Research Article

Rolling Bearing Fault Diagnostic Method Based on VMD-AR Model and Random Forest Classifier

Te Han and Dongxiang Jiang

State Key Lab of Control and Simulation of Power Systems and Generation Equipment, Department of Thermal Engineering, Tsinghua University, Beijing 100084, China

Correspondence should be addressed to Te Han; hant15@mails.tsinghua.edu.cn

Targeting the nonstationary and non-Gaussian characteristics of vibration signal from fault rolling bearing, this paper proposes a fault feature extraction method based on variational mode decomposition (VMD) and autoregressive (AR) model parameters. Firstly, VMD is applied to decompose vibration signals and a series of stationary component signals can be obtained. Secondly, AR model is established for each component mode. Thirdly, the parameters and remnant of AR model served as fault characteristic vectors. Finally, a novel random forest (RF) classifier is put forward for pattern recognition in the field of rolling bearing fault diagnosis. The validity and superiority of proposed method are verified by an experimental dataset. Analysis results show that this method based on VMD-AR model can extract fault features accurately and RF classifier has been proved to outperform comparative classifiers.

1. Introduction

Rolling bearings are widely used in industrial field, so the study on the method of rolling bearing fault diagnosis has a great significance. Fault diagnosis process mainly includes two important aspects: one is the fault feature information extraction and the other is pattern recognition and classification [1].

Nowadays, the research about fault feature extraction of rolling bearing receives widespread attention [2]. A considerable number of theoretical and experimental studies verified that autoregressive (AR) model is closely related to the characteristics of mechanical systems and AR model parameters are very sensitive to condition change [3–6]. Hence, the condition of system can be evaluated by the feature vector structured by AR model parameters exactly. However, the vibration signal of rolling bearing is typical nonstationary and non-Gaussian signal and AR model analysis is based on stationary random signal. Targeting this problem, some research combined AR model with empirical mode decomposition (EMD) [7, 8]. EMD is proposed by Huang et al. and this method is based on the local characteristic time

scale of signal. By applying EMD method to original signal, a series of stationary intrinsic mode functions (IMFs) can be acquired [9]. Then AR model is established for some selected IMFs and the AR model parameters are treated as feature vectors. Nevertheless, due to the defects of the algorithm, mode mixing is an inevitable problem in EMD. When mode mixing arises, single IMF contains widely different characteristic time scales or a similar time scale appears in adjacent IMFs. The problem would cause each IMF to not be able to reflect the real physical meaning. To overcome this shortcoming, Wu et al. presented the ensemble empirical mode decomposition (EEMD) method using the statistical properties of white noise in 2004 [10, 11]. Dragomiretskiy and Zosso put forward variational mode decomposition (VMD) in 2013 which is entirely nonrecursive and the modes are extracted concurrently [12]. Recently, some researchers focus on VMD method combining with engineering practice [13–16]. An and Zeng analyzed the pressure fluctuation signal from a hydroturbine by using VMD method. Compared with EMD, the research showed that VMD could overcome the shortcoming of mode mixing effectively [13]. Liu et al. applied VMD and EEMD to detect monotonic component

in the degradation signal from a wind turbine gearbox and VMD method got better performance [14]. Wang et al. proposed a method to detect multiple signatures caused by rotor-to-stator rubbing with VMD. VMD extracts all impact signatures successfully, which is superior to conventional EMD, EEMD, and empirical wavelet transform (EWT) [15].

The other key aspect of rolling bearing fault diagnosis is pattern recognition and classification. The pattern recognition method based on artificial neural network (ANN) classifier has been widely studied in mechanical fault diagnosis with its strong ability of self-organizing, self-learning, and nonlinear pattern classification performance [17]. While ANN needs a lot of typical fault data samples and experience knowledge to ensure the accuracy of the network, it brings about great difficulties in practical engineering application [18]. Also, with high-dimensional sample, the results are always unsatisfactory if dimension is not reduced or features are not preselected (e.g., with genetic algorithm (GA)) [19]. Support vector machine (SVM) has stronger generalization ability than neural network. It can get better results when solving the problem of high-dimensional sample classification [19]. However, the problem is that classification accuracy is affected by structural parameters of SVM such as penalty factor and kernel function parameter [20]. To have an ideal effect, scholars studied a series of optimization methods. Both the traditional optimization algorithm like cross validation (CV) and heuristic methods, such as GA and particle swarm optimization (PSO), can optimize the SVM parameters to some extent, while this process requires a lot of calculation and time [21–24].

In addition to ANN and SVM, random forest (RF), which is proposed by Breiman in 2001, has an excellent performance in field of pattern recognition [25]. Different from structural risk minimization principle of SVM, the learning method of RF is ensemble learning [25]. Essentially, RF contains multiple decision tree classifiers and there is no connection between any two decision trees. When the test data is sent into the random forest, every decision tree makes the decision of classification. The most popular class is voted and the final result depends on the majority of trees [19]. Recently, with the advantages of admirable generalization ability, simple structure parameters, and high classification accuracy, RF has received increased interest of scientists in the field of electronic tongue [26], digital soil mapping [27], land cover mapping [28, 29], corrosion monitoring [30], hyperspectral data [31], urban area classification [32], and fault diagnosis in spur gears [33]. Meanwhile, in the field of rolling bearing fault diagnosis, there is few research of RF classifier.

The main contribution of this paper is listed as follows. (1) A recently proposed method VMD is utilized for decomposing vibration signal from rolling bearing to obtain a series of stationary component signals. AR model is established for selected component signals and the parameters and remnant of AR models served as fault characteristic vectors. A fault feature extraction method based on VMD-AR model is presented in this work. (2) A novel random forest classifier is studied in the pattern recognition and classification of rolling bearing. To validate the superiority of RF, the comparisons between RF and some classification methods mentioned above, such as SVM, GA-SVM, and PSO-SVM, are conducted. (3) The effectiveness of proposed VMD-AR-RF method in this paper is confirmed by the experimental dataset which is from Case Western Reserve University Bearing Data Center. This dataset has become a criterion for testing algorithms in the field of rolling bearing fault diagnosis [34]. The remaining part of this paper is organized as follows. VMD algorithm, AR model principle, and RF classifier are investigated in Section 2. In Section 3, we present the rolling bearing fault diagnosis method based on VMD-AR-RF. The experimental dataset validation and results analysis are shown in Section 4. Finally the conclusions are drawn in Section 5.

