

# Intelligent fault diagnosis of roller bearings with multivariable ensemble-based incremental support vector machine



XiaoLi Zhang <sup>a,b</sup>, BaoJian Wang <sup>a</sup>, XueFeng Chen <sup>a,\*</sup>

<sup>a</sup>State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, PR China

<sup>b</sup>Key Laboratory of Road Construction Technology and Equipment, Ministry of Education, Chang'an University, Xi'an 710064, PR China

## ARTICLE INFO

### Article history:

Received 17 October 2014

Received in revised form 22 May 2015

Accepted 24 June 2015

Available online 29 June 2015

### Keywords:

Multivariable

Ensemble

Incremental

Support vector machine

Intelligent fault diagnosis

Roller bearing

## ABSTRACT

Since roller bearings are the key components in rotating machinery, detecting incipient failure occurring in bearings is an essential attempt to assure machinery operational safety. With a view to design a well intelligent system that can effectively correlate multiple monitored variables with corresponding defect types, a novel intelligent fault diagnosis method with multivariable ensemble-based incremental support vector machine (MEISVM) is proposed, which is testified on a benchmark of roller bearing experiment in comparison with other methods. Moreover, the proposed method is applied in the intelligent fault diagnosis of locomotive roller bearings, which proves the capability of detecting multiple faults including complex compound faults and different severe degrees with the same fault. Both experimental and engineering test results illustrate that the proposed method is effective in intelligent fault diagnosis of roller bearings from vibration signals.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

Roller bearing is the most common component in industrial rotating machinery. Defects in roller bearings can lead to machinery malfunctioning, even to economical loss and human casualties. Therefore, fault diagnosis of roller bearing plays a very important role in machinery maintenance and process automation. Various signal processing and pattern recognition approaches are proposed as available diagnostics tools [1]. It is worthy to note that time-frequency analysis including wavelet analysis [2–5] and Hilbert-Huang transform [6–8] can reveal defect frequency characteristics of the vibration signal. However, there are a number of factors affecting the complexity of the bearing signal [9]. Firstly, it is almost impossible to precisely determine the bearing characteristic frequencies with variation of bearing geometry and assembly. Secondly, different locations of a bearing defect cause different behaviors in transient signal response, which is easily buried in the wide band resonance and noise signals. Thirdly, it is hard to distinguish signals and extract their representative features in different severity stages with a same defect type. Fourthly, operational speed and load of the shaft greatly affect the way and the degree of a machinery vibration, which also affect the measured vibration signals. Finally, it is also impossible to obtain

the baseline information about some particular bearing, which makes signal processing methods impractical. All above, there is a great demand of constructing a reliable, intelligent and automated procedure for condition monitoring and diagnosis of roller bearing.

Pattern recognition techniques, such as neural network [10–12], decision tree [13–15], support vector machine [16–18], can offer diagnostic decision by learning the latent rules from the observed signals or measured variables with a fault type. However, there is a bottleneck that machinery fault samples are often very scarce. Support vector machine is of specialties for small samples based on statistical learning theory and therefore exhibits good performance in intelligent fault diagnosis of roller bearings [19,20].

Support vector machine usually require representative training samples to generate an appropriate decision boundary among different classes, so the whole training data is usually attained in prior and trained in one batch [21]. When a new data belonging to a new class occurs, a typical approach to learn new knowledge involves discarding the existing classifier and retraining another classifier with representative samples that have been accumulated so far. This learning style of support vector machine results in loss of previously acquired information, which is known as catastrophic forgetting phenomenon [22]. Moreover, it brings data storage burden and computation waste for repetitious retraining. The terrible case is that the previous training data was lost in the past and the new unseen data is not coming, which is infeasible for support

\* Corresponding author. Tel./fax: +86 29 82663689.

E-mail address: chenxf@mail.xjtu.edu.cn (X. Chen).

vector machine to learn adequate knowledge so as to recognize the whole patterns. To tackle the problem, incremental based learning methods have been proposed so far. Laskov et al. focuses on the design and analysis of efficient incremental support vector machine learning with the aim of providing a fast, numerically stable and robust implementation [23]. A detailed analysis of convergence and algorithmic complexity of incremental support vector machine learning is carried out in Ref. [23]. Moreover, Parikh et al. develop an ensemble-based incremental learning approach for data fusion application [24]. The experimental results indicate that the ensemble-based incremental learning approach often performs better than an ensemble classification system or a single classifier [24]. Erdem et al. proposed an ensemble of support vector machine for incremental learning on optical character recognition and volatile organic compounds dataset [25], which is proved to be capable of learning new information from subsequent datasets including new instances of previously unseen classes.

Since machinery fault samples are generally attained little by little as machinery condition varies in operation and representative fault samples are very scarce, it is important to learn fault information incrementally so as to make correct machinery condition assessment utilizing previously learned knowledge and newly attained data in an online fashion without requiring access to previous data. With a view to design a well intelligent system that can effectively correlate multiple monitored variables with defect types, a multivariable ensemble-based incremental support vector machine (MEISVM) is proposed to improve classification performance of typical support vector machine and learn new information gradually. The highlights and contributions are as follows:

- Multivariable ensemble-based incremental support vector machine is proposed to improve classification performance of support vector machine, since machinery fault samples are gradually attained as machinery condition varies in operation and the representative fault samples are very scarce.
- Multivariable ensemble-based incremental support vector machine has well capacity of correlating the measured multiple variables with a corresponding fault type of roller bearing from sensor information, which exhibits good performance in comparison with other proposed methods in literature on a benchmark of roller bearing experiment.
- Multivariable ensemble-based incremental support vector machine is applied in the intelligent fault diagnosis of locomotive roller bearings, which can detect multiple faults including complex compound faults and different severe degrees with a same fault type.

The organization of the paper is as follows. The basic theory of support vector machine is briefly reviewed in Section 2. The multivariable ensemble-based incremental support vector machine is explained in Section 3. Experiments are conducted to testify the effectiveness of the proposed methods in contrast to other methods in Section 4. The proposed method is finally conducted to detect complex faults of locomotive roller bearings in Section 5. General conclusions are drawn in Section 6.

## 2. The basic theory of support vector machine

Support vector machine is a powerful machine learning method for classification and regression problems based on statistical learning theory and structural risk minimization principle [19,26]. Most cases in practical are multi-classes, such as fault diagnosis. Many approaches have been proposed to extend the binary support vector machine to multi-class problems. The most common strategies are called “one-against-one” and

“one-against-all”. Many comparison studies and arguments show that the one-against-one approach outperforms other approaches in many cases [27–30]. Therefore, classification is done using one-against-one approach, which is briefly introduced as follows.

Given a training sample set in the input space

$$ST = \{(\mathbf{x}_t, y_t) | \mathbf{x}_t \in H, y_t \in \{1, 2, \dots, k\}, t = 1, \dots, l\} \quad (1)$$

where  $\mathbf{x}_t$  is an input vector,  $y_t$  is the corresponding label of  $\mathbf{x}_t$ ,  $l$  is the number of the training samples, and  $k$  is the number of different classes. One-against-one support vector machine constructs  $k(k - 1)/2$  binary classifiers to recognize different classes. For training data from  $i$ th and  $j$ th classes, the following binary classification problem can be solved:

$$\text{Minimize} \quad \frac{1}{2} (\mathbf{w}^{ij})^T \mathbf{w}^{ij} + C \sum_t \xi_t^{ij} (\mathbf{w}^{ij})^T \\ (\mathbf{w}^{ij})^T \Phi(\mathbf{x}_t) + b^{ij} \geq 1 - \xi_t^{ij} \quad \text{if } y_t = i, \quad (2)$$

$$\text{Subject to} \quad (\mathbf{w}^{ij})^T \Phi(\mathbf{x}_t) + b^{ij} \leq 1 - \xi_t^{ij} \quad \text{if } y_t = j, \\ \xi_t^{ij} \geq 0. \quad (3)$$

where  $\Phi(\mathbf{x}_i)$  is the vector mapped from the input space of  $\mathbf{x}_i$ ,  $\mathbf{w}^{ij}$  is the weight vector,  $\xi_t^{ij}$  is a slack variable,  $C$  is a penalty constant. By incorporating kernels and rewriting it in Lagrange multipliers, the above binary classification problem can be transformed into a dual quadratic optimization problem and finally forms decision function.

$$f^{ij}(\mathbf{x}) = \text{sign} \left( \sum_{t=1}^l y_t \alpha_t^{ij} \cdot K(\mathbf{x}, \mathbf{x}_t) + b^{ij} \right) \quad (4)$$

where  $\alpha_t^{ij}$  is the Lagrange multiplier,  $K(\mathbf{x}, \mathbf{x}_t)$  is kernel function,  $b^{ij}$  is bias value.

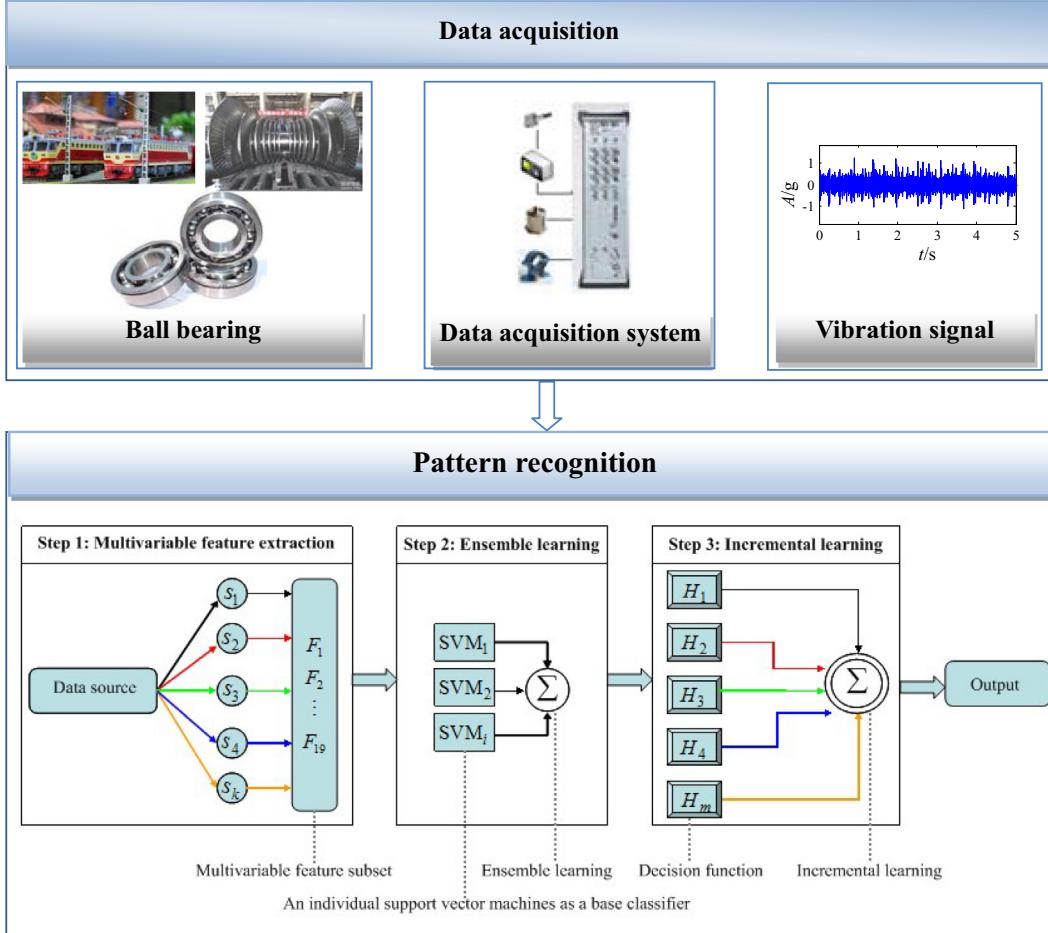
After all the  $k(k - 1)/2$  classifiers are constructed, the classification decision of the one-against-one support vector machine is made using the following strategy: if  $f^{ij}(\mathbf{x})$  says sample  $\mathbf{x}$  is in the  $i$ th class, then the vote for the  $i$ th class is added by one. Otherwise, the  $j$ th is increased with one. After being tested with the  $k(k - 1)/2$  classifiers respectively,  $\mathbf{x}$  belongs to the class which has maximal votes.

## 3. Multivariable ensemble-based incremental support vector machine (MEISVM)

The ensemble-based incremental learning approach is firstly proposed by Parikh et al. for data fusion application [24]. Inspired by the approach, multivariable ensemble-based incremental support vector machine is proposed for intelligent fault diagnosis of roller bearing, which mainly contains three parts as illustrated in Fig. 1: multivariable feature extraction, ensemble learning and incremental learning.

### 3.1. Multivariable feature extraction

Identifying significant variables or extracting fault features from large amount of measured sensory information is challenging and has been a focal point in the fault diagnosis domain. It is desirable that variables and features extracted from sensors are sensitive to machinery faults and robust to the varying machinery running conditions and background noise [9]. So there has been a lot of signal processing approach to obtain desirable features for machinery fault diagnosis, among which Fast Fourier Transform (FFT) is one of the most widely used and well-established methods. When a fault occurs, new frequency components may appear and a change of the convergence of frequency spectrum may take place.



**Fig. 1.** The illustration of the proposed intelligent fault diagnosis with multivariable ensemble-based incremental support vector machine.

**Table 1**  
Statistical features in the time domain and frequency domain.

Feature 1	Feature 2	Feature 3
$F_1 = \frac{\sqrt{\frac{1}{N} \sum_{n=1}^N x(n)^2}}{\frac{1}{N} \sum_{n=1}^N  x(n) }$	$F_2 = \frac{\max x(n) }{\sqrt{\frac{1}{N} \sum_{n=1}^N x(n)^2}}$	$F_3 = \frac{\max x(n) }{\frac{1}{N} \sum_{n=1}^N  x(n) }$
Feature 4	Feature 5	Feature 6
$F_4 = \frac{\max x(n) }{(\frac{1}{N} \sum_{n=1}^N \sqrt{ x(n) })^2}$	$F_5 = \frac{\frac{1}{N} \sum_{n=1}^N (x(n)-\bar{x})^4}{\left(\sqrt{\frac{1}{N} \sum_{n=1}^N (x(n)-\bar{x})^2}\right)^4}$	$F_6 = \frac{\frac{1}{N} \sum_{n=1}^N (x(n)-\bar{x})^3}{\left(\sqrt{\frac{1}{N} \sum_{n=1}^N (x(n)-\bar{x})^2}\right)^3}$
Feature 7	Feature 8	Feature 9
$F_7 = \frac{\sum_{k=1}^K s(k)}{K}$	$F_8 = \frac{\sum_{k=1}^K (s(k)-F_7)^2}{K-1}$	$F_9 = \frac{\sum_{k=1}^K (s(k)-F_7)^3}{K \sqrt{F_8^3}}$
Feature 10	Feature 11	Feature 12
$F_{10} = \frac{\sum_{k=1}^K (s(k)-F_7)^4}{K F_8^2}$	$F_{11} = \frac{\sum_{k=1}^K f_k s(k)}{\sum_{k=1}^K s(k)}$	$F_{12} = \sqrt{\frac{\sum_{k=1}^K (f_k - F_{11})^2 s(k)}{K}}$
Feature 13	Feature 14	Feature 15
$F_{13} = \sqrt{\frac{\sum_{k=1}^K f_k^2 s(k)}{\sum_{k=1}^K s(k)}}$	$F_{14} = \sqrt{\frac{\sum_{k=1}^K f_k^4 s(k)}{\sum_{k=1}^K f_k^2 s(k)}}$	$F_{15} = \frac{\sum_{k=1}^K f_k^2 s(k)}{\sqrt{\sum_{k=1}^K s(k) \sum_{k=1}^K f_k^4 s(k)}}$
Feature 16	Feature 17	Feature 18
$F_{16} = \frac{F_{12}}{F_{11}}$	$F_{17} = \frac{\sum_{k=1}^K (f_k - F_{11})^3 s(k)}{K F_{12}^2}$	$F_{18} = \frac{\sum_{k=1}^K (f_k - F_{11})^4 s(k)}{K F_{12}^4}$
Feature 19	where $x(n)$ represents a signal series for $n = 1, 2, \dots, N$ . $N$ denotes the number of data points; $s(k)$ is a spectrum for $k = 1, 2, \dots, K$ . $K$ represents the number of spectrum lines; $f_k$ denotes the frequency value of the $k$ th spectrum line.	
$F_{19} = \frac{\sum_{k=1}^K (f_k - F_{11})^{1/2} s(k)}{K \sqrt{F_{12}}}$		

As a result, the amplitude and the distribution of frequency spectrum are likely to exhibit different characteristics in frequency-domain. Statistical variables extracted from the time and frequency domain provide information on the defects and the evolution of their values indicates the level of aggravation of defects. It is shown that statistical variables are robust to varying operational conditions and are good indicators of sharp impulses generated by the contact of a defect with the bearing mating surfaces. Multiple variables extracted from the measured sensory information are shown in Table 1.

