iRF has high classification accuracy.

There are a lot of nonstationary nonlinear data in real life. The time-frequency analysis method is a new idea to deal with these data. Its application can be seen in different fields

such as electricity, meteorology, medical treatment, and so on. Therefore, the method proposed in this paper can be extended to other fields.

Data Availability

The data of roller bearing vibration are from the website of Case Western Reserve Lab, and they are all available at <https://csegroups.case.edu/bearingdatacenter/pages/download-data-file>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (nos. 11301036, 11226335, and 11571051) and the Scientific Project of Education Department of Jilin Province (nos.2014127 and JJKH20170540KJ). The authors are very grateful for the supports.

Supplementary Materials

The supporting material is the result of the new data experiment in Section 4.4 of the revised manuscript. In this experiment, the training set and the test set are modeled in a ratio of 9:1. The data in Excel are the classification results of test sets of 12 models. (*Supplementary Materials*)

References

- [1] J. Lou, C. Hu, and J. Zhao, "Research on EEMD sample entropy in bearing fault SVM identification," *Mechanical Transmission*, vol. 38, no. 3, pp. 41–45, 2014.
- [2] B. Zhong and R. Huang, *Mechanical Fault Diagnosis*, China Machine Press, Beijing, China, 3rd edition, 2007.
- [3] J. Guo, X. Liu, and H. Wei, "A fault diagnosis method for rolling bearings based on VMD and MED," *Bearing*, no. 6, pp. 50–54, 2018.
- [4] H. Jiang, C. Li, and H. Li, "An improved EEMD with multiwavelet packet for rotating machinery multi-fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 36, no. 2, pp. 225–239, 2013.
- [5] A. Brkovic, D. Gajic, J. Gligorijevic, I. Savic-Gajic, O. Georgieva, and S. Di Gennaro, "Early fault detection and diagnosis in bearings for more efficient operation of rotating machinery," *Energy*, vol. 136, pp. 63–71, 2017.
- [6] X. Wu, Z. Su, and L. Gao, "The development of wavelet and its application in fault diagnosis of mechanical equipment," *Equipment Management and Maintenance*, no. 9, pp. 94–97, 2016.
- [7] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings A*, vol. 454, no. 1971, pp. 903–995, 1998.
- [8] Y. Zhang, *Research on Gearbox Fault Diagnosis Method Based on Hilbert-Huang Transform*, Lanzhou University of Technology, Lanzhou, China, 2014.
- [9] S. Junshan, Y. Cao, and X. Shi, "Fault diagnosis of wind turbine planetary gear box based on EMD and resonance

- remodulation,” in *IEEE 2017 12th International Conference on Computer Science and Education (ICCSE)*, pp. 757–762, Houston, TX, USA, August 2017.
- [10] J. S. Smith, “The local mean decomposition and its application to EEG perception data,” *Journal of the Royal Society Interface*, vol. 2, no. 5, pp. 443–454, 2005.
 - [11] T. Xu, “Study on fault feature extraction method of rolling bearing based on LMD-ICA noise reduction,” *Journal of Beijing University of Posts and Telecommunications*, vol. 40, no. 1, pp. 111–116, 2017.
 - [12] Y. Fu, “Fault diagnosis method for rolling bearings based on LMD-CM-PCA,” *Journal of Vibration, Testing and Diagnosis*, vol. 37, no. 2, pp. 249–255, 2017.
 - [13] Z. Wu and N. E. Huang, “Ensemble empirical mode decomposition: a noise-assisted data analysis method,” *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.
 - [14] O. Ozgonenel, T. Yalcin, I. Guney, and U. Kurt, “A new classification for power quality events in distribution systems,” *Electric Power Systems Research*, vol. 95, no. 95, pp. 192–199, 2013.
 - [15] Q. Xiwen, Q. Li, X. Dong, and S. Lv, “The fault diagnosis of rolling bearing based on Ensemble empirical mode decomposition and random forest,” *Shock and Vibration*, vol. 2017, Article ID 2623081, 9 pages, 2017.
 - [16] K. Dragomiretskiy and D. Zosso, “Variational mode decomposition,” *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.
 - [17] F. Li, “Data-driven time-frequency analysis method based on variational mode decomposition and its application to gear fault diagnosis in variable working conditions,” *Mechanical Systems and Signal Processing*, no. 116, pp. 463–479, 2019.
 - [18] Y. Zhou and Z. Zhu, “A hybrid method for noise suppression using variational mode decomposition and singular spectrum analysis,” *Journal of Applied Geophysics*, vol. 161, pp. 105–115, 2019.
 - [19] T. Gong, X. Yuan, Y. Yuan, X. Lei, and X. Wang, “Application of tentative variational mode decomposition in fault feature detection of rolling element bearing,” *Measurement*, vol. 135, pp. 481–492, 2019.
 - [20] J. R. Quinlan, “Induction of decision trees,” *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986.
 - [21] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
 - [22] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
 - [23] M. Amarnath, V. Sugumaran, and H. Kumar, “Exploiting sound signals for fault diagnosis of bearings using decision tree,” *Measurement*, vol. 46, no. 3, pp. 1250–1256, 2013.
 - [24] M. Cerrada, G. Zurita, D. Cabrera, R.-V. Sánchez, M. Artés, and C. Li, “Fault diagnosis in spur gears based on genetic algorithm and random forest,” *Mechanical Systems and Signal Processing*, vol. 70–71, pp. 87–103, 2016.
 - [25] W. Dong, C. Zuo, and Z.-Y. Xing, “Fault diagnosis and safety region estimation based on SVM and EMD of roller bearing in metro vehicle,” in *Proceedings of the 2017 36th Chinese Control Conference (CCC)*, July 2017.
 - [26] S. Basu, K. Kumbier, J. B. Brown, and B. Yu, “Iterative random forests to discover predictive and stable high-order interactions,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 8, pp. 1943–1948, 2018.