2. Theoretical Background

2.1. Variational Mode Decomposition. The VMD method decomposes a real signal f into a series of modes u_k which have specific sparsity properties. All these modes u_k can reconstruct the original signal. It can be assumed that each mode is compact around center pulsation ω_k which depends on the decomposition. Hence, the sparsity property of each mode is determined by its bandwidth in frequency spectrum. To have access to the bandwidth of every mode, the following scheme is proposed. (1) Hilbert transform firstly is applied to every mode u_k to gain a unilateral frequency spectrum. (2) Then add an exponential tuned to each estimated center frequency and shift the frequency spectrum of every mode to "baseband." (3) Observe H^1 Gaussian smoothness of demodulated signal and estimate the bandwidth of each mode. Then the problem in the decomposition process is defined as follows [12]:

$$
\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}
$$
$$
\text{s.t.} \quad \sum_k u_k = f, \tag{1}
$$

where f is the original signal, $\{u_k\} = \{u_1, \ldots, u_k\}$ are the modes, and $\{\omega_k\} = \{\omega_1, \ldots, \omega_k\}$ are their center frequency. $\delta(t)$ is the Dirac distribution. In order to confirm the reconstruction constraint, a quadratic penalty term and Lagrangian multipliers, λ, are introduced. The problem can be rewritten in the following format [12]:

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

where α is the balancing parameter of the data-fidelity constraint. Then the problem in (2) can be solved by means of the alternate direction method of multipliers (ADMM). The ADMM algorithm is listed in the following steps [12]:

(a) ADMM optimization algorithm for VMD:

Initialize $\{u_k^1\}, \{\omega_k^1\}, \lambda^1, n \leftarrow 0$

Repeat

$n \leftarrow n + 1$

For $k = 1 : K$ do

Update u_k

$$u_k^{n+1} \longleftarrow \underset{u_k}{\arg\min} L\left(\{u_{i<k}^{n+1}\}, \{u_{i \geq k}^n\}, \{\omega_i^n\}, \lambda^n\right) \quad (3)$$

End for

For $k = 1 : K$ do

Update ω_k

$$\omega_k^{n+1} \longleftarrow \underset{\omega_k}{\arg\min} L\left(\{u_i^{n+1}\}, \{u_{i<k}^n\}, \{\omega_{i \geq k}^n\}, \lambda^n\right) \quad (4)$$

End for

Dual ascent

$$\lambda^{n+1} \longleftarrow \lambda^n + \tau\left(f - \sum_k u_k^{n+1}\right) \quad (5)$$

Until convergence: $\sum_k \|u_k^{n+1} - u_k^n\|_2^2 / \|u_k^n\|_2^2 < \varepsilon$.

(b) Minimization with respect to u_k:

In (3), u_k should be updated. And the subproblem can be equal to the formulation below:

$$u_k^{n+1} = \underset{u_k \in X}{\arg\min} \left\{\alpha \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t)\right] e^{-j\omega_k t} \right\|_2^2 \right. \\ \left. + \left\| f(t) - \sum_i u_i(t) + \frac{\lambda(t)}{2} \right\|_2^2 \right\}. \quad (6)$$

And the solution can be drawn by using Parseval/Plancherel Fourier isometry under L2 norm:

$$\hat{u}_k^{n+1}(\omega) \\ = \left(\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}\right) \frac{1}{1 + 2\alpha(\omega - \omega_k)^2} \quad (7)$$

(c) Minimization with respect to ω_k:

The minimization with respect to ω_k can be given as follows:

$$\omega_k^{n+1} \\ = \underset{\omega_k}{\arg\min} \left\{\left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t)\right] e^{-j\omega_k t} \right\|_2^2\right\},$$

$$\omega_k^{n+1} = \underset{\omega_k}{\arg\min} \left\{\int_0^\infty (\omega - \omega_k)^2 |\hat{u}_k(\omega)|^2 d\omega\right\}, \quad (8)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega}.$$

To summarize the above steps, the complete optimization algorithm for VMD can be illustrated in the following steps:

Initialize $\{\hat{u}_k^1\}, \{\omega_k^1\}, \hat{\lambda}^1, n \leftarrow 0$

Repeat

$n \leftarrow n + 1$

For $k = 1 : K$ do

Update u_k for all $\omega \geq 0$

$$\hat{u}_k^{n+1}(\omega) \\ \longleftarrow \frac{\hat{f}(\omega) - \sum_{i<k} \hat{u}_i^{n+1}(\omega) - \sum_{i>k} \hat{u}_i^n(\omega) + \hat{\lambda}^n(\omega)/2}{1 + 2\alpha(\omega - \omega_k^n)^2} \quad (9)$$

Update ω_k

$$\omega_k^{n+1} \longleftarrow \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 d\omega} \quad (10)$$

End for

Dual ascent for all $\omega \geq 0$

$$\hat{\lambda}^{n+1}(\omega) \longleftarrow \hat{\lambda}^n(\omega) + \tau\left(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega)\right) \quad (11)$$

Until convergence: $\sum_k \|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2 / \|\hat{u}_k^n\|_2^2 < \varepsilon$.

Detailed theory and advantage of VMD were given in the literature [12]. For example, a mixed signal $x(t)$ is defined by the following equation:

$$x(t) = \cos(2\pi f_1 t) + 2\cos(2\pi f_2 t) + \text{noise}, \quad (12)$$

where $f_1 = 50$, $f_2 = 15$, and the noise obeys uniform distribution within $[0 \ 0.1]$ and the mean is zero. EMD and VMD are adopted to decompose the mixed signal, respectively. Results are presented in Figure 1. Eight-component signal c_k and a residual are produced by EMD. In this figure, only the front six components are shown. Another six components on the right side are from VMD. It is clear to observe the severe mode mixing in EMD modes, c_4, c_5, c_6. Compared to EMD, VMD can handle mode mixing effectively. Two real components from $x(t)$ corresponding to u_1, u_2 can be separated accurately.
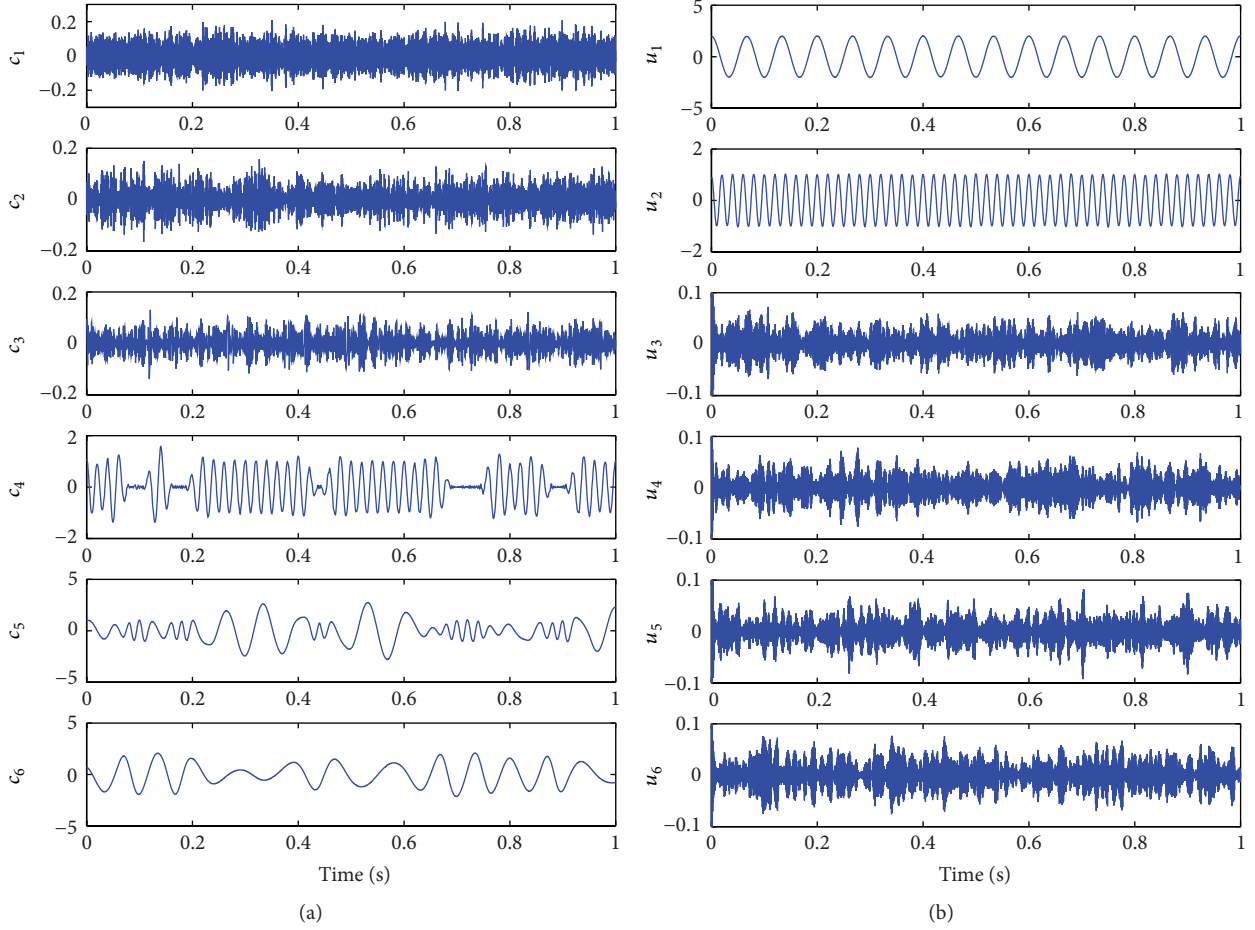
FIGURE 1: Component signals decomposed by (a) EMD and (b) VMD.

2.2. AR Model and Parameters. For system $H(z)$, assume $u(n)$ is a white noise sequence to motivate the system and $x(n)$ is output. Equation (13) can be established:

$$x(n) = -\sum_{k=1}^{p} a_k x(n-k) + \sum_{k=0}^{q} b_k u(n-k), \qquad (13)$$

where a_k $(k = 1, 2, \ldots, p)$ and b_k $(k = 1, 2, \ldots, p)$ are the model parameters of this system, $b_0 = 1$, and p is model order. If all b_k $(k = 1, 2, \ldots, p)$ are zero, the model is called AR model. And it can be illustrated by the identity below:

$$x(n) = -\sum_{k=1}^{p} a_k x(n-k) + u(n), \qquad (14)$$

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 + \sum_{1}^{p} a_k z^{-k}}. \qquad (15)$$

Multiply $x(n+m)$ on both sides of (14) and get average. The result can be expressed as follows:

$$r_x(m) = E\{x(n)x(n+m)\}$$

$$= E\left\{\left[-\sum_{k=1}^{p} a_k x(n+m-k) + u(n+m)\right]x(n)\right\}$$

$$= -\sum_{k=1}^{p} a_k E\{x(n+m-k)x(n)\}$$

$$+ E\{u(n+m)x(n)\}, \qquad (16)$$

where $r_x(m)$ is autocorrelation function of $x(n)$. According to the properties of white noise, (17) can be proved:

$$E\{u(n)x(n-m)\} = \begin{cases} 0, & m \neq 0, \\ \sigma^2, & m = 0, \end{cases} \qquad (17)$$

where σ^2 is the variance of input $u(n)$. The relation between AR model parameters and the autocorrelation function of $x(n)$ can be established. Equation (14) can be transformed into the following form:

$$r_x(m) = \begin{cases} -\sum_{k=1}^{p} a_k r_x(m-k), & m \geq 0, \\ -\sum_{k=1}^{p} a_k r_x(k) + \sigma^2, & m = 0, \end{cases} \qquad (18)$$
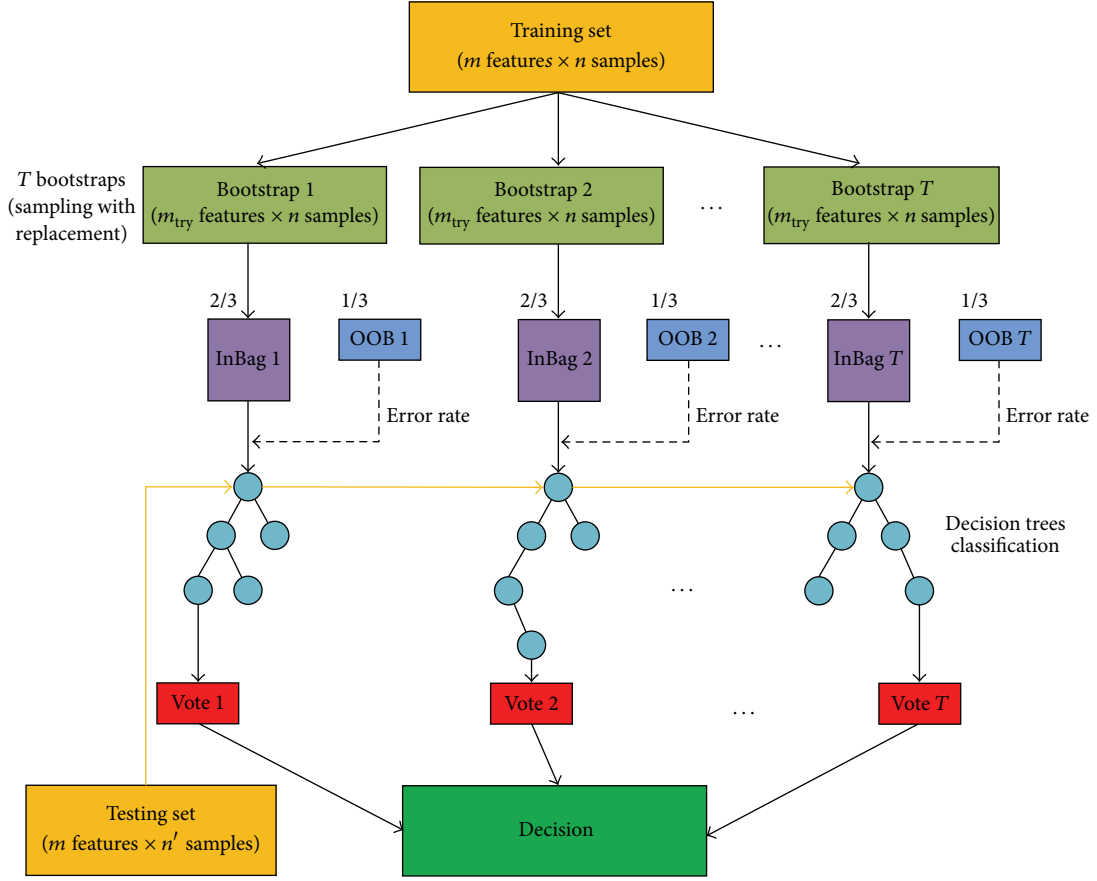
FIGURE 2: Workflow of RF algorithm.

where $\delta(m)$ is the Kronecker delta function. Rewrite (18) into matrix form as follows:

$$
\begin{bmatrix}
r_x(0) & r_x(1) & r_x(2) & \cdots & r_x(p) \\
r_x(1) & r_x(0) & r_x(1) & \cdots & r_x(p-1) \\
r_x(2) & r_x(1) & r_x(0) & \cdots & r_x(p-2) \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
r_x(p) & r_x(p-1) & r_x(p-2) & \cdots & r_x(0)
\end{bmatrix}
\begin{bmatrix}
1 \\ a_1 \\ a_2 \\ \vdots \\ a_p
\end{bmatrix}
$$

$$
=
\begin{bmatrix}
\sigma^2 \\ 0 \\ 0 \\ \vdots \\ 0
\end{bmatrix}.
\tag{19}
$$

In the process of formula reasoning, the even symmetry property of autocorrelation function is used: $r_x(m) = r_x(-m)$. Equations (18) and (19) are called Yule-Walker equations of AR model. It can be seen that AR model with p order consists of $p+1$ parameters, $a_1, a_2, \ldots, a_p, \sigma^2$. When we know the $p+1$ autocorrelation functions, $r_x(0), r_x(1), \ldots, r_x(p)$, the $p+1$ parameters of AR model can be calculated. These parameters can reflect the system condition sensitively [3]. So the AR

model parameters can be adopted to structure the feature vector of rolling bearing system. As for the order p for an AR model, we can get a best value based on Akaike information criterion (AIC). The AIC value of a model is given as follows:

$$
\text{AIC} = n\ln\left(\sigma^2\right) + 2K,
\tag{20}
$$

where σ^2 is the variance of the residuals of the model, K is the number of model parameters, and n is the length of signal. The best model order p is that with the lowest AIC value. The detailed theory of AIC is explained in the literature [6].

2.3. Random Forest (RF) Classifier. RF is an ensemble learning algorithm, which was proposed by Breiman in 2001. Before RF, decision tree method and bagging method are also based on this idea [25]. In fact, RF is an ensemble of unpruned decision trees and the growing algorithm is the same as decision trees. Also, RF selects the best split mode in a randomly chosen feature subset on the basis of bagging idea. RF consists of T trees $\{t_1(x), t_2(x), \ldots, t_T(x)\}$, where $x \in \{x_1, x_2, \ldots, x_m\}$. x is m-dimensional feature vector composed of feature parameters of signal. The outputs $\{\widehat{y}_1 = t_1(x), \widehat{y}_2 = t_2(x), \ldots, \widehat{y}_k = t_k(x)\}$, where $\widehat{y}_k$, $k \in \{1, 2, \ldots, T\}$, is the prediction for a classified object by the kth tree, and a collection of all individuals make a final classification decision. Figure 2 illustrates the workflow of RF
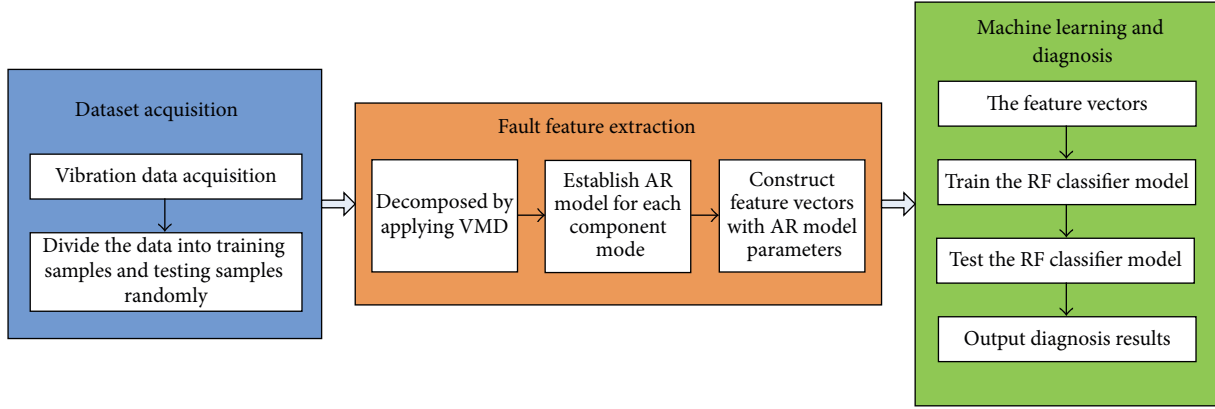
FIGURE 3: Procedures of rolling bearing diagnosis method based on VMD-AR model and RF classifier.