The variables  $F_1 \sim F_6$  are time domain statistical features. Feature  $F_1$  is the shape factor. Feature  $F_2$  is the crest factor. Feature  $F_3$  is the impulse factor. Feature  $F_4$  is the margin factor. Feature  $F_5$  is the kurtosis factor. Feature  $F_6$  is the skewness factor. And the variables  $F_7 \sim F_{19}$  are frequency domain statistical

**Table 2**  
Bearing information.

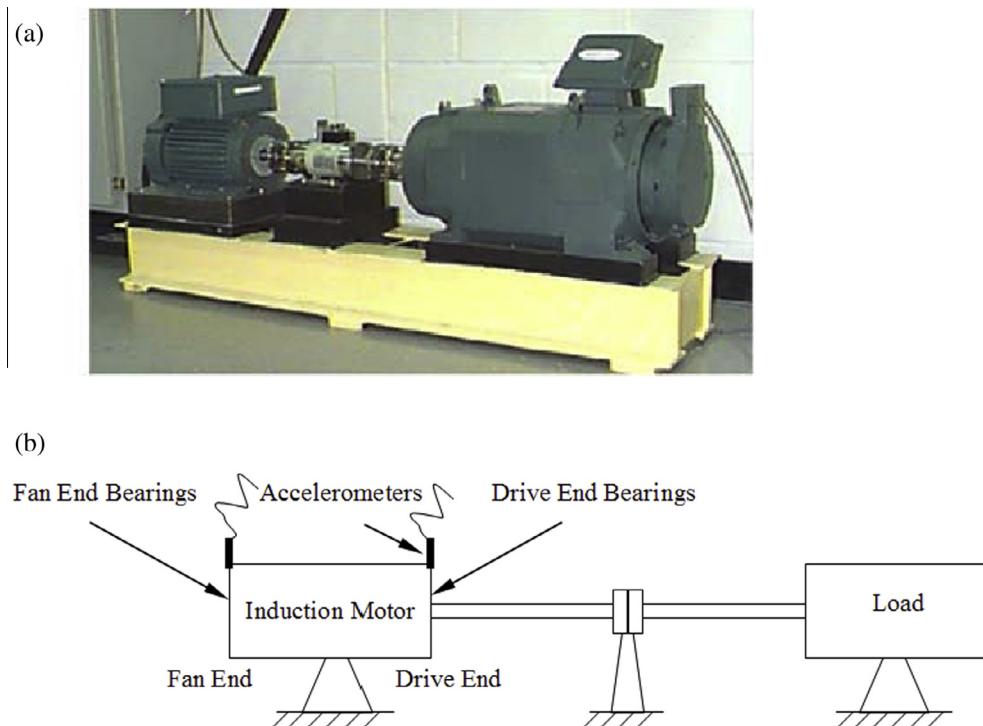
Bearing type	Drive-end bearing	Fan-end bearing
Bearing specs	6205-2RS JEM SKF, deep groove ball bearing	6203-2RS JEM SKF, deep groove ball bearing
Inside diameter (size: in.)	0.9843	0.6693
Outside diameter (size: in.)	2.0472	1.5748
Thickness (size: in.)	0.5906	0.4724
Ball diameter (size: in.)	0.3126	0.2656
Pitch diameter (size: in.)	1.537	1.122

variables. Variable  $F_7$  indicates the vibration energy in frequency domain. Variables  $F_8 \sim F_{10}$ ,  $F_{12}$  and  $F_{16} \sim F_{19}$  describe the convergence of frequency spectrum power. Variables  $F_{11}$  and  $F_{13} \sim F_{15}$  show the position change of main frequencies [31].

The training dataset and testing dataset are respectively constructed from signals reflecting different operational conditions of roller bearing, which assures the independent requirement of information. Therefore, the training dataset and testing dataset are independently constructed. When training process or testing process is conducted, each training sample or testing sample is respectively computed to extract variables  $F_1 \sim F_{19}$  according to formulas in Table 1.

### 3.2. Ensemble learning

Recently, the ensemble based methods have been applied successfully in machine learning fields to improve the performance of an individual support vector machine. The idea of the ensemble support vector machine has been proposed in Ref. [19], in which the boosting technique is used to train each individual support vector machine and takes another support vector machine to combine several support vector machine. Assuming that there is an ensemble of  $n$  individual support vector machine  $\{f_1, f_2, \dots, f_n\}$ , then the ensemble performance will be as the same as the individual classifiers if all of the individual support vector machines are



**Fig. 2.** The test bench of the roller bearing in the motor driven mechanical system: (a) an experiment picture, (b) an schematic illustration.

**Table 3**  
Fault specifications of the bearings.

Bearing	Fault location	Defect diameter (size: in.)	Defect depth (size: in.)	Approx. motor speed (r/min)	Motor load (hp)
–	Normal	–	–	1797/1772/1750/1730	0/1/2/3
Drive-end bearings	Outer Raceway	0.007	0.011	1797/1772/1750/1730	0/1/2/3
		0.014	0.011	1797/1772/1750/1730	0/1/2/3
		0.021	0.011	1797/1772/1750/1730	0/1/2/3
	Inner Raceway	0.007	0.011	1797/1772/1750/1730	0/1/2/3
		0.014	0.011	1797/1772/1750/1730	0/1/2/3
		0.021	0.011	1797/1772/1750/1730	0/1/2/3
	Ball	0.007	0.011	1797/1772/1750/1730	0/1/2/3
		0.014	0.011	1797/1772/1750/1730	0/1/2/3
		0.021	0.011	1797/1772/1750/1730	0/1/2/3
Fan-end bearings	Outer Raceway	0.007	0.011	1797/1772/1750/1730	0/1/2/3
		0.014	0.011	1797/1772/1750/1730	0/1/2/3
		0.021	0.011	1797/1772/1750/1730	0/1/2/3
	Inner Raceway	0.007	0.011	1797/1772/1750/1730	0/1/2/3
		0.014	0.011	1797/1772/1750/1730	0/1/2/3
		0.021	0.011	1797/1772/1750/1730	0/1/2/3
	Ball	0.007	0.011	1797/1772/1750/1730	0/1/2/3
		0.014	0.011	1797/1772/1750/1730	0/1/2/3
		0.021	0.011	1797/1772/1750/1730	0/1/2/3

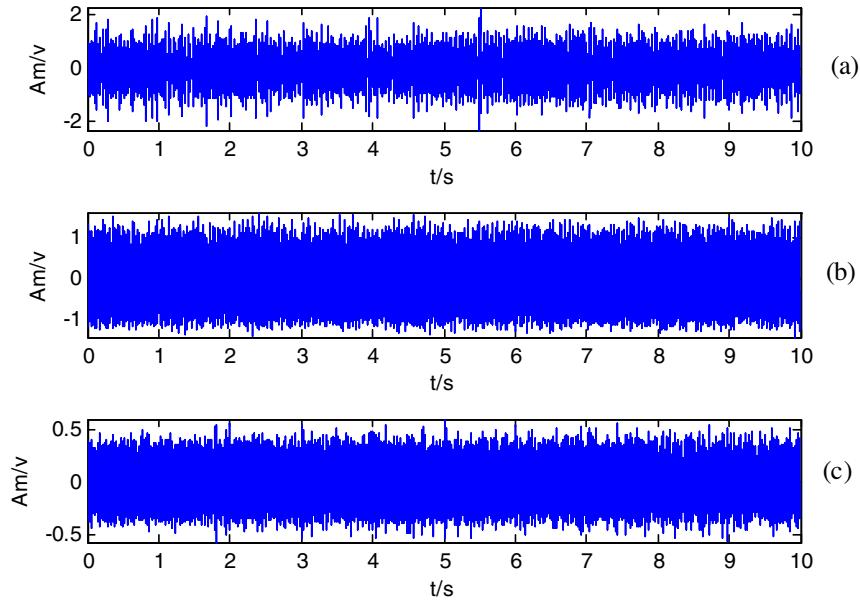
identical. If the individual support vector machines are different and their errors are irrelevant, most of other support vector machines may be correct except for support vector machine  $f_i(\mathbf{x})$ . More precisely, supposing that the error of an individual support vector machine is  $p < 1/2$  since the error of random guessing for a

binary class problem is  $1/2$ , the error of the ensemble by majority voting is

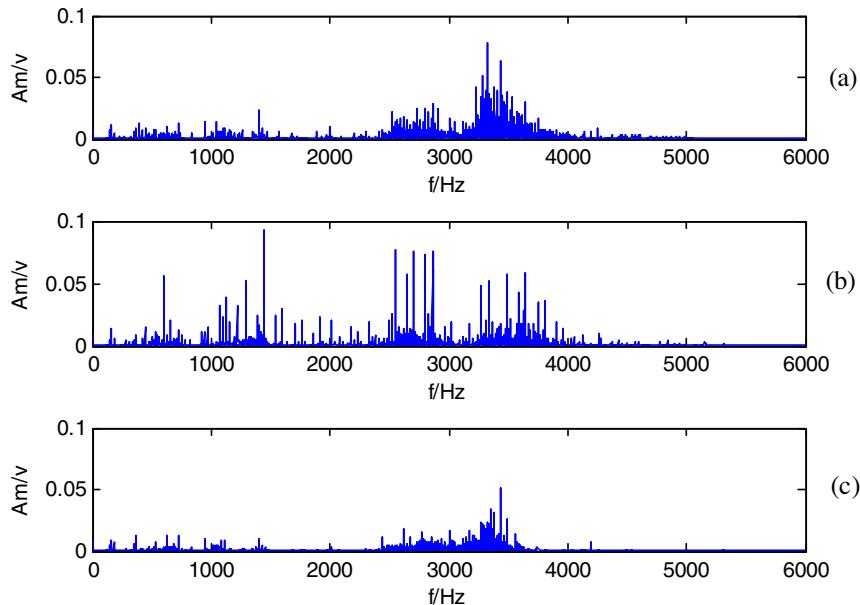
$$P_E = \sum_{k=\lceil n/2 \rceil}^n p^k (1-p)^{n-k} \quad (5)$$

**Table 4**  
Experimental data description.

Bearing	Fault location	Defect diameter (size: in.)	Approx. motor speed (r/min)	Motor load (hp)	State label
Drive-end bearings	Outer Raceway	0.007	1750	2	1
	Inner Raceway	0.007	1750	2	2
	Ball	0.007	1750	2	3



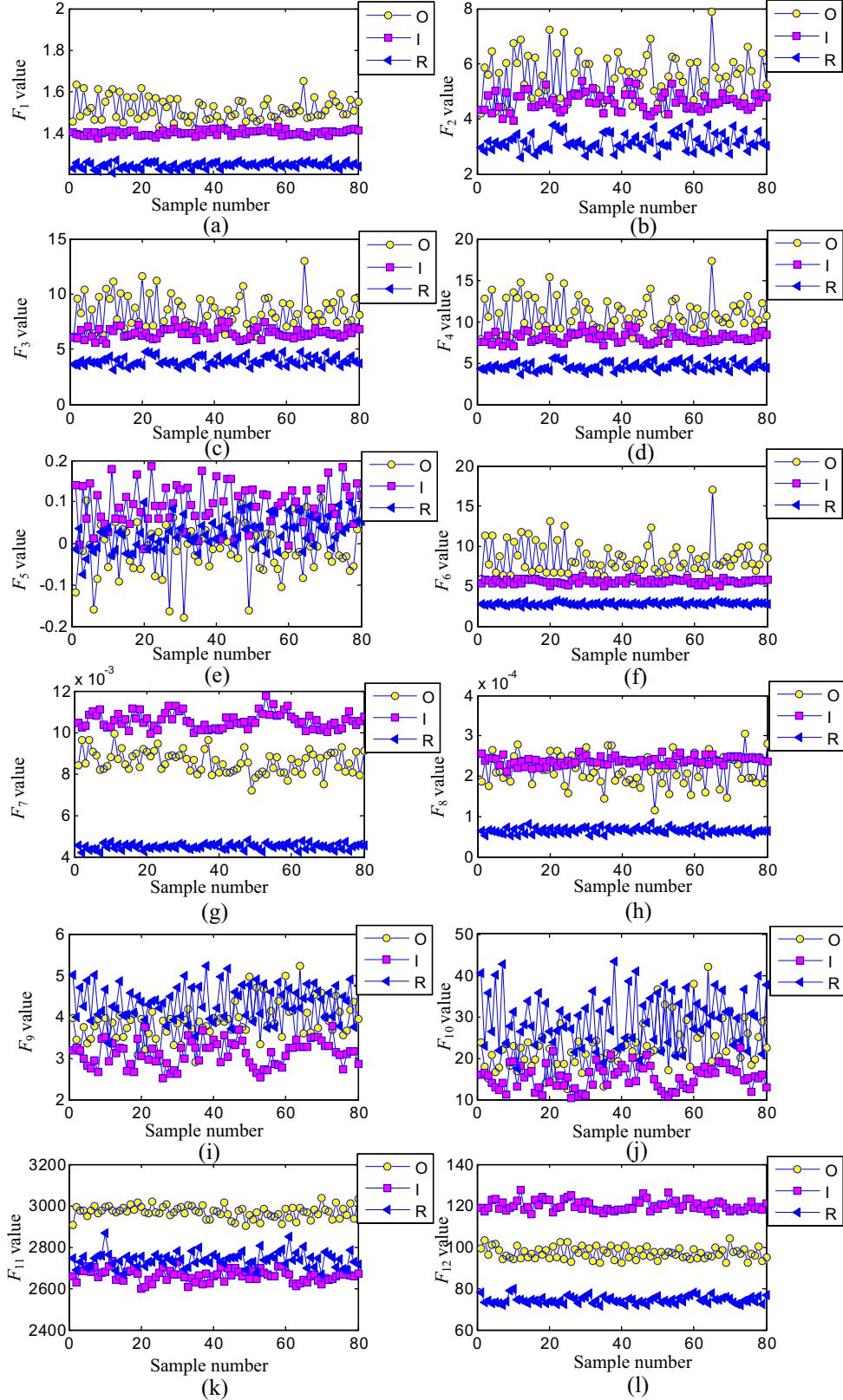
**Fig. 3.** Vibration signals of the roller bearings in three fault conditions: (a) outer race fault with 0.007 in., (b) inner race fault with 0.007 in., (c) ball fault with 0.007 in.



**Fig. 4.** Frequency spectrum of signals in three fault conditions: (a) outer race fault with 0.007 in., (b) inner race fault with 0.007 in., (c) ball fault with 0.007 in.

Since  $p < 1/2$ ,  $1 - p < 1$ ,  $P_E < \sum_{k=[n/2]}^n (1/2)^k (1)^{(n-k)}$ , thus  $P_E < \sum_{k=[n/2]}^n (1/2)^k$ . When the size of support vector machine  $n$  is large,  $P_E$  becomes very small. Because a support vector machine that

makes entirely random guessing has the error equal to  $1/2$ , the error of the final ensemble hypothesis will drop exponentially fast if the weak hypotheses that are slightly better than random guessing can be consistently found [32].



**Fig. 5.** Nineteen statistical features ( $F_1 \sim F_{19}$ ) of the vibration signals in three fault conditions (O: outer race fault, I: inner race fault, R: ball fault).