algorithm and the algorithm process is interpreted as follows [19, 25]:

(1) From a training dataset of n samples and m features, draw a bootstrap sample. Sampling method is randomly sampling with replacement. Each bootstrap also has n samples.

(2) For each bootstrap sample, grow a tree with the following modification: firstly, select a subset of m_{try} from m features randomly at each node. Then choose the best split mode. In the process of growing, no pruning is conducted. Finally, the tree is grown to the maximum size.

(3) Repeat the steps above until the number of grown trees reaches T.

(4) Send the testing dataset into RF and aggregate the outputs from T trees. And the classification result is determined by the majority vote.

Beyond the steps mentioned above, actually, in a bootstrap sample, some samples from training dataset are left out, while some others are repeated. It can be calculated that only 2/3 of the training molecules are applied to build the tree for each bootstrap, which is InBag data. The remaining 1/3 are called Out-of-Bag (OOB) samples. For the left 1/3 samples which have not been used to train tree, the classification error rate of corresponding tree can be estimated with these independent ones [19, 25].

Generally speaking, there are only two variables in RF. One is the number of trees in the forest (T) and the other one is the random subset of each node (m_{try}). Furthermore, some researches have shown that the sensitivity of the parameters is weak [26]. This is a significant advantage to dealing with the engineering practice. Another point of concern is the random selected subset of features, m_{try}. It makes the structure of the tree less complete and greedy and increases the possibility that some weak features can have access to the tree and combine with other features. Thus, the local characteristics of each sample can be magnified and the probability of wrong judgement caused by information loss can be reduced. All the votes by trees have a comprehensive assessment for a sample

[30]. RF is a novel and powerful statistical classifier that is well used in other domains but is relatively unknown in the field of rolling bearing fault diagnosis [35].

3. Rolling Bearing Fault Diagnosis Method Based on VMD-AR Model and RF Classifier

Summarizing the method and theory mentioned above, a novel rolling bearing fault diagnosis method based on VMD-AR and RF classifier is proposed. The working process is shown in Figure 3. The main steps are given as follows:

(1) Acquire the vibration signal of rolling bearing under different conditions. For fault location problem, conditions include normal, inner race fault, outer race fault, and ball fault. For problem of fault severity degree, the conditions are concerned with the crack width or depth. Divide data of each condition into 2 groups randomly. One is training dataset and the other is testing dataset.

(2) Decompose the vibration signal into several stationary component modes, u_k. The exact number of u_k depends on the specific characteristics of signal. Usually, the mode mixing problem will arise when the number of decomposed modes by VMD is too small. Also, too large number of modes will cost much computing time.

(3) Establish AR model for each component mode. The order p can be determined by AIC. The parameters $a_1, a_2, \ldots, a_p$ and remnant variance σ^2 served as the variable of feature vector:

$$A_i = \left[a_{i1}, a_{i2}, \ldots, a_{ip}, \sigma_i^2 \right] \quad (i = 1, 2, \ldots, K), \quad (21)$$

where K is the number of component modes.

(4) The feature vector A for a sample is constructed as follows:

$$A = [A_1, A_2, \ldots, A_K]. \quad (22)$$

TABLE 1: Classification results for fault location problem using VMD-AR model and RF classifier.

Number of modes	Number of training samples	Number of testing samples	Classification accuracy (%)				Total accuracy (%)
			Normal	IR fault	B fault	OR fault	
4	80	320	100	100	100	100	100
	160	240	100	100	100	100	100
	240	160	100	100	100	100	100
	320	80	100	100	100	100	100
5	80	320	100	100	100	98.25	99.56
	160	240	100	100	100	98.83	99.71
	240	160	100	100	100	100	100
	320	80	100	100	100	100	100
6	80	320	100	100	100	100	100
	160	240	100	100	100	100	100
	240	160	100	100	100	100	100
	320	80	100	100	100	100	100
7	80	320	100	100	100	99.5	99.88
	160	240	100	100	100	99.83	99.96
	240	160	100	100	100	100	100
	320	80	100	100	100	100	100

(5) Train RF classifier by means of feature vectors corresponding to all conditions. According to literature [25], T, which is the number of trees in forest, is set to 500. Another parameter, m_{try}, in RF classifier which means the number of random subsets from all features can be set to the recommended value in literature [19, 25]:

$$m_{try} = \lfloor \sqrt{m} \rfloor, \tag{23}$$

where m is the number of total features. And check out the performance of trained RF classifier with the testing dataset.

4. Experimental Analysis

To validate the effectiveness of the proposed approach, the experimental dataset from Case Western Reserve University Bearing Data Center (CWRU) is analyzed. This dataset is extensively adopted to test algorithms. The sampling frequency is 48 kHz and rotational speed of rotor is 1772 rpm. The bearing type is deep groove ball bearing (6205-2RS JEM SKF) and single point faults were introduced to test bearings [34].

4.1. Case 1: Results for Fault Location Problem. Firstly, the study to distinguish different fault locations is performed. The dataset contains 4 different conditions (normal, inner race fault, outer race fault, and ball fault). The fault diameter of each fault type is 0.18 mm. There are 100 samples under each condition and 400 samples are acquired totally. In every condition, some samples selected randomly served as training data. The remaining samples are testing data. All the sample signals are decomposed by VMD algorithm. Then AR model is established for every component mode. In fact, the

system condition is mainly decided by the first several AR parameters and the remnant variance. In this paper, we select the first 6 ones and remnant variance to construct feature vector. All training feature vectors are adopted to train RF classifier. In this work, we test different numbers of modes decomposed by VMD, respectively. Different proportions of the training samples to testing samples are also taken into account. The numbers of modes and the proportions of the training samples to testing samples are listed. Meanwhile, we performed 10 times for each test and the average accuracy can be achieved.