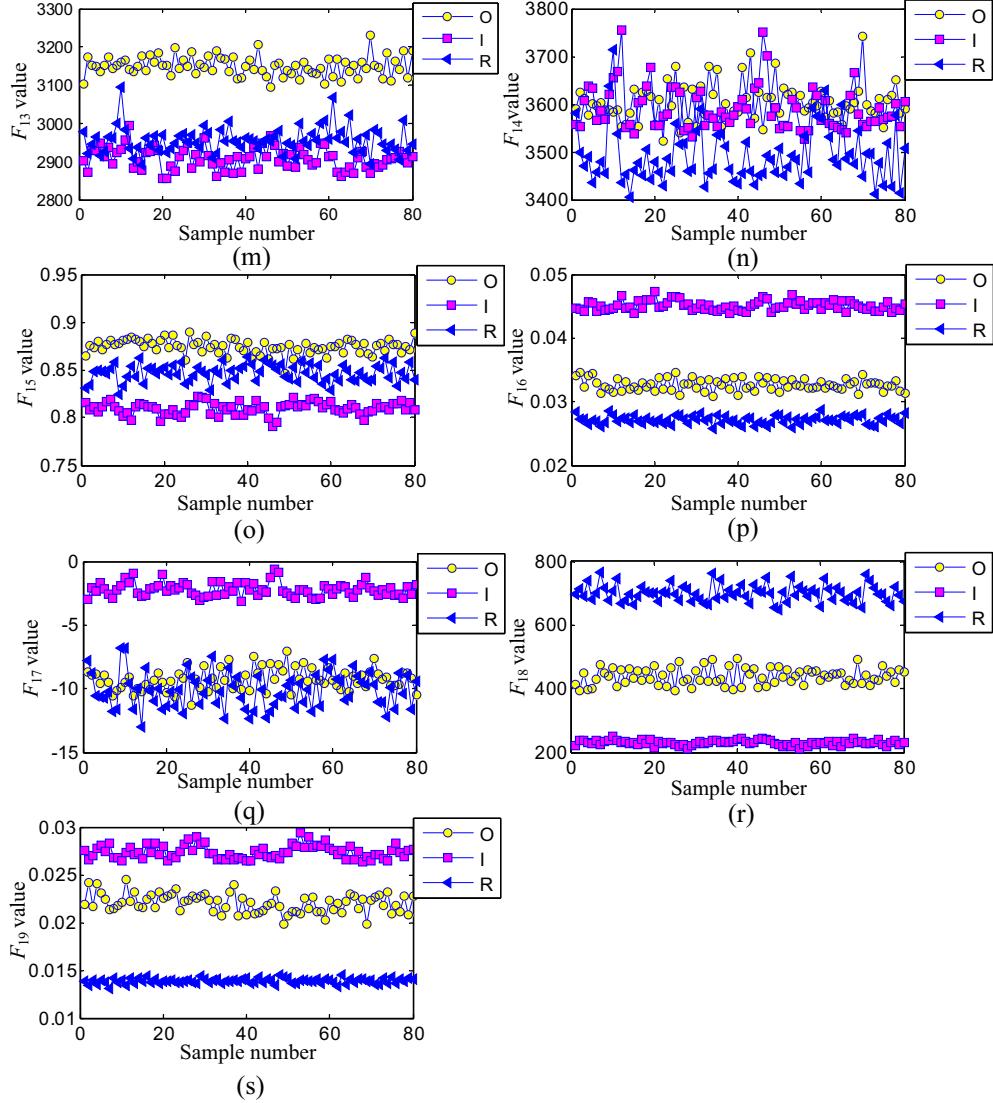


Fig. 5 (continued)

### 3.3. Ensemble-based incremental learning

Each support vector machine is thought as a hypotheses  $h$  from the input space  $X$  to the output space  $Y$ . For each iteration  $t = 1, 2, \dots, T_k$ , samples in dataset  $S_k$  ( $k = 1, 2, \dots, n$ ) is divided into a training subset  $TR_t$  and a testing subset  $TE_t$  according to the current distribution  $D_t$  ( $t = 1, 2, \dots, T_k$ ), and an individual support vector machine is adopted to generate a hypothesis  $h_t : X \rightarrow Y$  using training subset  $TR_t$ . The distribution  $D_t$  is obtained by normalizing a set of weights assigned to each sample based on the performance of the individual support vector machine on the sample. In general, samples that are difficult to classify receive higher weights so as to increase their chances of being selected into next training dataset. The distribution  $D_1$  for the first iteration is initialized to  $1/m$  ( $m$  is the total number of the samples in the dataset  $S_k$ ), giving equal likelihood to each sample to be selected into the first training subset, unless there is sufficient reason to initialize otherwise. The error of  $h_t$  on dataset  $S_k = TR_t + TE_t$  is defined as

$$\varepsilon_t = \sum_{i:h_t(\mathbf{x}_i) \neq y_i} D_t(i) \quad (6)$$

The error  $\varepsilon_t$  is simply the sum of distribution weights of misclassified instances. If  $\varepsilon_t > 1/2$ , the current hypotheses  $h_t$  is discarded and a new training subset  $TR_t$  and a new testing subset  $TE_t$  are constituted. That is, an individual support vector machine is only expected to achieve a 50% (or better) empirical classification accuracy on the dataset  $S_k$ . If  $\varepsilon_t < 1/2$  is satisfied, then the normalized error  $\beta_t$  is computed as

$$\beta_t = \frac{\varepsilon_t}{1 - \varepsilon_t} \quad (7)$$

All hypotheses generated in previous  $t$  iterations are then combined using weighted majority voting. The voting weights are computed as the logarithms of the reciprocals of normalized errors  $\beta$ . Therefore, those hypotheses that perform well on their own training and testing data are given larger voting powers. A classification decision is then made based on the combined outputs of individual hypotheses, which constitutes the composite hypotheses  $H_t$

$$H_t = \arg \max_{y \in Y} \sum_{t:h_t(\mathbf{x})=y} \log \frac{1}{\beta_t} \quad (8)$$

Note that  $H_t$  decides on the class that receives the highest total vote from all hypotheses. The composite error made by  $H_t$  is then computed as

$$E_t = \sum_{i:H_t(\mathbf{x}_i) \neq y_i} D_t(i) = \sum_{i=1}^m D_t(i)[|H_t(\mathbf{x}_i) \neq y_i|] \quad (9)$$

where  $[\cdot]$  is 1 if the predicate is true, and 0 otherwise. If  $E_t > 1/2$ , the current  $h_t$  is discarded, a new training subset is constituted and a new  $h_t$  is generated. It is noted that  $E_t$  can only exceed this threshold during an immediate iteration after a new database  $S_{k+1}$  is

introduced. At all other times,  $E_t < 1/2$  will be satisfied, since all hypotheses  $h_t$  that make up the composite hypotheses have already been verified in formula (8) to achieve a minimum of 50% accuracy on the dataset  $S_k$ . If  $E_t < 1/2$ , composite normalized error is computed as

$$B_t = \frac{E_t}{1 - E_t} \quad (10)$$

The weight  $\omega_t(i)$  is then updated to compute the next distribution  $D_{t+1}$ , which is in turn used in selecting the next training and testing subsets  $TR_{t+1}$  and  $TE_{t+1}$ , respectively.

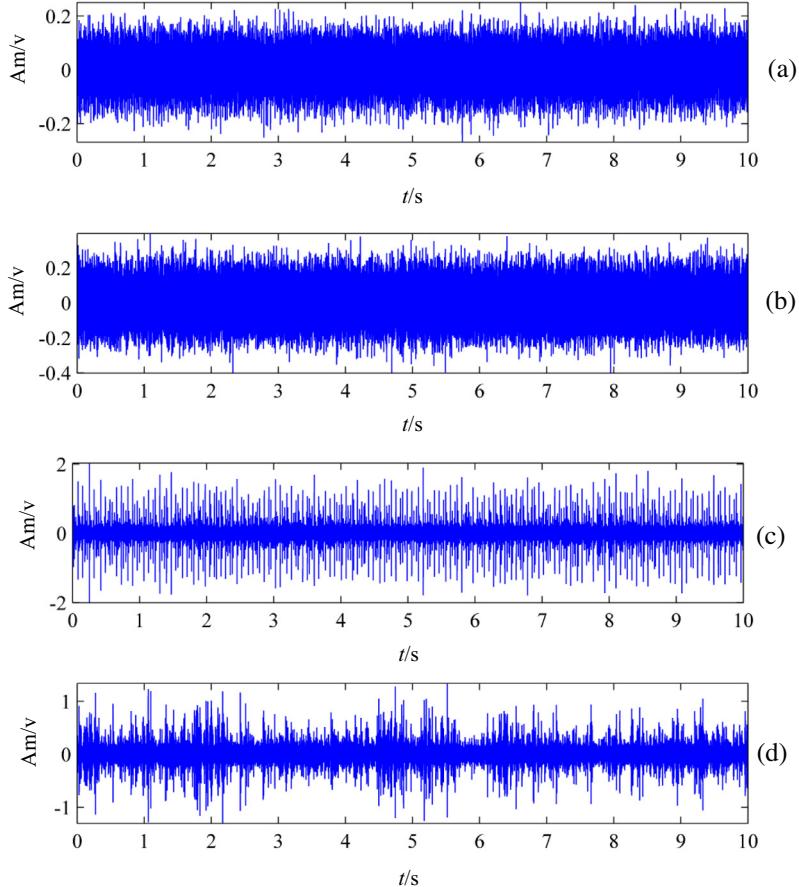
**Table 5**

Experimental results of the proposed intelligent fault diagnosis method with the multivariable ensemble-based incremental support vector machine in contrast to the results of the seven methods mentioned in Ref. [34].

Bearing condition	Method	Accuracy (%)		CPU time
		Mean	Standard deviation	
1 Outer Raceway defect	Discrete Cosine Transform	85	–	–
2 Inner Raceway defect	Daubechies wavelet	78	–	–
3 Ball defect	Symlets wavelet	74	–	–
	Walsh transform	78	–	–
	Fast Fourier Transform	84	–	–
	Walsh-Rough set theory	80	–	–
	Fast Fourier Transform –Rough set theory	86	–	–
	SVM <sup>a</sup> (10-fold cross validation)	100	0	60.18
	SVM (Ant colony optimization)	100	0	114.87
	MEISVM <sup>b</sup> (10-fold cross validation)	100	0	70.63
	MEISVM (Ant colony optimization)	100	0	125.39

<sup>a</sup> SVM: support vector machine.

<sup>b</sup> MEISVM: multivariable ensemble-based incremental support vector machine.



**Fig. 6.** Vibration signals of the bearings in four conditions: (a) normal, (b) outer race fault with 0.014 in., (c) inner race fault with 0.014 in., (d) ball fault with 0.014 in.

$$\begin{aligned} \omega_{t+1}(i) &= \omega_t(i) \times \begin{cases} B_t, & \text{if } H_t(\mathbf{x}_i) = y_i \\ 1, & \text{otherwise} \end{cases} \\ &= \omega_t(i) \times B_t^{1 - [H_t(\mathbf{x}_i) \neq y_i]} \end{aligned} \quad (11)$$

According to the rule, if a sample  $\mathbf{x}_i$  is correctly classified by the composite hypotheses  $H_t$ , its weight is multiplied by a factor of  $B_t$  which is less than 1. If  $\mathbf{x}_i$  is misclassified, its distribution weight is kept unchanged. The rule reduces the probability of correctly classified samples being chosen into  $TR_{t+1}$ , while increasing the probability of misclassified samples to be selected into  $TR_{t+1}$ . The multivariable ensemble-based incremental support vector machine algorithm focuses more on samples that are repeatedly misclassified, especially for the new samples belonging to a previously unseen data.

After  $T_k$  hypotheses are generated for each dataset  $S_k$ , the final hypotheses is obtained by the weighted majority voting of all composite hypotheses

$$H_{final}(\mathbf{x}) = \arg \max_{y \in Y} \sum_{k=1}^K \left( \sum_{t: H_t(\mathbf{x})=y} \log \left( \frac{1}{\beta_t} \right) \right) \quad (12)$$

**Table 6**  
Experimental data description.

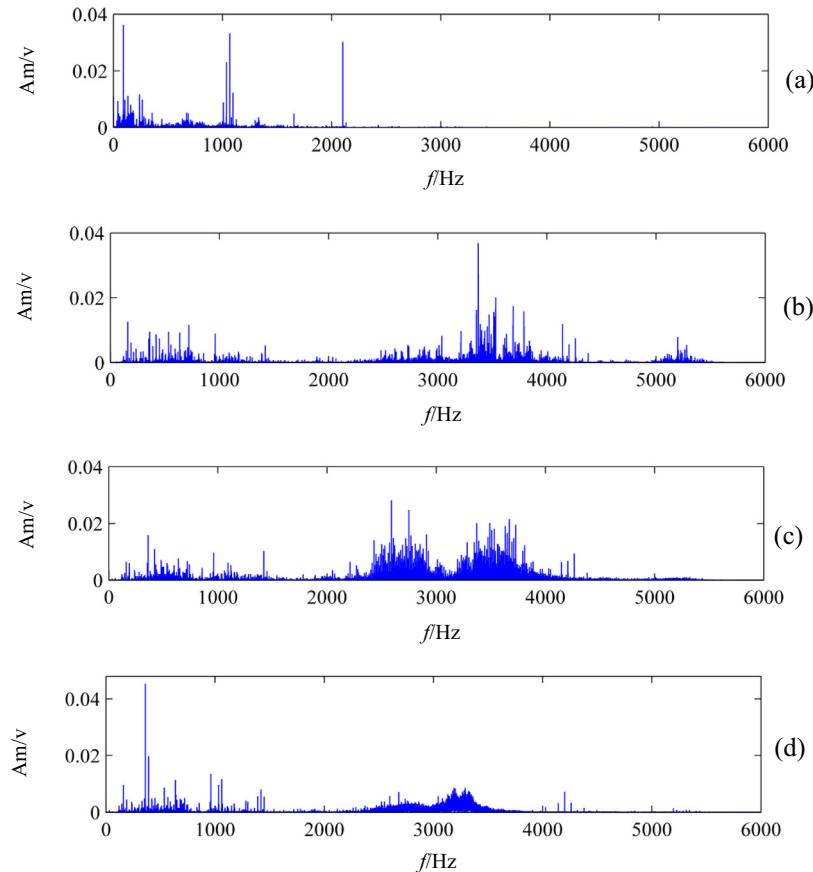
Bearing	Fault location	Defect diameter (in.)	Approx. motor speed (r/min)	Motor load (HP)	State label
Drive-end bearings	Normal	–	1772	1	1
	Outer Raceway	0.014	1772	1	2
	Inner Raceway	0.014	1772	1	3
	Ball	0.014	1772	1	4

#### 4. Intelligent fault diagnosis experiments for roller bearings with multivariable ensemble-based incremental support vector machine

In order to test the effectiveness of the proposed intelligent fault diagnosis method for roller bearings with multivariable ensemble-based incremental support vector machine, experiments are conducted on the dataset of the Bearing Data Center of Case Western Reverse University (CWRU) [33] which have been widely used to testify theory and application of fault diagnosis method in many papers. Experiments of the proposed method and detailed comparisons with some other methods are respectively conducted as follows.

##### 4.1. Experimental system description

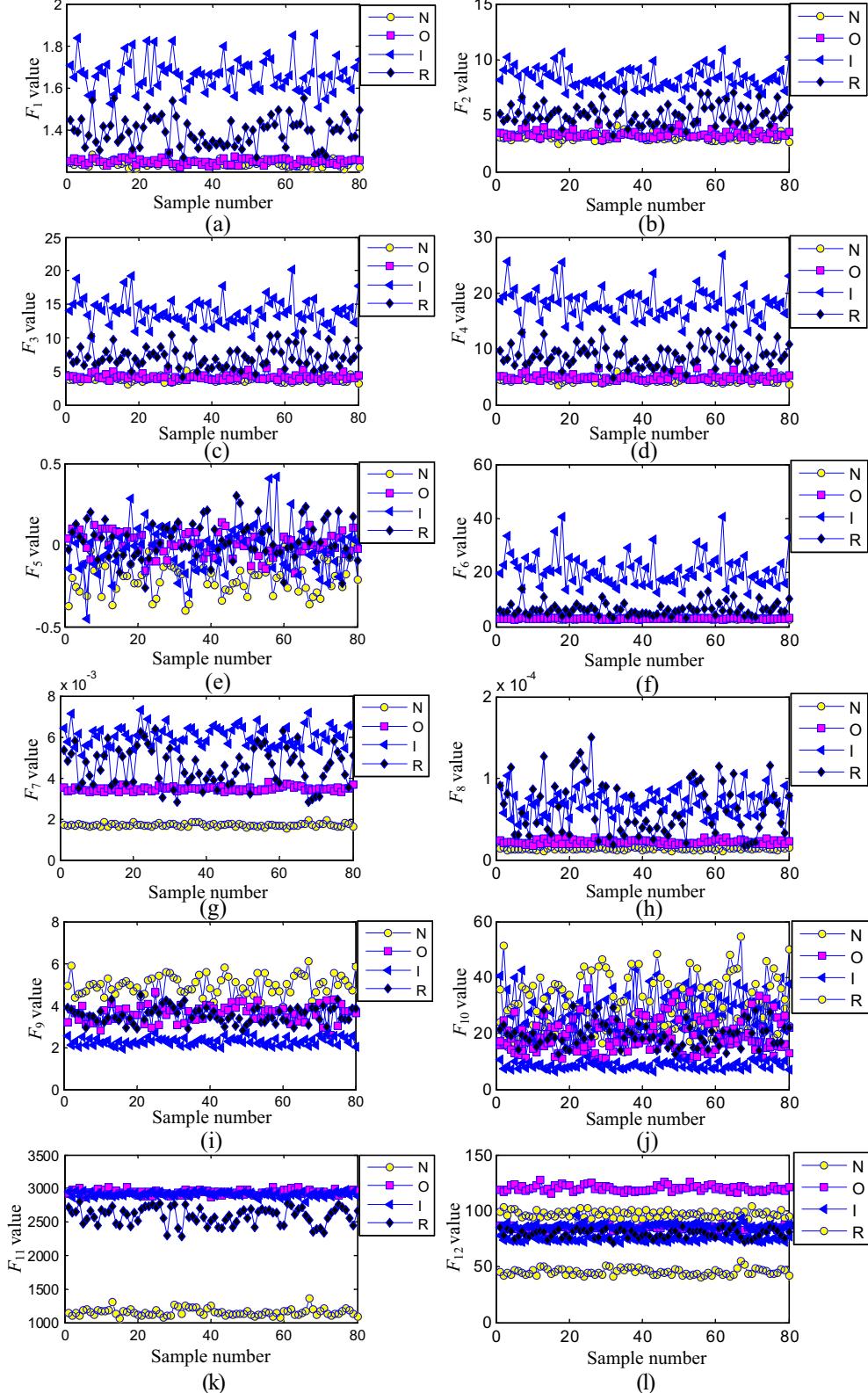
The roller bearing test stand of the experiment is illustrated in Fig. 2, which consists of a 2hp motor (left), a torque transducer/encoder (center), a dynamometer (right). The test bearings including the drive-end bearings and the fan-end bearings are