From Table 1, we can see the diagnostic accuracies are almost close to 100%. The results show that the method based on VMD-AR model and RF classifier has a high accuracy in the classification problem of fault location. We can also observe that the number of component modes has almost no effect on the classification rates. It is worth mentioning that high diagnostic accuracies can be acquired when only 80 training samples are applied to train the RF classifier. It indicates that RF classifier has an excellent generalization ability and self-learning ability.

Furthermore, to verify the superiority of RF classifier, the diagnosis results of SVM classifier using the same fault feature extraction method and calculating process are presented in Table 2. In this step, the state-of-the-art software LibSVM is adopted and all the default structural parameters are used [36]. One can see that both the amount of the training samples and the number of modes have an obvious influence on the diagnostic results. For 4 component modes, the recognition rates are relatively low. All the rates are smaller than 94%. The diagnostic accuracy can only reach 88.03% when the number of training sets is 80. Nevertheless, the accuracy will increase to 93.13% when the amount of training samples is 320. From another aspect, the diagnostic accuracies have marked improvement when using 5, 6, or 7

TABLE 2: Classification result for fault location problem using VMD-AR model and SVM classifier.

Number of modes	Number of training samples	Number of testing samples	Classification accuracy (%)				Total accuracy (%)
			Normal	IR fault	B fault	OR fault	
4	80	320	99	100	100	53.13	88.03
	160	240	100	100	100	50.83	87.71
	240	160	100	100	100	53.25	88.31
	320	80	100	100	100	72.5	93.13
5	80	320	100	87	98.25	95	95.06
	160	240	100	86.5	98.83	100	96.33
	240	160	100	88.25	99.25	100	96.88
	320	80	100	89.5	99	100	97.13
6	80	320	100	100	88.88	100	97.22
	160	240	100	98	99.17	100	99.29
	240	160	100	100	100	100	100
	320	80	100	100	100	100	100
7	80	320	99.63	96.75	92.13	91	94.88
	160	240	99.67	99.83	97.67	93.5	97.67
	240	160	100	100	100	95.75	98.94
	320	80	100	100	100	94.5	98.63

TABLE 3: Classification results for fault severity level problem using VMD-AR model and RF classifier.

Number of modes	Number of training samples	Number of testing samples	Classification accuracy (%)				Total accuracy (%)
			Normal	0.18 mm	0.36 mm	0.53 mm	
4	80	320	100	97.63	98.38	91.25	96.81
	160	240	100	99.67	98.83	95.83	98.58
	240	160	100	99.5	98.5	96	98.5
	320	80	100	99	98.5	97	98.63
5	80	320	100	99.63	97	98.25	98.72
	160	240	100	100	99.17	99.83	99.75
	240	160	100	100	99	99.5	99.63
	320	80	100	100	100	100	100
6	80	320	100	100	99.5	98.13	99.41
	160	240	100	100	100	98.67	99.67
	240	160	100	100	100	99	99.75
	320	80	100	100	99.5	97.5	99.25
7	80	320	100	100	95.75	94.25	97.5
	160	240	100	100	96.17	97.33	98.38
	240	160	100	99.25	96.5	97.75	98.38
	320	80	100	100	97.5	100	99.38

component modes. Compared with Table 1, the classification rates of RF are considerably higher than SVM.

4.2. Case 2: Results for Fault Severity Level Problem. The second case study is the fault severity level problem. In the dataset from CWRU, the inner race fault data is selected. The fault diameters are 0.18 mm, 0.36 mm, and 0.53 mm, respectively. Following the same procedure above, the diagnostic results are shown in Table 3. From Table 3, we can observe that the classification rates are still high. The diagnostic accuracies are more than 98% in most cases. The results prove that the proposed fault diagnosis method can

also handle the rolling bearing severity level classification effectively. Similarly, the accuracies have no evident changes with different numbers of component modes and proportions of the training samples to testing samples.

Similar tests are carried out on the SVM classifier and the results are given in Table 4. The diagnostic accuracies change with the number of modes. In particular, when the signal is decomposed into 4 component modes, the classification rates are low. When the training samples are 80, we can only acquire a diagnostic accuracy of 81.96%. Also, the amount of training samples has a strong influence on the diagnostic accuracy. Certainly, the superiority of RF is confirmed again.

TABLE 4: Classification results for fault severity level problem using VMD-AR model and SVM classifier.

| Number of modes	Number of training samples	Number of testing samples	Classification accuracy (%)				Total accuracy (%)
			Normal	0.18 mm	0.36 mm	0.53 mm	
4	80	320	97.25	94.75	47.88	88	81.97
	160	240	96.5	100	56.5	94.83	86.96
	240	160	98.25	100	63.25	95	89.13
	320	80	100	100	66	95.5	91
5	80	320	100	89.5	88.88	75.63	88.5
	160	240	100	91.83	89.5	85	91.58
	240	160	100	92.75	90	91.2	93.56
	320	80	100	100	90.5	93	95.88
6	80	320	100	99.13	92	85.38	94.13
	160	240	100	100	94	89.17	95.79
	240	160	100	100	92	91.25	95.81
	320	80	100	100	94.5	89.5	96
7	80	320	98.13	100	89	83.63	92.69
	160	240	99.33	100	93.5	87	94.96
	240	160	98.75	100	94.25	87.25	95.06
	320	80	99.5	100	95	86.5	95.25
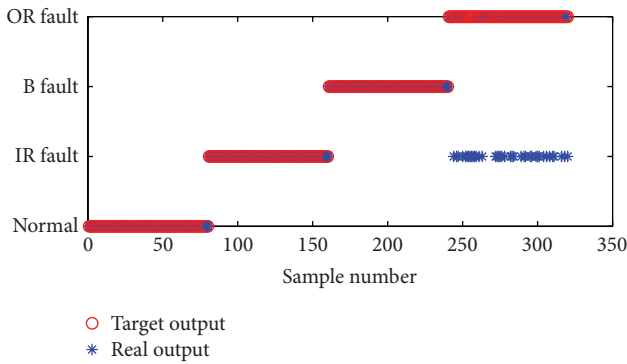
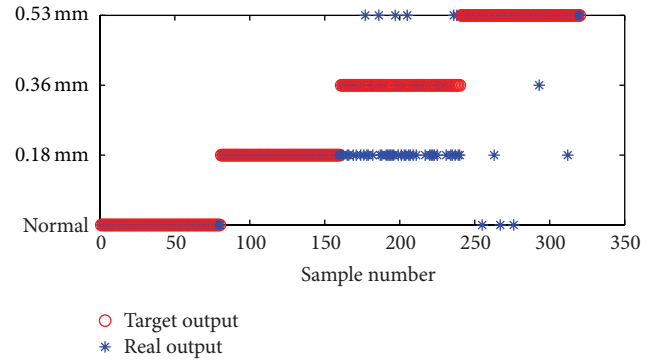FIGURE 4: Real output for fault location problem with SVM classifier under 4 modes.