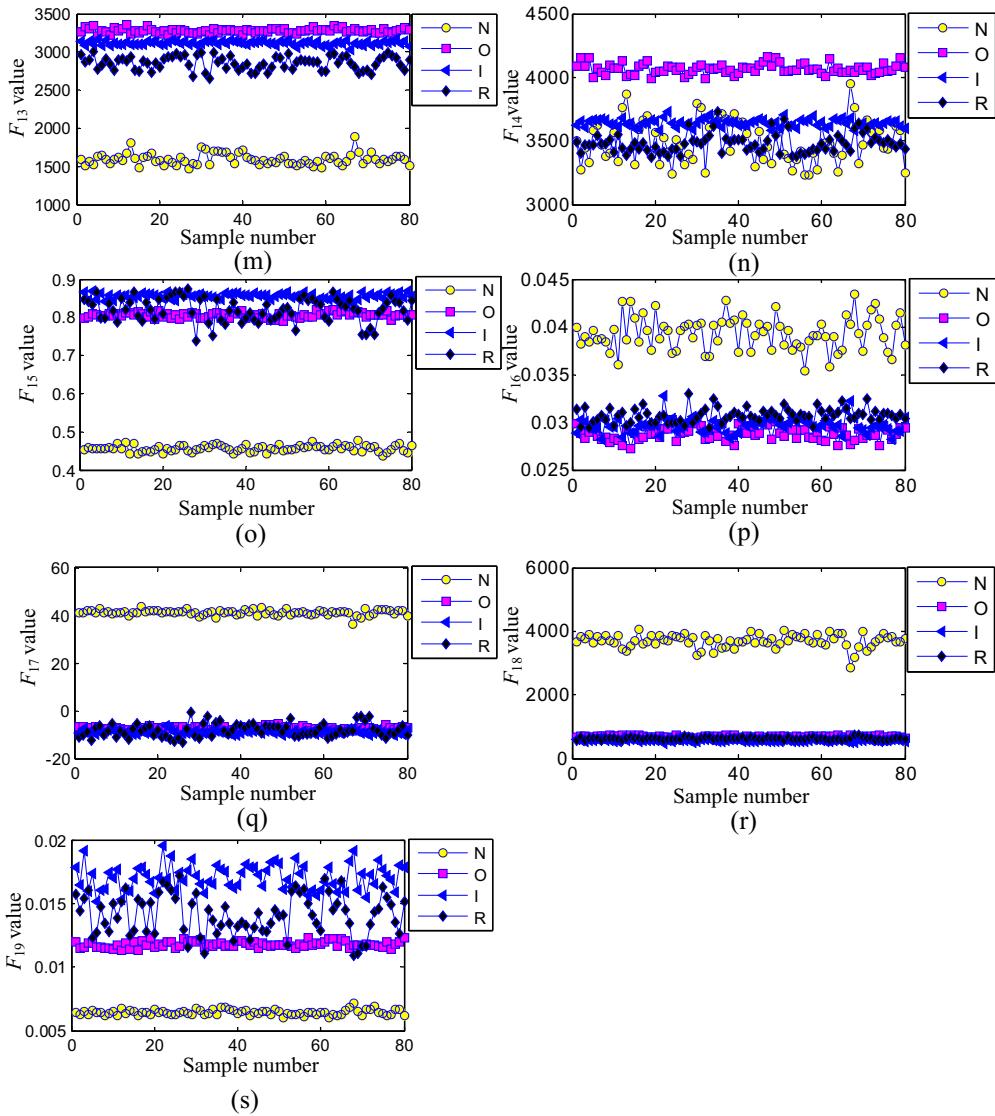
**Fig. 7.** Frequency spectrum of signals in four fault conditions: (a) normal, (b) outer race fault with 0.014 in., (c) inner race fault with 0.014 in., (d) ball fault with 0.014 in.

installed in the test stand, which support the motor shaft. The information of the test bearings is listed in **Table 2**. Single point faults are processed to the test bearings using electro-discharge machining with fault diameters of 0.007, 0.014, 0.021 in., and the fault depth is 0.011 in. Fault specifications of the drive-end bearings and fan-end bearings are given in **Table 3**, which include

different fault locations with different defect diameter and different rotating speed (motor load). Two accelerometers are respectively mounted on the motor housing at drive-end and fan-end of the motor to acquire vibration signals of the test bearings under different states. The data collection system consists of a high bandwidth amplifier particularly designed for the vibration



**Fig. 8.** Nineteen statistical features ( $F_1 \sim F_{19}$ ) of the vibration signals in four conditions (N: normal, O: outer race fault, I: inner race fault, R: ball fault).

**Fig. 8 (continued)****Table 7**

Experimental results of the proposed intelligent fault diagnosis method with the multivariable ensemble-based incremental support vector machine in contrast to the results of the four methods mentioned in Ref. [35].

Bearing condition	Method	Accuracy (%)		CPU time (s)
		Mean	Standard deviation	
1 Normal	Fuzzy ARTMAP with original feature set	79.228	–	–
2 Outer Raceway defect	Fuzzy ARTMAP with optimal feature set	91.185	–	–
3 Inner Raceway defect	Improved fuzzy ARTMAP with original feature set	89.382	–	–
4 Ball defect	Improved fuzzy ARTMAP with optimal feature set	99.541	–	–
	SVM (10-fold cross validation)	87	3.49	74.69
	SVM (Ant colony optimization)	93	4.2164	125.54
	MEISVM (10-fold cross validation)	96	3.1623	83.51
	MEISVM (Ant colony optimization)	98.5	2.4152	147.68

signals and a data recorder with a sampling frequency of 12,000 Hz. The data recorder is equipped with low-pass filters at the input stage for anti-aliasing.

#### 4.2. Experimental analysis and comparisons

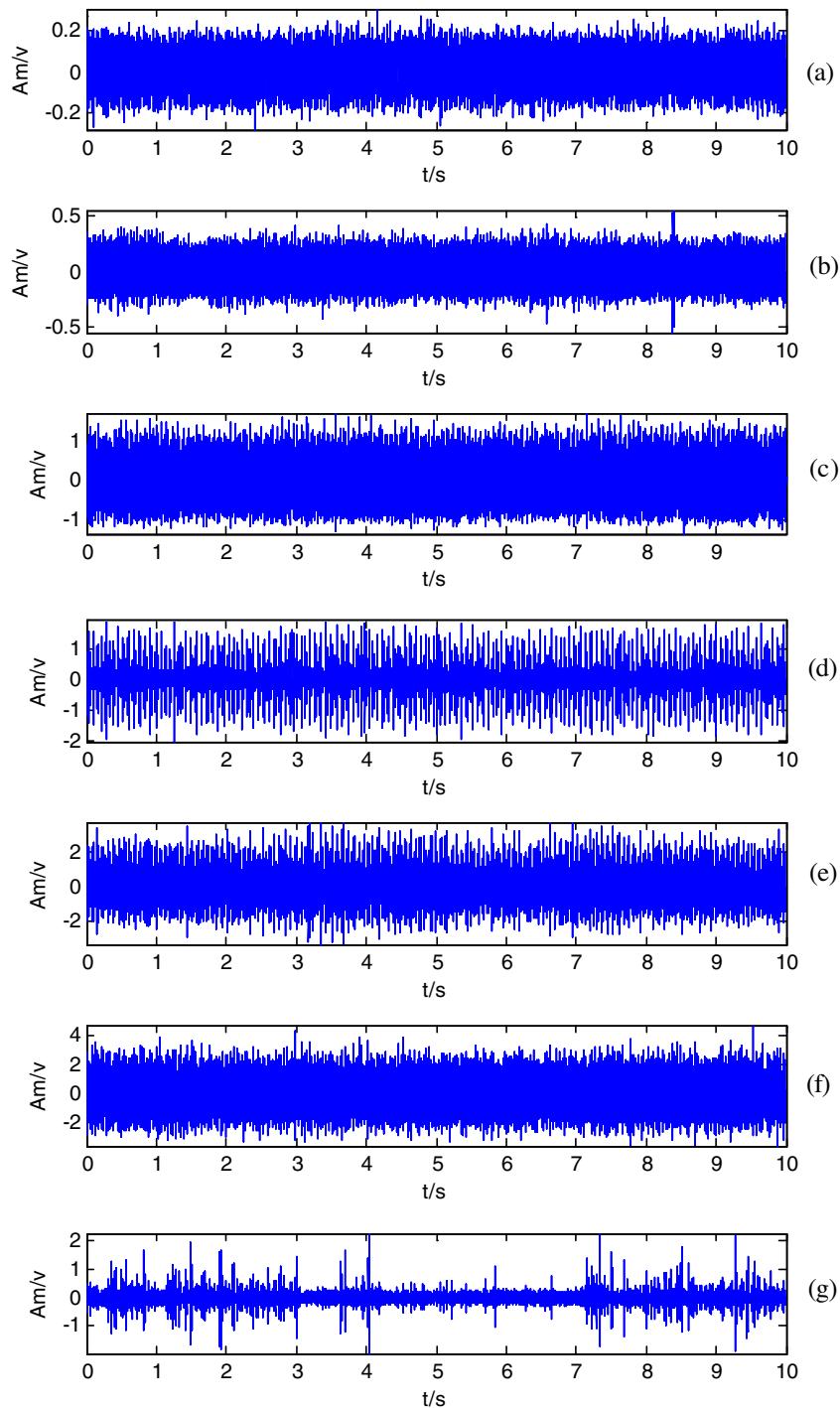
Since the experiment of roller bearing conducted in the Bearing Data Center of Case Western Reverse University is widely used to

testify the fault diagnosis approach of roller bearings, detailed comparisons of the proposed method with some other methods are carried out on the experimental dataset with a same experimental setup. Since the detailed information about the experimental dataset and the specific setup of the experiment are clearly shown and highly cited in Refs. [34–37], experiments are respectively followed the experimental setup as the work they done. Our implementation is carried out in the Matlab of 7.11.0.584

**Table 8**

Experimental data description.

Bearing	Fault location	Defect diameter (size: in.)	Approx. motor speed (r/min)	Motor load (HP)	State label
Drive-end bearings	Normal	–	1772	1	1
	Outer Raceway	0.014	1772	1	2
	Inner Raceway	0.007	1772	1	3
	Inner Raceway	0.014	1772	1	4
	Inner Raceway	0.021	1772	1	5
	Inner Raceway	0.028	1772	1	6
	Ball	0.014	1772	1	7



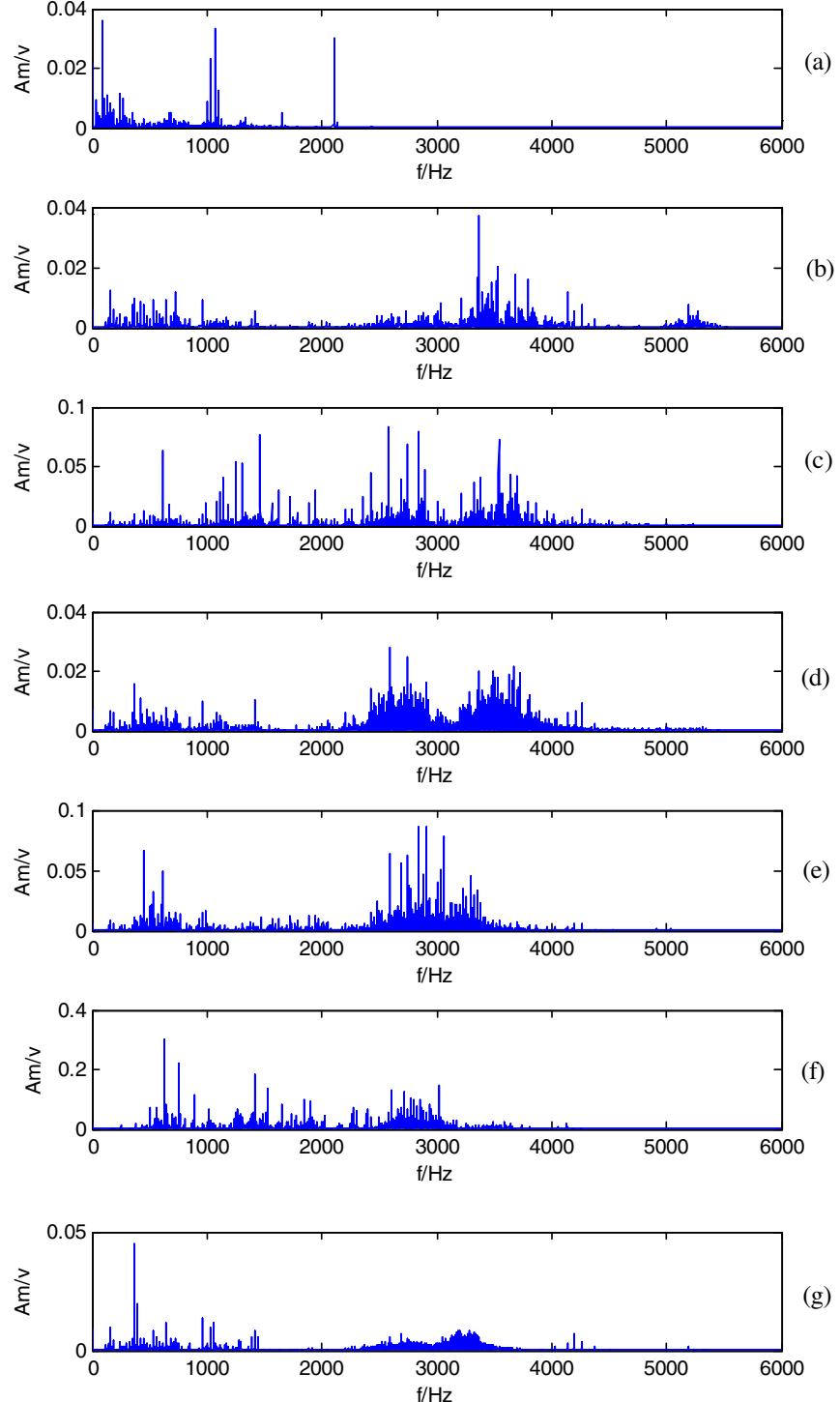
**Fig. 9.** Vibration signals of the bearings in seven conditions: (a) normal, (b) outer race fault with 0.014 in., (c) inner race fault with 0.007 in., (d) inner race fault with 0.014 in., (e) inner race fault with 0.021 in., (f) inner race fault with 0.028 in., (g) ball fault with 0.014 in.

version environment and Windows 7 operating system on Inter(R) core(TM) i3-4150 CPU @ 3.5 GHz Processor running at 8.0 GB RAM.

#### 4.2.1. Case 1: A comparison research of the proposed intelligent fault diagnosis method with multivariable ensemble-based incremental support vector machine to the methods proposed in Ref. [34]

With the same experimental setup as mentioned in Ref. [34], three classes of fault signals including ball fault, inner race fault and outer race fault (Load Zone Centered at 12:00) with the defect

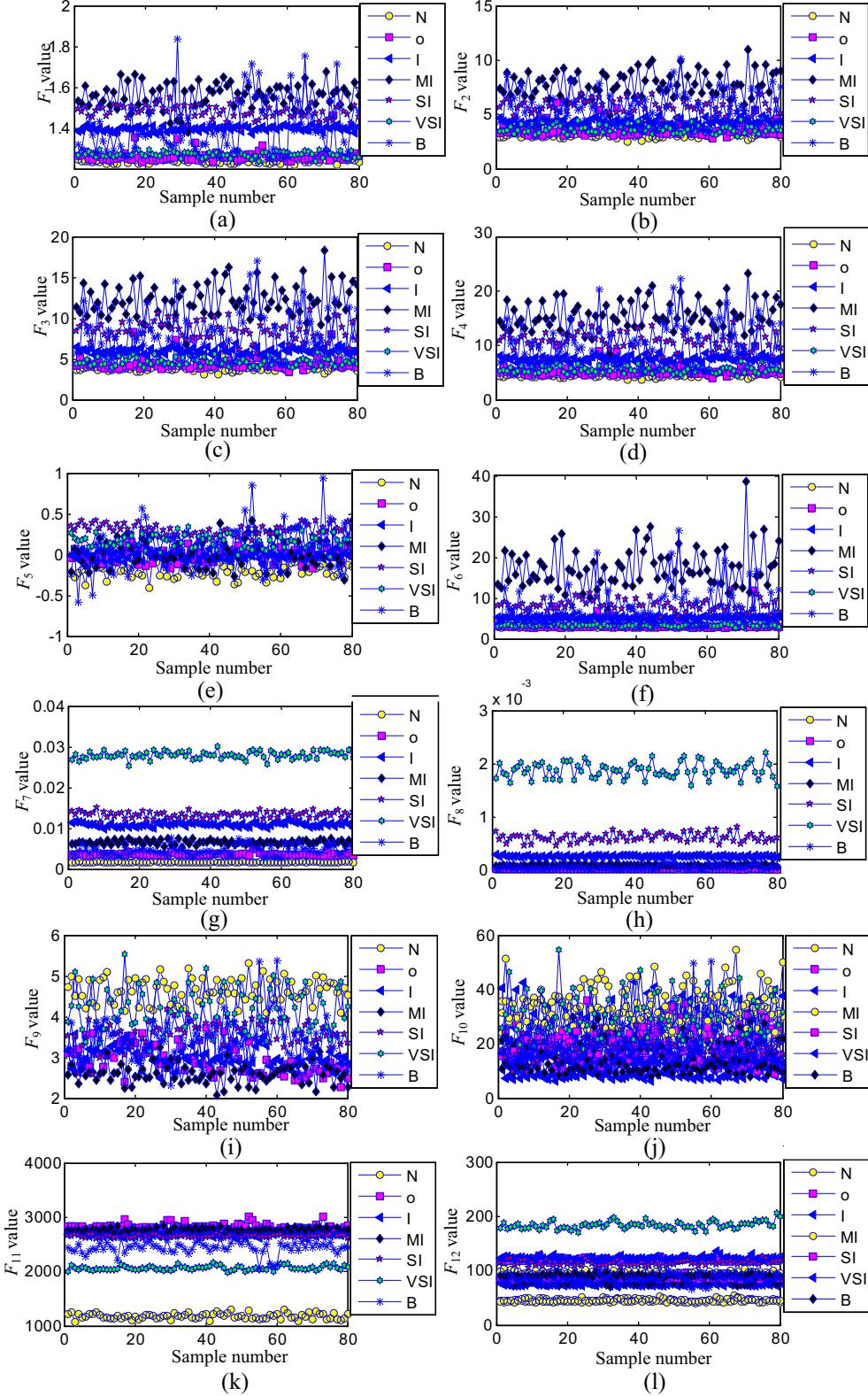
sizes of 0.007 in. at the drive end bearing are collected for experiment, which are described in Table 4. Vibration signals in the time domain are respectively measured for bearing defect on the outer race, inner race and ball shown in Fig. 3. The frequency spectrums of the vibration signals by Fast Fourier Transform in the three states are shown in Fig. 4. The waveforms in time domain and spectrums in frequency domain reveal a few characteristics in three different fault states. For example, the amplitudes of the vibration signal in three fault states are different. The maximum



**Fig. 10.** Frequency spectrum of signals in seven conditions: (a) normal, (b) outer race fault with 0.014 in., (c) inner race fault with 0.007 in., (d) inner race fault with 0.014 in., (e) inner race fault with 0.021 in., (f) inner race fault with 0.028 in., (g) ball fault with 0.014 in.

amplitude of signal in outer race fault is about 2v shown in Fig. 3(a), which is a little larger than inner race fault shown in Fig. 3(b) and ball fault shown in Fig. 3(c). The maximum amplitude of signal in ball fault is the smallest among them, which is about

0.5v. Moreover, spectrum distributions also have some differences in three fault states. For example, some frequency amplitudes in low frequency range are much larger than that in outer race fault shown in Fig. 4(a) and ball fault shown in Fig. 4(c). Thirty-five



**Fig. 11.** Nineteen statistical features ( $F_1 \sim F_{19}$ ) of the vibration signals in different conditions (N: normal, O: outer race fault, I: inner race fault with 0.007 in., MI: inner race fault with 0.014 in., SI: inner race fault with 0.021 in., VSI: inner race fault with 0.028 in., B: ball fault).

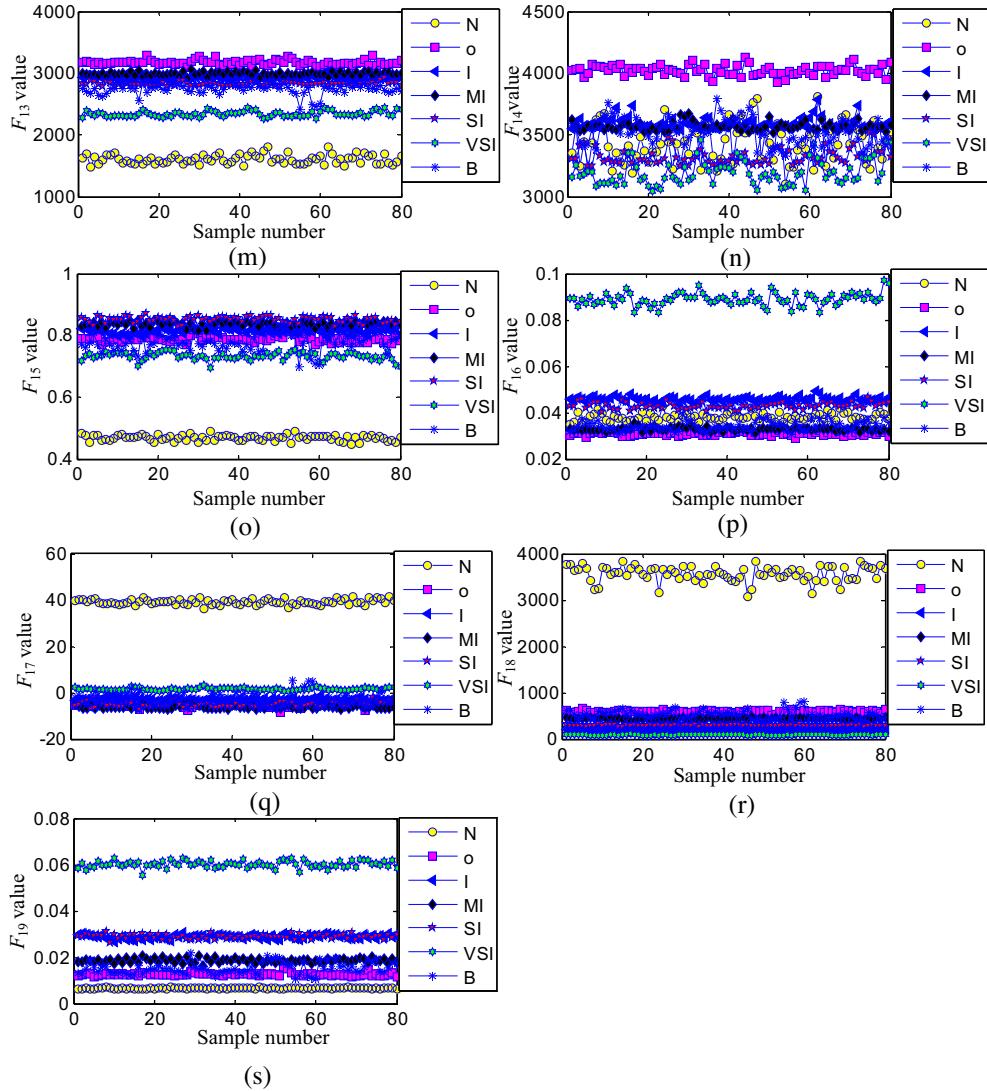


Fig. 11 (continued)

samples in each fault conditions are acquired for training, and fifteen samples in each fault conditions are independently acquired for testing, which are set as the same as other methods in Ref. [34]. Each sample includes 1024 points. Nineteen variables  $F_1 \sim F_{19}$  of each sample are respectively computed according to formulas in Table 1, which are shown in Fig. 5. From Fig. 5(a) to (s), feature  $F_1$ ~feature  $F_{19}$  of three vibration signals under three different fault conditions are respectively revealed, which are more or less irregular. For example, each feature characteristic of eighty samples is changed unsteadily and rapid

fluctuated. Features  $F_7$ ,  $F_{12}$ ,  $F_{16}$ , and  $F_{18}$ – $F_{19}$  among three different fault conditions are quite different which is useful for classification.

Gaussian kernel is used. Parameters ( $C$ ,  $\sigma$ ) of support vector machine are selected from [0.5, 50] and 10-fold cross validation method is used. Ant colony algorithm [38,39] is also adopted for parameters optimization of  $C$  and  $\sigma$ . Table 5 gives the experimental results in contrast to the results attained by seven other methods mentioned in Ref. [34]. SVM is the abbreviation of support vector machine and MEISVM is the abbreviation of the proposed

**Table 9**

Experimental results of the proposed intelligent fault diagnosis method with the multivariable ensemble-based incremental support vector machine in contrast to the methods proposed in Ref. [36].

Bearing condition (Defect size: in.)	Method	Accuracy (%)		CPU time (s)
		Mean	Standard deviation	
1 Normal	Modified fuzzy ARTMAP with original feature set	77.551	–	–
2 Outer race fault (0.014)	Modified fuzzy ARTMAP with optimal feature set	84.898	–	–
3 Inner race fault (0.014)	Modified fuzzy ARTMAP and feature-weight learning	87.302	–	–
4 Inner race fault (0.0007)	SVM (10-fold cross validation)	85.1786	6.9442	85.32
5 Inner race fault (0.021)	SVM (Ant colony optimization)	89.4643	3.0872	132.47
6 Inner race fault (0.028)	MEISVM (10-fold cross validation)	92.1429	2.3810	92.19
7 Ball fault (0.014)	MEISVM (Ant colony optimization)	96.4214	1.9196	155.52

intelligent fault diagnosis method with multivariable ensemble-based incremental support vector machine, which are used in the following Tables. From Table 5, it is seen that the proposed method and typical support vector machine exhibit better performance than other seven methods (Discrete Cosine Transform, Daubechies wavelet, Symlets wavelet, Walsh transform, FFT, Walsh-Rough set theory, FFT-Rough set theory) in identifying three common fault conditions (ball fault, inner race fault and outer race fault), which hit the highest accuracy of 100%. The mean accuracy is 100%, and the standard deviation is zero since three fault conditions are relatively easy to identify. The effectiveness of the proposed method is further testified in the following experiments.

#### 4.2.2. Case 2: A comparison research of the proposed intelligent fault diagnosis method with multivariable ensemble-based incremental support vector machine to the methods proposed in Ref.[35]

Another comparison research with the methods proposed in Ref. [35] is conducted. The normal condition of the roller bearing is added with three fault conditions of the roller bearings: ball

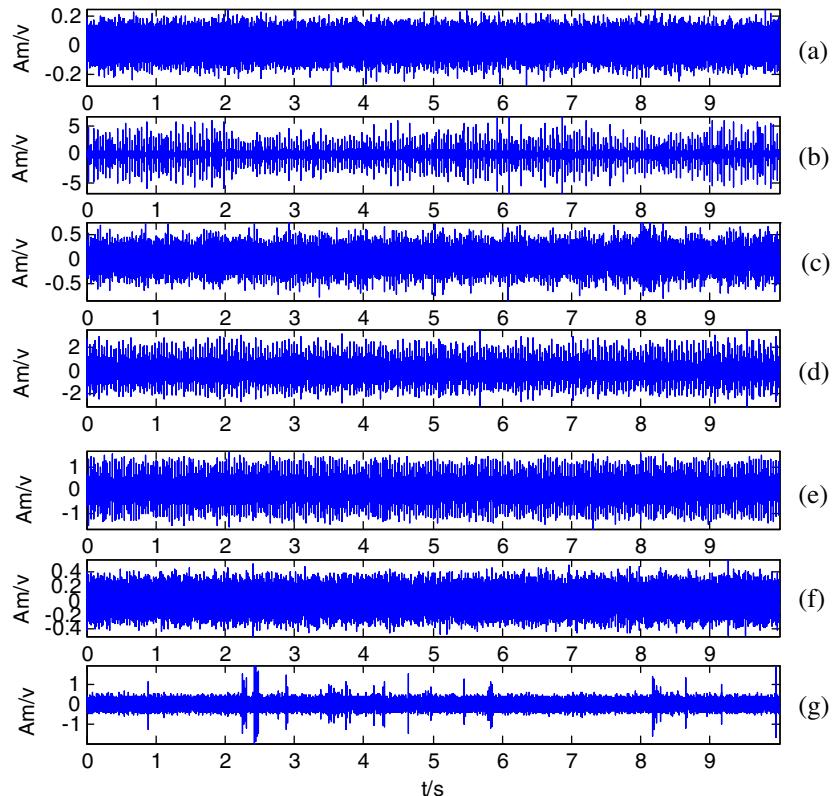
fault, inner race fault and outer race fault. Several differences between current experiment and case 1 are that the motor load and the motor speed are higher and the defect sizes of the ball fault, inner race fault and outer race fault are about 0.014 in. The experimental setup is also as the same as the one mentioned in Ref. [35] in order to justly compare, which are described in Table 6. Vibration signals in time domain are respectively plotted in Fig. 6. The frequency spectrums of the vibration signals in four states are shown in Fig. 7.

When a fault occurs in roller bearing, some impulses occurs and the phenomenon of amplitude modulation happens, such as Fig. 6(c) and (d). However, it is seen that the waveform in normal state shown in Fig. 6(a) and outer race fault shown in Fig. 6(b) are quite similar in time domain, which are hard to classify. The spectrums in frequency domain reflect a few different characteristics in four states as shown in Fig. 7. For example, the amplitudes of the vibration signals in four fault states are mainly concentrated in a frequency band from 2000 Hz to 4000 Hz. But the amplitude of the vibration signal in normal state is mainly concentrated in a low frequency range from 0 to 2000 Hz. Thirty-five samples in each fault conditions are acquired for training, and fifteen samples in each fault conditions are independently acquired for testing, which are set as the same as other methods in Ref. [35]. Each sample includes 1024 points. Nineteen variables  $F_1 \sim F_{19}$  of each sample are respectively computed according to formulas in Table 1, which are shown in Fig. 8. From Fig. 8(a) to (s), feature  $F_1 \sim$  feature  $F_{19}$  of the four different conditions are respectively revealed, which is more or less irregular. Each feature characteristic is changed unsteadily and rapid fluctuated, which are hard to classify them.

Gaussian kernel is used. Parameters ( $C, \sigma$ ) in support vector machine are selected from [0.5, 50] and 10-fold cross validation method is used. Ant colony algorithm is also adopted for parameters optimization of  $C$  and  $\sigma$ . The computed results are shown in

**Table 10**  
Description of bearing dataset in the experiment (size: in.).

Bearing location	Fault type	Defect diameter	Defect depth	Approx. motor speed (r/min)	State label
Drive End	Normal	–	–	1772	1
Drive End	Outer Race fault	0.021	0.011	1772	2
Fan End	Outer Race fault	0.021	0.011	1772	3
Drive End	Inner Race fault	0.021	0.011	1772	4
Fan End	Inner Race fault	0.021	0.011	1772	5
Drive End	Ball fault	0.021	0.011	1772	6
Fan End	Ball fault	0.021	0.011	1772	7



**Fig. 12.** Vibration signals of the bearings in different conditions: (a) normal, (b) outer race fault in drive end bearing, (c) outer race fault in fan end bearing, (d) inner race fault in drive end bearing, (e) inner race fault in fan end bearing, (f) ball fault in drive end bearing, (g) ball fault in fan end bearing.

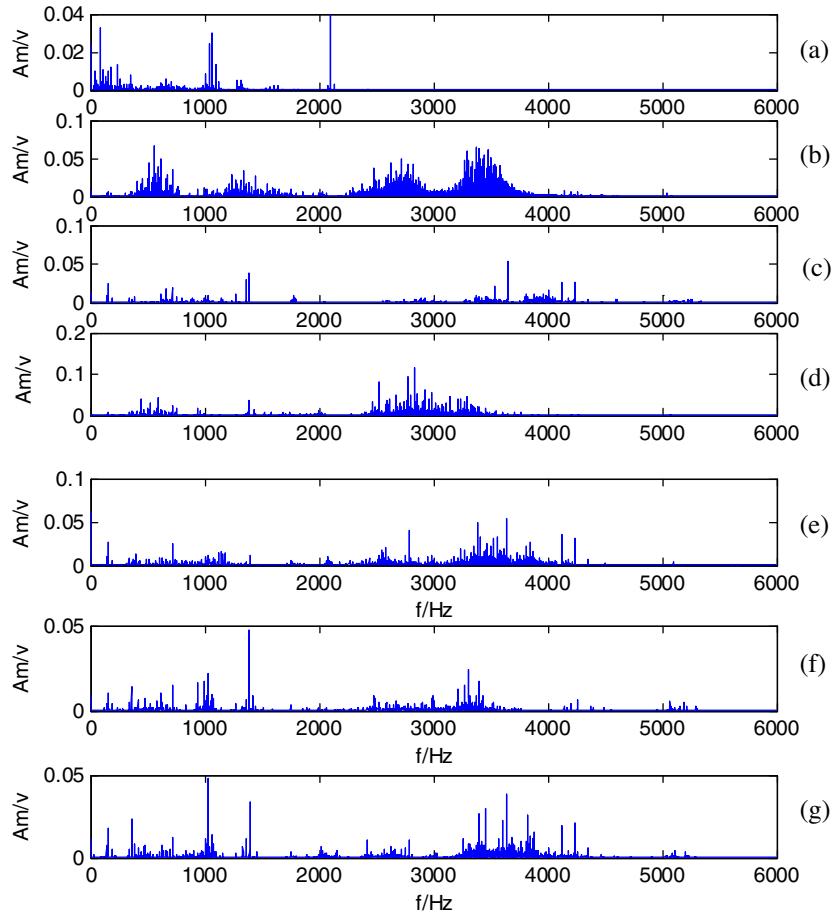
**Table 7** in contrast with other four methods based on Fuzzy ARTMAP. It is seen that the performance of the proposed multivariable ensemble-based incremental support vector machine attains better diagnosis result than the results of typical support vector machine and other three methods based on the Fuzzy ARTMAP mentioned in Ref. [35]. Since the final diagnosis result not only rely on building an effective intelligent model but also rely on inputting optimal features, the performance of the proposed method is a little worse than the performance of the improved fuzzy ARTMAP in which optimal features are processed by modified distance discrimination technique in Ref. [35].