FIGURE 5: Real output for fault severity level problem with SVM classifier under 4 modes.

4.3. Discussion.
From Table 2, it is clear to find that the classification rates of outer race fault can only be about 53% when using 4 component modes and fewer training samples. Figure 4 is the real output with 320 testing samples in one time test. A considerable amount of outer race fault testing samples is classified into inner race fault incorrectly. From Table 4, we can also observe the classification rates of fault diameters of 0.36 mm are unsatisfactory with 4 component modes and fewer training samples. A real output in one time test is displayed in Figure 5. On one hand, it is plausible that there is mode mixing problem when the signal is decomposed into 4 modes. This defect will bring more difficulties of classification for SVM. On the other hand, the classification success rate of SVM classifier is affected by structural parameters severely. In the further work, we use two optimization algorithms (GA and PSO) to select optimal penalty factor and kernel function parameter for SVM. Referring to literature [22, 23], the maximum generation is

50 and the number of populations is 20 in GA-SVM. For PSO, the maximum generation count is 50, the number of particles is set to 20, acceleration constants $c_1 = 1.5$ and $c_2 = 1.7$, and inertia weight $W = 1$. The diagnostic accuracies are relatively low when using 4 component modes. Hence, the datasets with 4 modes are still adopted for further comparative study. The diagnostic accuracies and consuming time under different classifiers are expressed in Tables 5 and 6. Obviously, the classification performance and generalization ability of SVM can be significantly promoted by GA and PSO optimization algorithms. Nevertheless, the optimization procedure needs a mass of calculation and time. RF classifier has obvious advantages both in computation time and in accuracy. It is worth mentioning that RF is insensitive to the classifier parameters. A quite satisfactory result can be obtained when we select the recommended parameter value.

In general, the fault diagnosis success rates are all over 95% when the fault feature extraction method based on VMD-AR model is combined with different classifiers (RF,

TABLE 5: The performance for fault location problem using different classifiers under 4 modes.

Number of training samples	Number of testing samples	Total accuracy (%)				Cost time (s)			
		RF	SVM	GA-SVM	PSO-SVM	RF	SVM	GA-SVM	PSO-SVM
80	320	100	88.03	99.06	98.97	0.0771	0.009	1.719	4.108
160	240	100	87.71	99.67	99.96	0.1153	0.0104	3.349	9.274
240	160	100	88.31	99.75	99.88	0.1511	0.0107	6.026	16.89
320	80	100	93.13	100	99.88	0.187	0.011	9.266	27.36

TABLE 6: The performance for fault severity level problem using different classifiers under 4 modes.

Number of training samples	Number of testing samples	Total accuracy (%)				Cost time (s)			
		RF	SVM	GA-SVM	PSO-SVM	RF	SVM	GA-SVM	PSO-SVM
80	320	96.81	81.97	95.47	95.69	0.0859	0.0093	2.06	4.041
160	240	98.58	86.96	97.29	96.58	0.1377	0.0103	4.79	9.640
240	160	98.50	89.13	98.06	96.81	0.1822	0.0115	7.77	18.04
320	80	98.63	91.00	98.38	97.63	0.2356	0.0119	11.53	28.34

GA-SVM, or PSO-SVM). It proves that the VMD-AR model method is suitable for extracting feature of rolling bearing vibration signal which is always nonstationary. The comparative analysis indicates that the RF has a better classification capacity compared with the SVM at the time of handling the rolling bearing fault diagnosis problem while costing little time. The training strategy based on ensemble learning of RF is superior to the structural risk minimization principle of SVM in this field.

5. Conclusion

In this paper, a rolling bearing fault diagnosis method based on VMD-AR model and RF classifier is put forward. Firstly, the VMD-AR model method is developed for feature extraction. Secondly, a novel RF classifier is applied to pattern recognition. Finally, the validity of this method is validated by an experimental dataset. The analysis result shows that the VMD-AR model method has a good performance in feature extraction of rolling bearing. Furthermore, the comparative analysis shows that the RF classifier combined with VMD-AR model has a higher diagnosis success rate compared with SVM, GA-SVM, and PSO-SVM, while the time consumption is little. This classifier is pretty suitable to deal with the rolling bearing fault classification problem.

In summary, the novel method based on VMD-AR model and RF classifier is valid for rolling bearing fault diagnosis and has a wide application prospect in engineering practice.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.
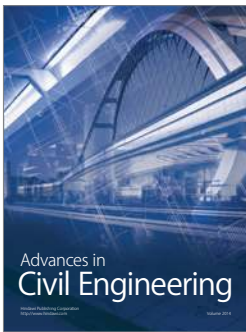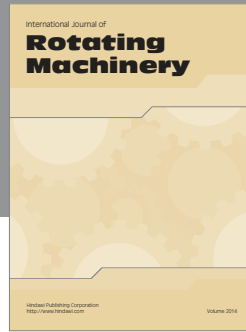
Acknowledgments

References

[1] D. Ho and R. B. Randall, "Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals," *Mechanical Systems and Signal Processing*, vol. 14, no. 5, pp. 763–788, 2000.

[2] Y. Ming, J. Chen, and G. Dong, "Weak fault feature extraction of rolling bearing based on cyclic Wiener filter and envelope spectrum," *Mechanical Systems and Signal Processing*, vol. 25, no. 5, pp. 1773–1785, 2011.