#### 4.2.3. Case 3: A comparison research of the proposed intelligent fault diagnosis method with multivariable ensemble-based incremental support vector machine to the methods proposed in Ref. [36]

Since multiple faults including defects with different severe degrees are hard to detect defecting frequency characteristic, an experiment is conducted on a complex case in order to further investigate the performance of the proposed method: seven different fault conditions of the roller bearings including normal condition, ball defect, outer race defect and inner race defect in different severe degree (slight defect, medium defect, severe defect and very severe defect). The experiment setup is set as the same as the work done in Ref. [36]. A detailed description of the data is shown in **Table 8**. Vibration signals in time domain are respectively measured for seven fault conditions in **Fig. 9**. It is shown that the vibration signals of inner race defect in different severe degree are very similar to each other. The amplitudes of the vibration signal in

normal bearing and outer race fault with 0.014 in. are smaller than others. Frequency spectrums of the vibration signals in seven fault conditions are shown in **Fig. 10**. The frequency components of vibration signal in normal state are mainly concentrated in a low frequency range from 0 to 2000 Hz. But the spectrums of vibration signals in fault states are distributed in a wider frequency band, which are mainly concentrated in a frequency band from 2000 Hz to 4000 Hz. It is seen that both of the signals in time domain and frequency domain are not enough to distinguish fault conditions. Therefore, the proposed method is applied. Forty samples in each condition are acquired for training, and forty samples are independently acquired for testing in contrast to other three methods in Ref. [36] with the same experimental setup. Each sample includes 1024 points. Nineteen variables  $F_1 \sim F_{19}$  of each sample are respectively computed according to formulas in **Table 1**, which are shown in **Fig. 11**. From **Fig. 11(a)** to **(s)**, feature  $F_1 \sim F_{19}$  of different conditions are more or less irregular, which are hard to classify them.

Gaussian kernel is used. Parameters ( $C, \sigma$ ) in support vector machine are selected from [0.5, 50] and 10-fold cross validation method is used. Ant colony algorithm is also adopted for parameters optimization of  $C$  and  $\sigma$ . Comparing with the results mentioned in Ref. [36], **Table 9** gives the experimental results of the proposed method in contrast to support vector machine and other three methods based on fuzzy ARTMAP proposed in Ref. [36]. It is seen that the proposed method attains the best result among other approaches, which proves that the proposed method not only can diagnose different multiple fault patterns (normal, outer race fault,

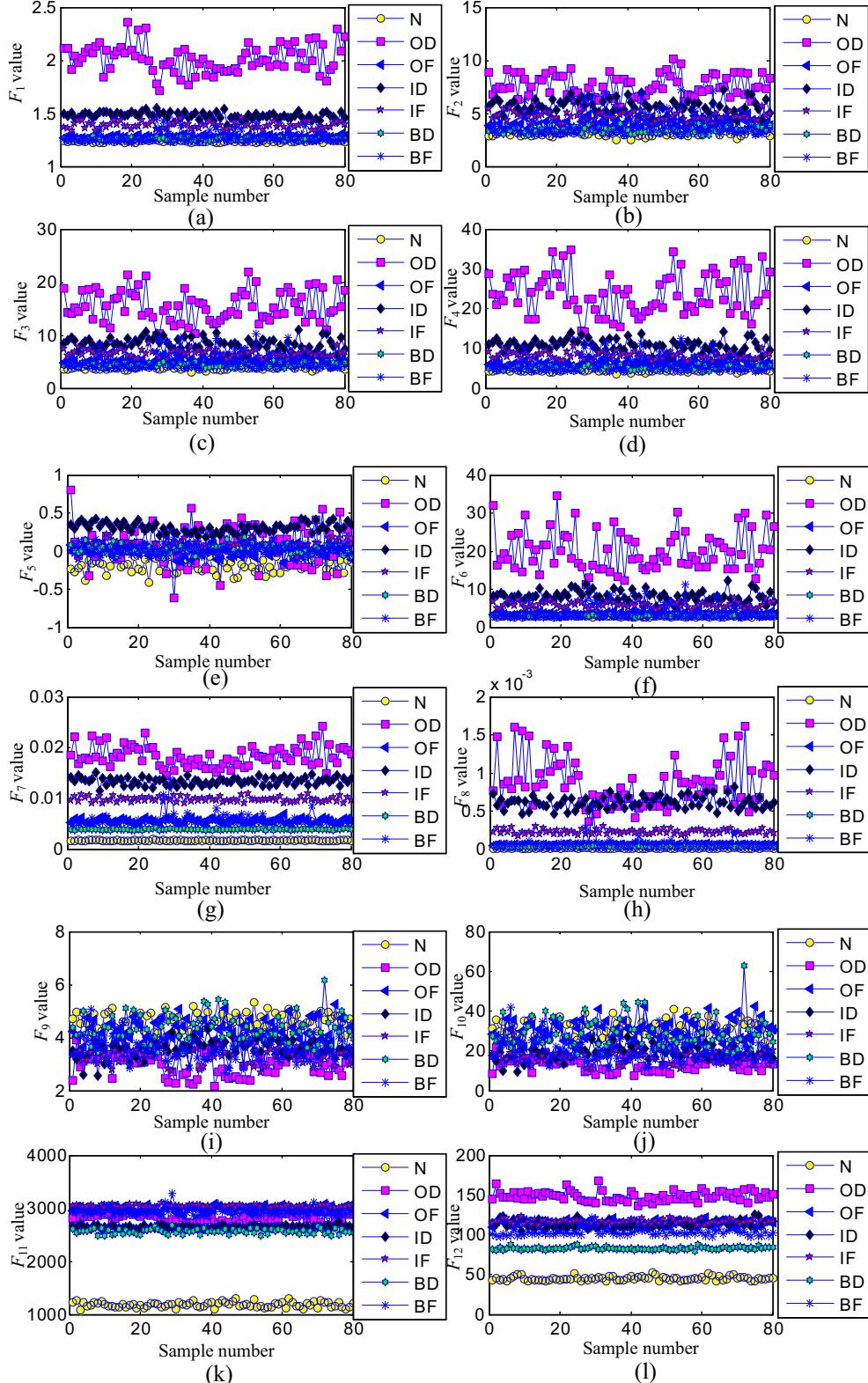


**Fig. 13.** Frequency spectrum of signals in different conditions: (a) normal, (b) outer race fault in drive end bearing, (c) outer race fault in fan end bearing, (d) inner race fault in drive end bearing, (e) inner race fault in fan end bearing, (f) ball fault in drive end bearing, (g) ball fault in fan end bearing.

inner race fault and ball fault), but also can effectively distinguish different severe degrees. But the CPU time of proposed method is longer than typical support vector machine in 10-fold cross validation and ant colony algorithm for parameter optimization, since the computation are more complex.

#### 4.2.4. Case 4: A comparison research of the proposed intelligent fault diagnosis method with multivariable ensemble-based incremental support vector machine to the methods proposed in Ref. [37]

Since multiple faults including defects in different positions are also difficult to classify, experiment is conducted on a complex



**Fig. 14.** Nineteen statistical features ( $F_1 \sim F_{19}$ ) of the vibration signals in different conditions (N: normal, OD: outer race fault in drive end bearing, OF: inner race fault in drive end bearing, ID: inner race fault in fan end bearing, IF: inner race fault in fan end bearing, BD: ball fault in drive end bearing, BF: ball fault in fan end bearing).

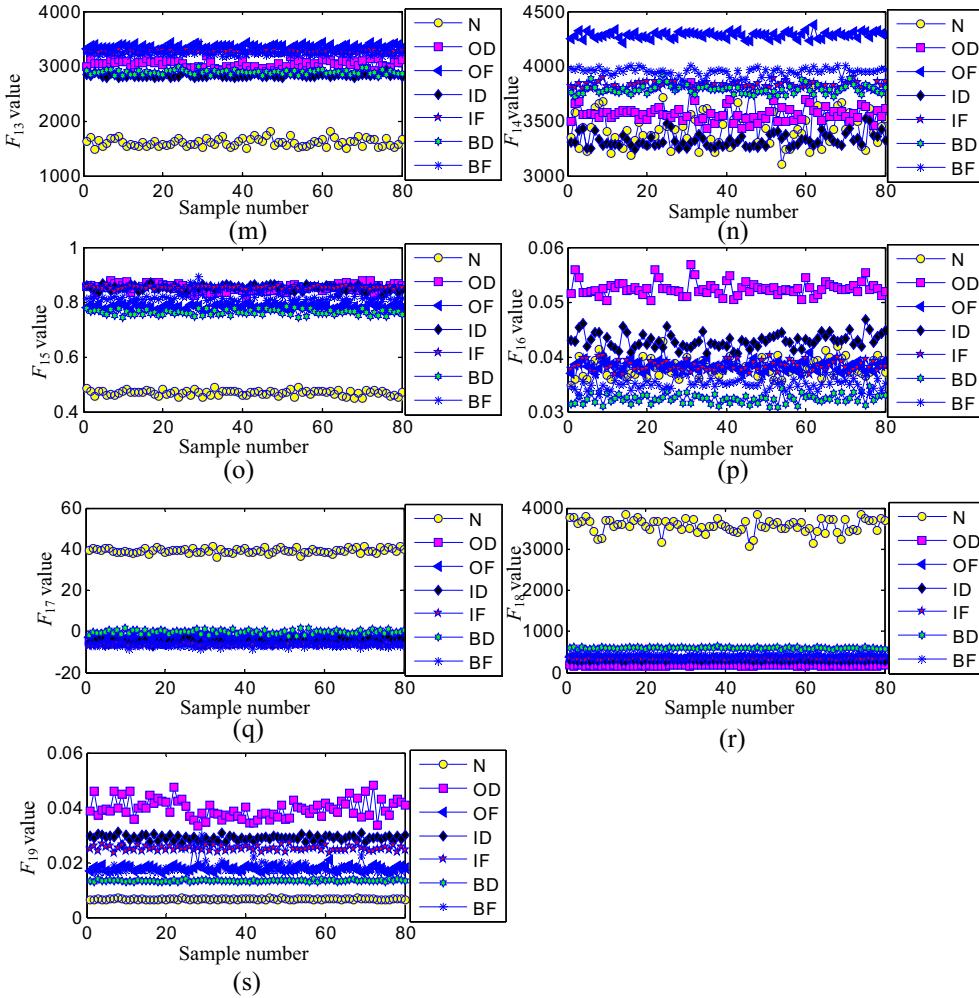


Fig. 14 (continued)

case in order to further investigate the performance of the proposed method: seven different fault conditions of the roller bearings including normal condition, ball defect, outer race defect and inner race defect in drive end bearing and fan end bearing. The experiment setup is set as the same as the work done in Ref.

[37]. A detailed description of the data is shown in Table 10. Vibration signals in time domain are respectively measured for the seven fault conditions in Fig. 12. The frequency spectrum of the vibration signals in seven fault conditions are shown in Fig. 13. It is seen that both of the signals in time domain and spectrums in frequency domain are not enough to distinguish from each other. From Fig. 14(a) to (s), feature  $F_1 \sim F_{19}$  of different conditions are more or less irregular. Each feature characteristic is changed unsteadily and fluctuated. And some features are overlapped, which are hard to classify. Therefore, the proposed method is applied to distinguish different fault conditions in

Table 11

Experimental results of the proposed intelligent fault diagnosis method with the multivariable ensemble-based incremental support vector machine in contrast to the methods proposed in Ref. [37].

Bearing condition	Method	Accuracy (%)		CPU time (s)
		Mean	Standard deviation	
1 Normal	1-NN	96.25	–	–
2 Outer race fault in drive end bearing	MLP	93.9	–	–
3 Outer race fault in fan end bearing	SVM (10-fold cross validation)	97.25	5.2541	84.17
4 Inner race fault in drive end bearing	SVM (Ant colony optimization)	97.65	3.4632	133.56
5 Inner race fault in fan end bearing	MEISVM (10-fold cross validation)	98.13	2.5217	93.29
6 Ball fault in drive end bearing	MEISVM (Ant colony optimization)	99.04	2.0218	158.46
7 Ball fault in fan end bearing				

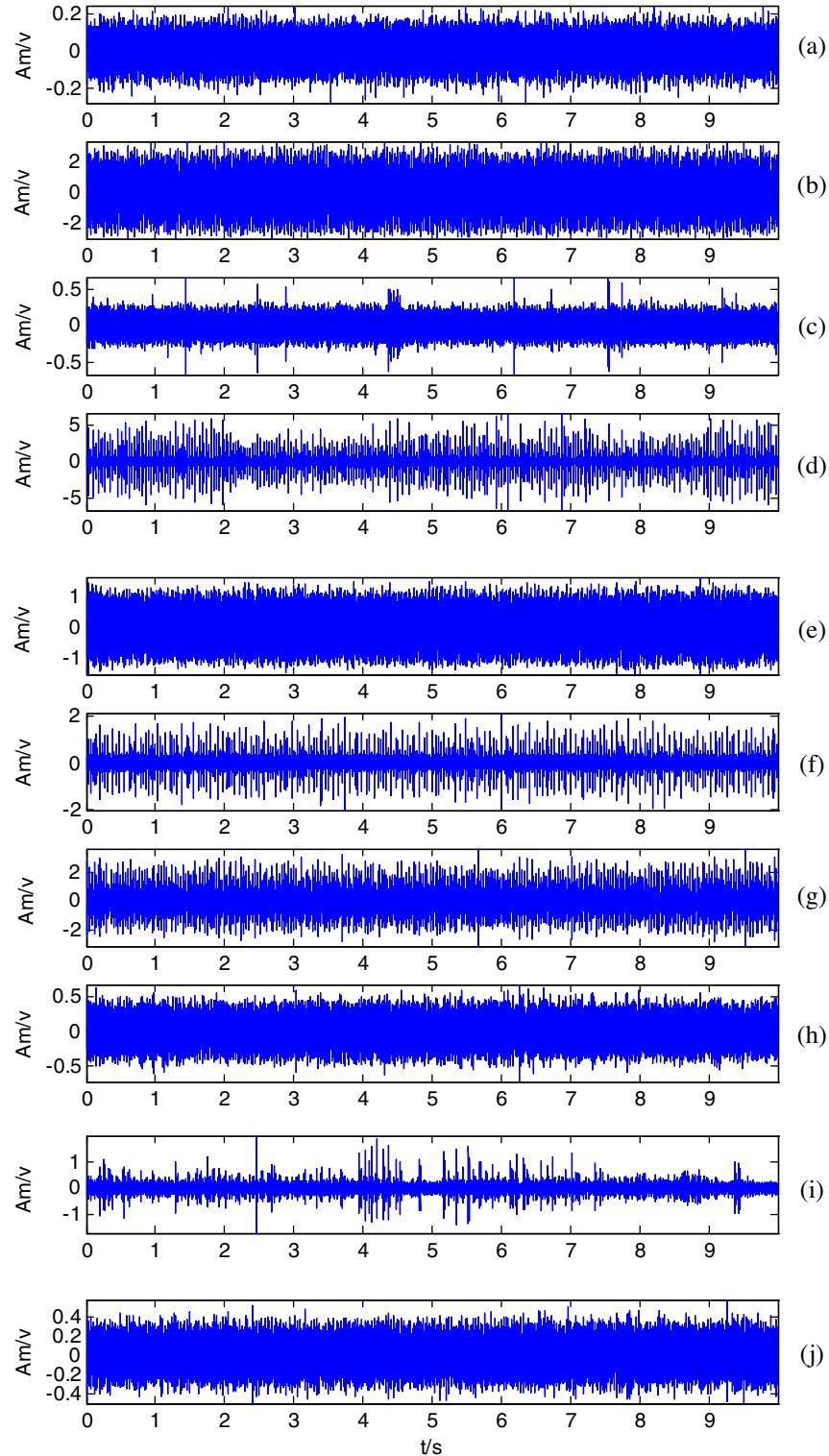
Table 12

Description of bearing dataset in the experiment (size: in.).

Bearing	Fault location	Defect diameter	Defect depth	Approx. motor speed (r/min)	State label
Drive End	Normal	–	–	1772	1
	Outer Raceway	0.007	0.011	1772	2
	Outer Raceway	0.014	0.011	1772	3
	Outer Raceway	0.021	0.011	1772	4
	Inner Raceway	0.007	0.011	1772	5
	Inner Raceway	0.014	0.011	1772	6
	Inner Raceway	0.021	0.011	1772	7
	Ball	0.007	0.011	1772	8
	Ball	0.014	0.011	1772	9
	Ball	0.021	0.011	1772	10

contrast to other methods in Ref. [37] with the same experimental setup. Gaussian kernel is used. Parameters ( $C, \sigma$ ) in support vector machine are selected from [0.5, 50] and 10-fold cross validation method is used. Ant colony algorithm is also adopted for parameters optimization of  $C$  and  $\sigma$ . Table 11 gives the experimental

results. It is seen that the proposed method attains the best result among other approaches, which proves that the proposed method can identify different multiple fault patterns (outer race fault, inner race fault and ball fault) in different position but the CPU time is longer than typical support vector machine.

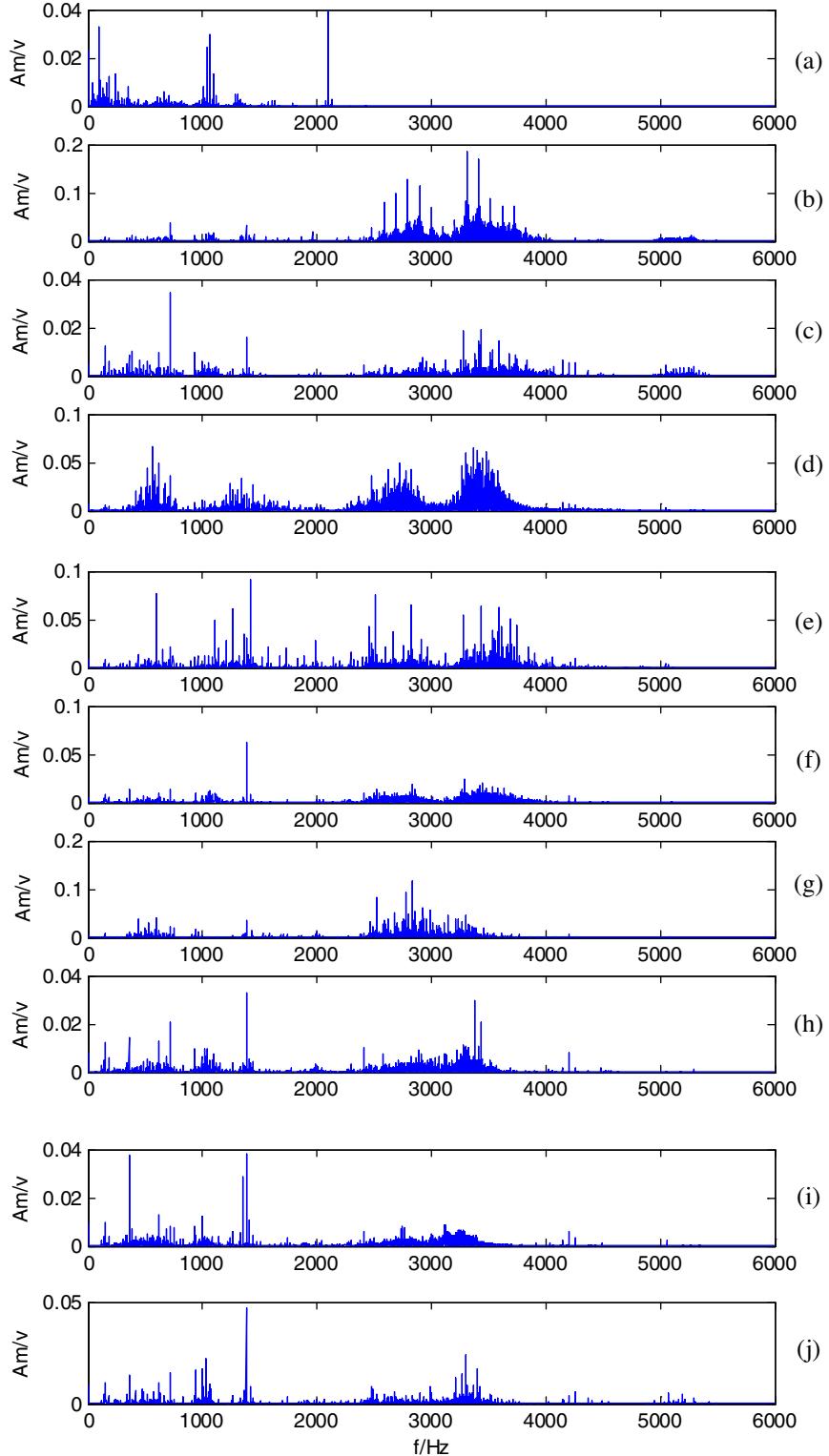


**Fig. 15.** Vibration signals of the bearings in different conditions: (a) normal, (b) outer race fault with 0.007 in., (c) outer race fault with 0.014 in., (d) outer race fault with 0.021 in., (e) inner race fault with 0.007 in., (f) inner race fault with 0.014 in., (g) inner race fault with 0.021 in., (h) ball fault with 0.007 in., (i) ball fault with 0.014 in., (j) ball fault with 0.021 in.

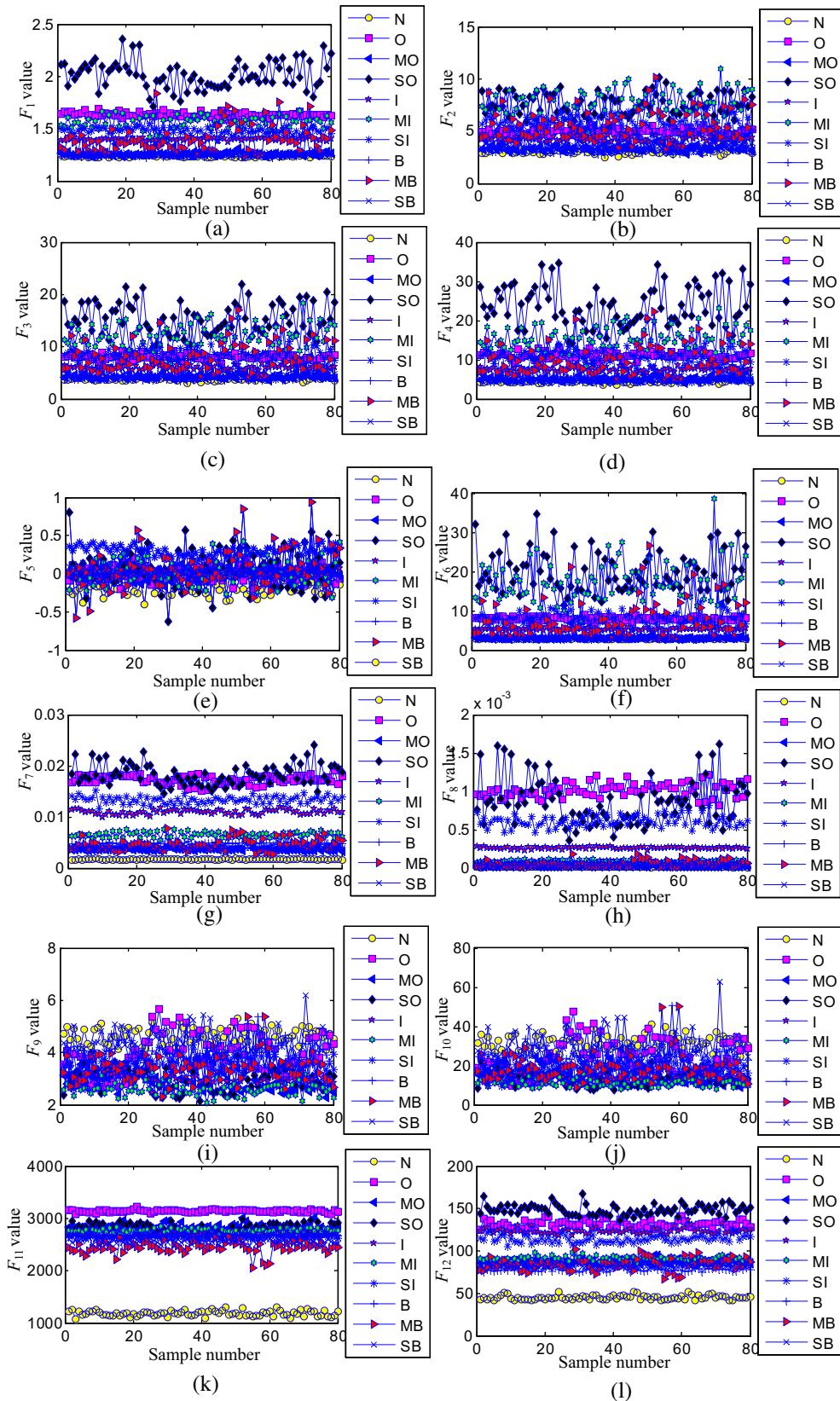
#### 4.2.5. Case 5: A research of the proposed intelligent fault diagnosis method with multivariable ensemble-based incremental support vector machine for gradually learning new information

Since machinery fault samples are generally attained little by little as machinery conditions varies in operation, it is important

to learn new information utilizing the previously learned knowledge and new attained data in an online fashion without discarding the existing classifier. The experiment is conducted on ten different fault conditions of roller bearings including normal condition and three defect types in different severe degree to simulate



**Fig. 16.** Frequency spectrum of signals in different conditions: (a) normal, (b) outer race fault with 0.007 in., (c) outer race fault with 0.014 in., (d) outer race fault with 0.021 in., (e) inner race fault with 0.007 in., (f) inner race fault with 0.014 in., (g) inner race fault with 0.021 in., (h) ball fault with 0.007 in., (i) ball fault with 0.014 in., (j) ball fault with 0.021 in.



**Fig. 17.** Nineteen statistical features ( $F_1 \sim F_{19}$ ) of the vibration signals in different conditions (N: normal, O: outer race fault with 0.007 in., MO: outer race fault with 0.014 in., SO: outer race fault with 0.021, I: inner race fault with 0.007 in., MI: inner race fault with 0.014 in., SI: inner race fault with 0.021 in., B: ball fault with 0.007 in., MB: ball fault with 0.014 in., SB: ball fault with 0.021 in.).

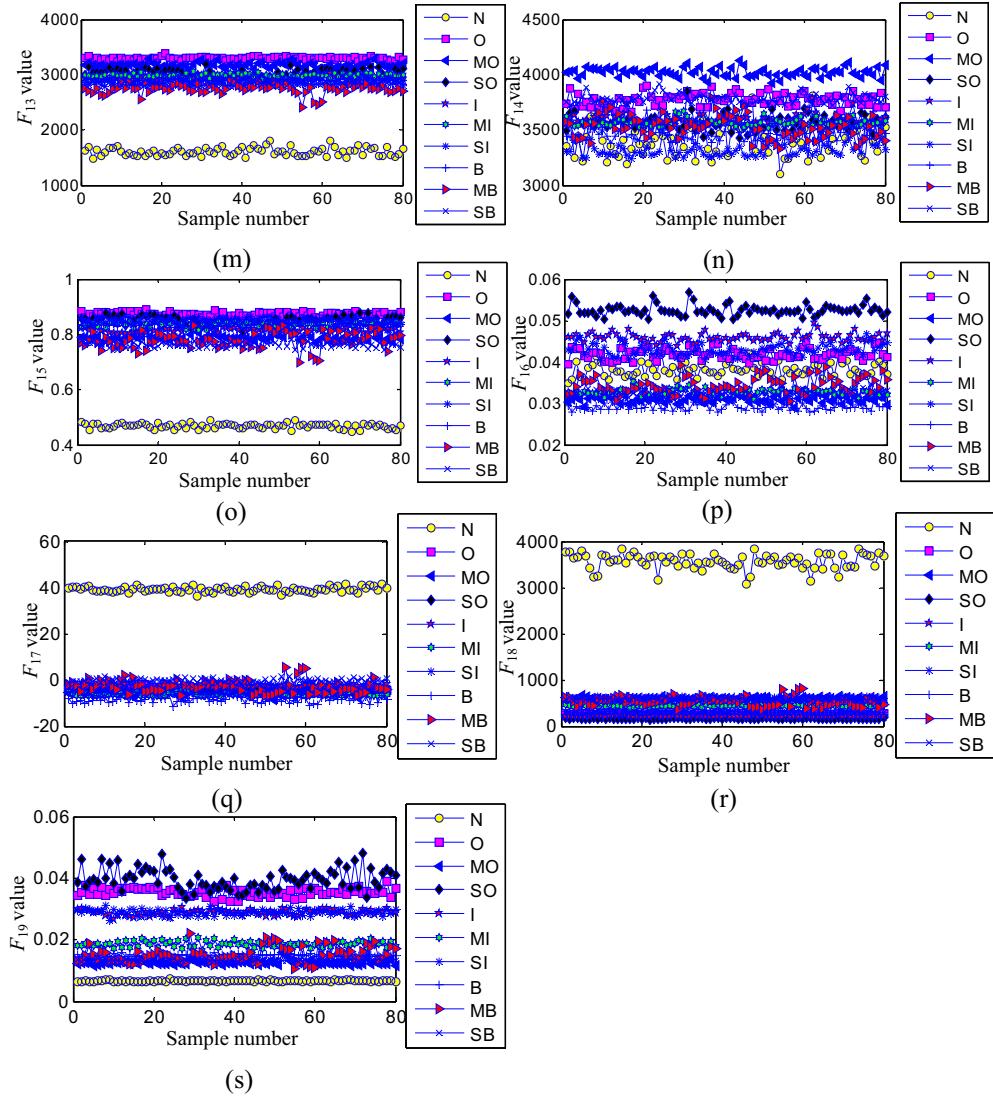


Fig. 17 (continued)

**Table 13**

Experimental results of the proposed intelligent fault diagnosis method with the multivariable ensemble-based incremental support vector machine for gradually learning new information.

Dataset	Sample number	Defect size (in.)	Bearing conditions	Accuracy (%)	CPU time (s)
S <sub>1</sub>	30	0.007	Normal/Outer race defect	100	59.68
S <sub>2</sub>	30	0.014	Outer race defect	100	62.13
S <sub>3</sub>	30	0.021	Outer race defect	100	63.52
S <sub>4</sub>	30	0.007	Inner race defect	100	67.19
S <sub>5</sub>	30	0.014	Inner race defect	100	66.32
S <sub>6</sub>	30	0.021	Inner race defect	100	67.41
S <sub>7</sub>	30	0.007	Ball defect	100	68.54
S <sub>8</sub>	30	0.014	Ball defect	100	68.63
S <sub>9</sub>	30	0.021	Ball defect	100	69.24
TEST	30	0.007/0.014/0.021	Normal/Outer race defect/Inner race defect/Ball defect	96.67	92.37

the case that these fault dataset are not trained in one batch, which is described in Table 12. Vibration signals are respectively plotted in Fig. 15. The frequency spectrums in ten fault conditions are

**Table 14**

Statistical analysis of the proposed multivariable ensemble-based incremental support vector machine (MEISVM) and ARTMP in Refs. [35,36].

	Case 2		Case 3	
	ARTMP	MEISVM	ARTMP	MEISVM
	79.228	93	77.551	92.1429
	91.185	98.5	84.898	96.4214
	89.382		87.302	
	99.541			
Wilcoxon signed-ranks test result	0.5333		0.2	

shown in Fig. 16. It is seen that both of signals in time domain and spectrums in frequency domain are not enough to distinguish ten fault conditions from each other. Each signal in ten different conditions is respectively processed to training samples and testing samples. Each sample includes 1024 points. The training dataset and testing dataset are independent. The multiple variables of each sample are computed according to formulas in Table 1, which are shown in Fig. 17. From Fig. 17(a) to (s), feature  $F_1$ ~ feature  $F_{19}$  of different conditions are more or less irregular.

Nine dataset from  $S_1$  to  $S_9$  is constructed to simulate the cases that each dataset is gradually measured one by one and finally

**Table 15**

Statistical analysis of the proposed multivariable ensemble-based incremental support vector machine (MEISVM) and support vector machine (SVM).

	SVM (10-fold cross validation)	MEISVM (10-fold cross validation)	SVM (Ant colony optimization)	MEISVM (Ant colony optimization)
Case 1	100	100	100	100
Case 2	87	96	93	98.5
Case 3	85.1786	92.1429	89.4643	96.4214
Case 4	97.25	98.13	97.65	99.04
Wilcoxon signed-ranks test result	0.5429		0.3714	

**Table 16**

Description of bearing dataset in the experiment (size: in.).

Bearing	Fault location	Defect diameter	Defect depth	Approx. motor speed (r/min)	State label
Drive End	Normal	–	–	1772	0
	Outer Raceway	0.007	0.011	1772	1
	Inner Raceway	0.007	0.011	1772	1
	Ball	0.007	0.011	1772	1

**Table 17**

Experimental results.