[3] W. Wang and A. K. Wong, "Autoregressive model-based gear fault diagnosis," *Journal of Vibration and Acoustics*, vol. 124, no. 2, pp. 172–179, 2002.

[4] A. G. Poulimenos and S. D. Fassois, "Parametric time-domain methods for non-stationary random vibration modelling and analysis—a critical survey and comparison," *Mechanical Systems and Signal Processing*, vol. 20, no. 4, pp. 763–816, 2006.

[5] C.-C. Wang, Y. Kang, P.-C. Shen, Y.-P. Chang, and Y.-L. Chung, "Applications of fault diagnosis in rotating machinery by using time series analysis with neural network," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1696–1702, 2010.

[6] J. A. F. Diniz-filho, T. F. L. V. B. Rangel, and L. M. Bini, "Model selection and information theory in geographical ecology," *Global Ecology and Biogeography*, vol. 17, no. 4, pp. 479–488, 2008.

[7] C. Junsheng, Y. Dejie, and Y. Yu, "A fault diagnosis approach for roller bearings based on EMD method and AR model," *Mechanical Systems and Signal Processing*, vol. 20, no. 2, pp. 350–362, 2006.

[8] L. Jiang, Y. Liu, X. Li, and A. Chen, "Degradation assessment and fault diagnosis for roller bearing based on AR model and fuzzy cluster analysis," *Shock and Vibration*, vol. 18, no. 1-2, pp. 127–137, 2011.

[9] 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 of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[10] Z. Wu and N. E. Huang, "A study of the characteristics of white noise using the empirical mode decomposition method," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 460, no. 2046, pp. 1597–1611, 2004.

[11] P. Flandrin, G. Rilling, and P. Gonçalvés, "Empirical mode decomposition as a filter bank," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 112–114, 2004.

[12] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.

[13] X. An and H. Zeng, "Pressure fluctuation signal analysis of a hydraulic turbine based on variational mode decomposition," *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*, vol. 229, no. 8, pp. 978–991, 2015.

[14] C. Liu, D. X. Jiang, and W. G. Yang, "Feature extraction in machine degradation with variational mode decomposition," in *Joint Conference: MFPT 2015 and ISA's 61st International Instrumentation Symposium*, Huntsville, Ala, USA, May 2015.

[15] Y. Wang, R. Markert, J. Xiang, and W. Zheng, "Research on variational mode decomposition and its application in detecting rub-impact fault of the rotor system," *Mechanical Systems and Signal Processing*, vol. 60-61, pp. 243–251, 2015.

[16] A. A. Abdoos, P. K. Mianaei, and M. R. Ghadikolaei, "Combined VMD-SVM based feature selection method for classification of power quality events," *Applied Soft Computing*, vol. 38, pp. 637–646, 2016.

[17] Y. Yu, YuDejie, and C. Junsheng, "A roller bearing fault diagnosis method based on EMD energy entropy and ANN," *Journal of Sound and Vibration*, vol. 294, no. 1-2, pp. 269–277, 2006.

[18] M. Zacksenhouse, S. Braun, M. Feldman, and M. Sidahmed, "Toward helicopter gearbox diagnostics from a small number of examples," *Mechanical Systems and Signal Processing*, vol. 14, no. 4, pp. 523–543, 2000.

[19] V. Svetnik, A. Liaw, C. Tong, J. Christopher Culberson, R. P. Sheridan, and B. P. Feuston, "Random forest: a classification and regression tool for compound classification and QSAR modeling," *Journal of Chemical Information and Computer Sciences*, vol. 43, no. 6, pp. 1947–1958, 2003.

[20] K. Duan, S. S. Keerthi, and A. N. Poo, "Evaluation of simple performance measures for tuning SVM hyperparameters," *Neurocomputing*, vol. 51, pp. 41–59, 2003.

[21] S. Luo, J. Cheng, and H. Ao, "Application of LCD-SVD technique and CRO-SVM method to fault diagnosis for roller bearing," *Shock and Vibration*, vol. 2015, Article ID 847802, 8 pages, 2015.

[22] H. L. Ao, J. Cheng, K. Li, and T. K. Truong, "A roller bearing fault diagnosis method based on LCD energy entropy and ACROA-SVM," *Shock and Vibration*, vol. 2014, Article ID 825825, 12 pages, 2014.

[23] J. J. Liang, A. K. Qin, P. N. Suganthan, and S. Baskar, "Comprehensive learning particle swarm optimizer for global optimization of multimodal functions," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 3, pp. 281–295, 2006.

[24] H. Xu and G. Chen, "An intelligent fault identification method of rolling bearings based on LSSVM optimized by improved PSO," *Mechanical Systems and Signal Processing*, vol. 35, no. 1-2, pp. 167–175, 2013.

[25] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[26] M. Liu, M. Wang, J. Wang, and D. Li, "Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: application to the recognition of orange beverage and Chinese vinegar," *Sensors and Actuators B: Chemical*, vol. 177, pp. 970–980, 2013.

[27] R.-M. Yang, G.-L. Zhang, F. Liu et al., "Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem," *Ecological Indicators*, vol. 60, pp. 870–878, 2016.

[28] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, no. 1, pp. 93–104, 2012.

[29] V. F. Rodriguez-Galiano, M. Chica-Olmo, F. Abarca-Hernandez, P. M. Atkinson, and C. Jeganathan, "Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture," *Remote Sensing of Environment*, vol. 121, pp. 93–107, 2012.

[30] N. Morizet, N. Godin, J. Tang, E. Maillet, M. Fregonese, and B. Normand, "Classification of acoustic emission signals using wavelets and Random Forests: application to localized corrosion," *Mechanical Systems and Signal Processing*, vol. 70-71, pp. 1026–1037, 2016.

[31] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 492–501, 2005.

[32] S. Hariharan, S. Tirodkar, and A. Bhattacharya, "Polarimetric SAR decomposition parameter subset selection and their optimal dynamic range evaluation for urban area classification using Random Forest," *International Journal of Applied Earth Observation and Geoinformation*, vol. 44, pp. 144–158, 2016.

[33] 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.

[34] http://csegroups.case.edu/bearingdatacenter.

[35] D. R. Cutler, T. C. Edwards Jr., K. H. Beard et al., "Random forests for classification in ecology," *Ecology*, vol. 88, no. 11, pp. 2783–2792, 2007.

[36] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, article 27, 2011.