	Sensitivity	Specificity	$F_1^*$	Accuracy	CPU time (s)
SVM	0.95	0.9167	0.9305	0.95	67.14
MEISVM	1	0.9667	0.9831	0.9917	75.32

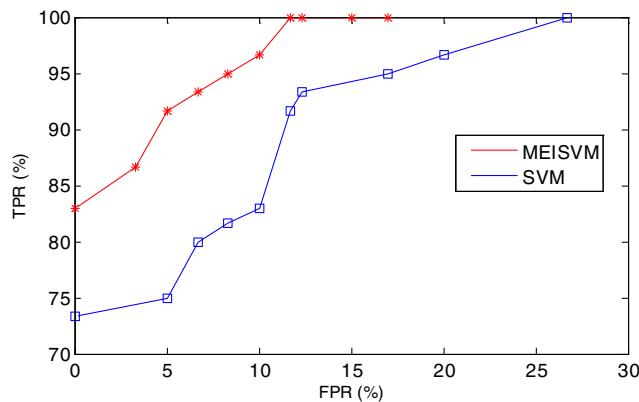


Fig. 18. ROC curves of performance evaluation from SVM and MEISVM.

the *TEST* dataset containing all the bearing conditions is used for testing. These datasets are described in Table 13. Firstly, dataset  $S_1$  including two states is input according to the rule of algorithm described in Section 3, and test accuracy of the testing samples from its own dataset  $S_1$  is 100%. Then dataset  $S_i (i = 2, \dots, 9)$  which contains new information (different defect or different defect size) is gradually input without discarding the existing classifier, the test accuracy attains 100%, which is tested by testing samples from its own dataset  $S_i (i = 2, \dots, 9)$  for learning new information. The experiment results are shown in Table 13. After all the different information is learned in batches, *TEST* dataset containing all the fault states is finally tested. Since features of different defect sizes are very similar and mixing them together increases learning difficulty, the test accuracy is reduced to 96.67%, which shows that the proposed multivariable ensemble-based incremental support vector machine can learn new information in batches, which is suitable for the case that new fault information is coming gradually and learning information without forgetting previously acquired knowledge.

#### 4.3. Statistical analysis

Statistical evaluation of experimental results has been considered as an essential part of validation for machine learning methods. Wilcoxon signed-ranks test (Wilcoxon, 1945) is a non-parametric alternative to the paired *t*-test, which ranks the differences in performances of two classifiers for each data set, ignoring the signs, and compares the ranks for the positive and the negative differences.

Let  $d_i$  be the difference between the performance scores of the two classifiers on  $i$ th out of  $N$  data sets. The differences are ranked according to their absolute values; average ranks are assigned in case of ties. Let  $R^+$  be the sum of ranks for the data sets on which the second algorithm outperformed the first, and  $R^-$  be the sum of ranks for the opposite. Ranks of  $d_i = 0$  are split evenly among the sums. If there is an odd number of them, one is ignored [40].

$$R^+ = \sum_{d_i > 0} rank(d_i) + \frac{1}{2} \sum_{d_i = 0} rank(d_i) \quad (13)$$

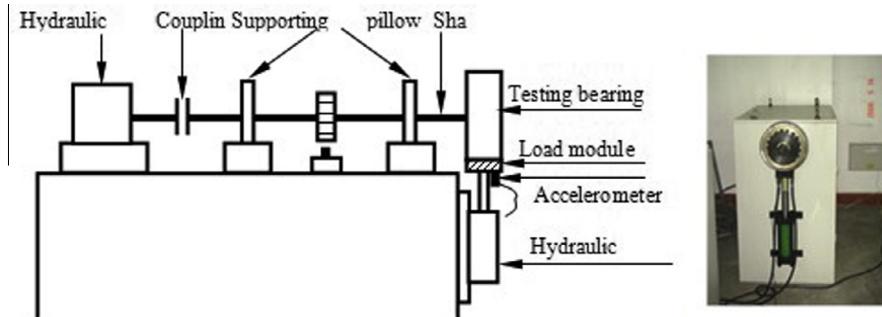


Fig. 19. Test bench of the locomotive roller bearing.

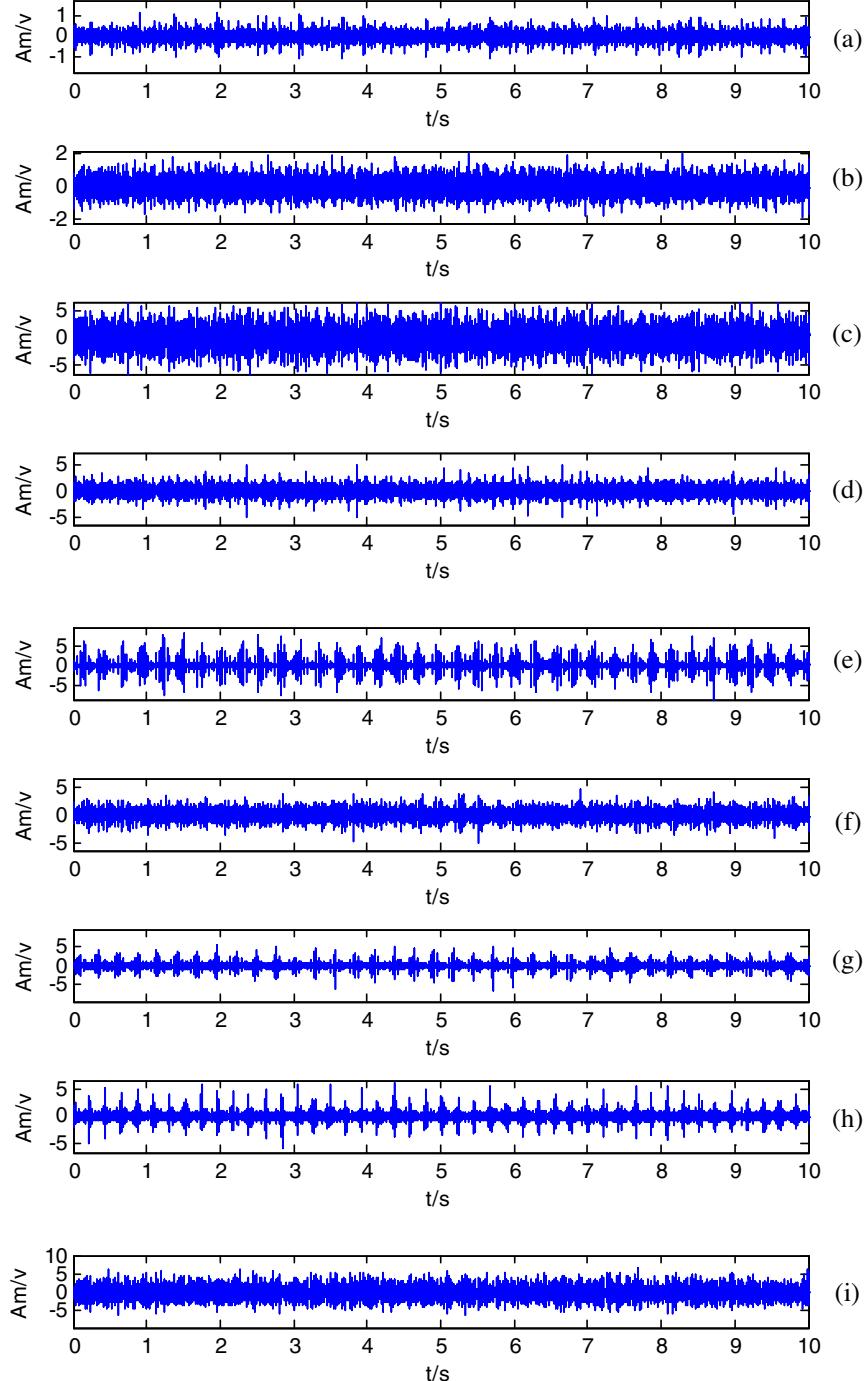
$$R^- = \sum_{d_i < 0} rank(d_i) + \frac{1}{2} \sum_{d_i=0} rank(d_i) \quad (14)$$

Let  $T$  be the smaller of the sums,  $T = \min(R^+, R^-)$ . Most books on general statistics include a table of exact critical values for  $T$  for  $N$  up to 25 (or sometimes more). For a larger number of data sets, the statistics is computed.

$$z = \frac{T - \frac{1}{4}N(N+1)}{\sqrt{\frac{1}{24}N(N+1)(2N+1)}} \quad (15)$$

which is distributed approximately normally. With  $\alpha = 0.05$ , the null-hypothesis can be rejected if  $z$  is smaller than  $-1.96$ .

**Table 14** shows the statistical analysis results of the proposed multivariable ensemble-based incremental support vector machine (MEISVM) and ARTMP in Case 2 and Case 3 [35,36]. The equivalent probability of two methods are 0.5333 and 0.2. **Table 15** shows the statistical analysis of the proposed MEISVM and support vector machine with 10-fold cross validation and ant colony algorithm for parameter selection from case 1 to case 4. The results show that the equivalent probability of the two methods are 0.5429 and 0.3714. The results prove that the



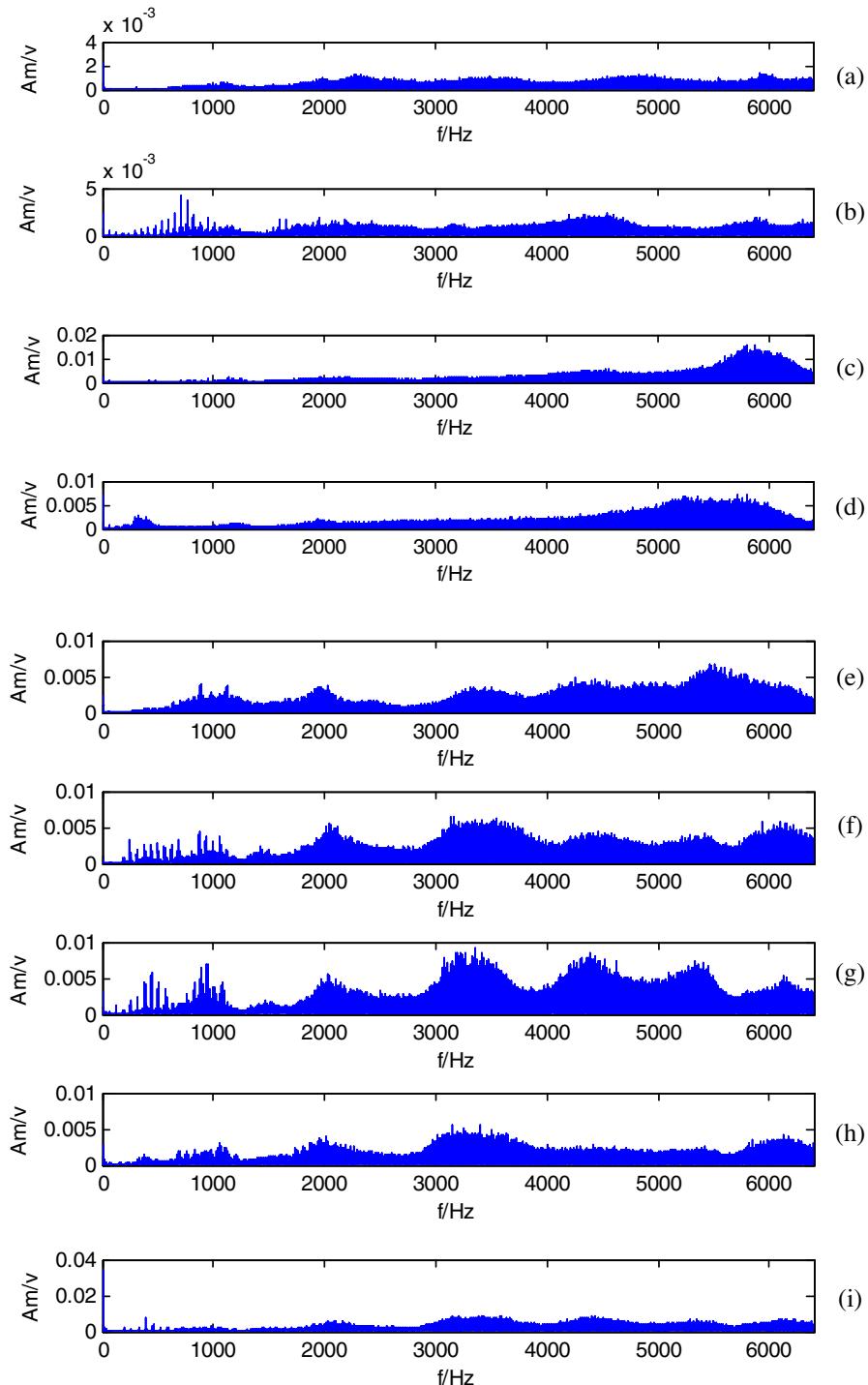
**Fig. 20.** Vibration signals of the locomotive roller bearings in nine condition: (a) normal, (b) slight fault in outer races, (c) serious fault in outer race, (d) inner race fault, (e) roller fault, (f) compound faults in outer race and inner race, (g) compound faults in outer race and rollers, (h) compound faults in inner race and rollers, (i) compound faults in outer race, inner race and rollers.

proposed multivariable ensemble-based incremental support vector machine is better than others.

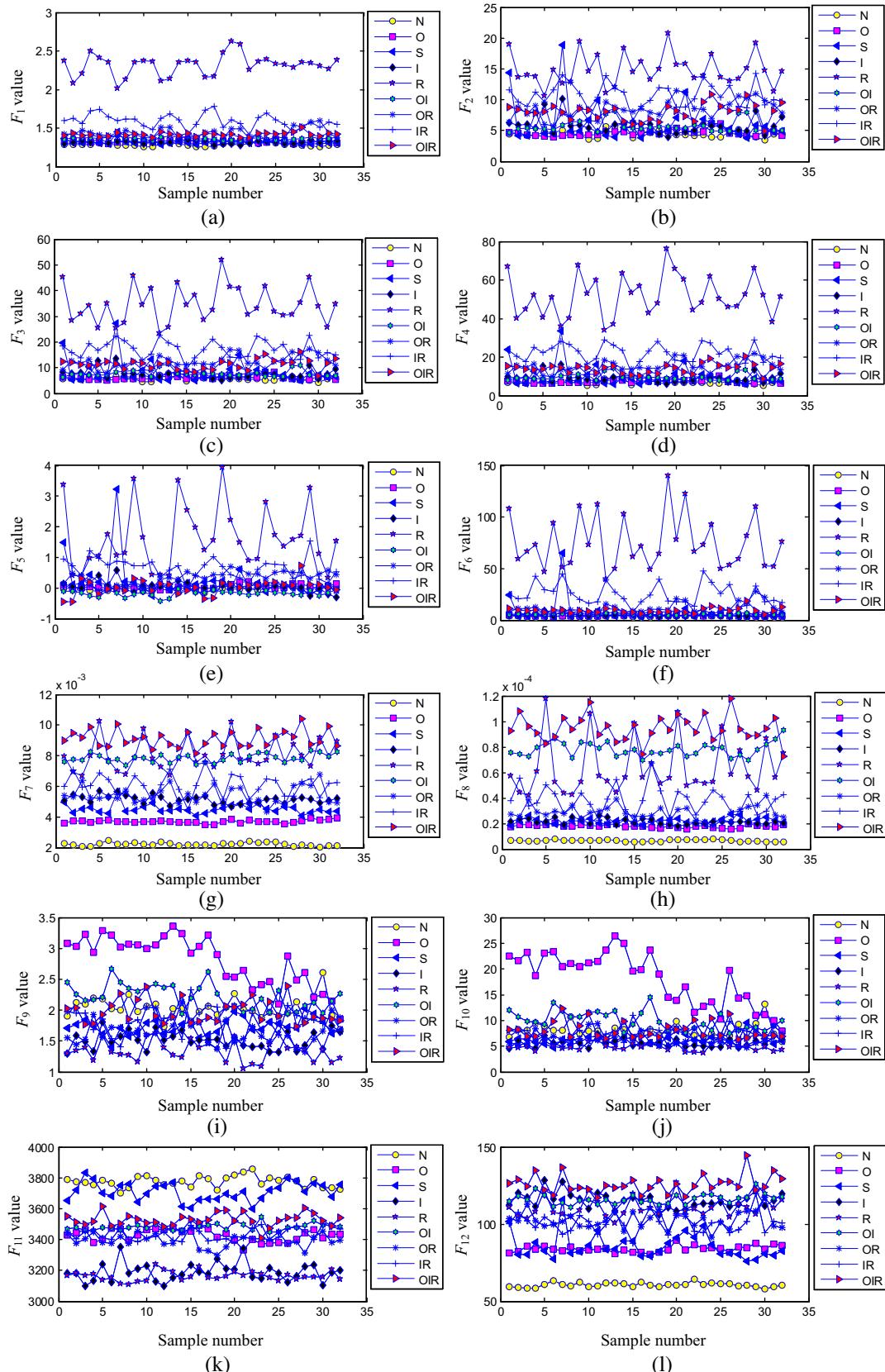
#### 4.4. Performance evaluation

Another important aspect of building classifiers is the evaluation of their performance. The accuracy represents the proportion of cases that are correctly classified, which constitutes the vast majority among the criteria in the considered literature and is also

used in this research. When two classes (positive and negative) are considered, several performance evaluation criterions are suitable. Sensitivity is the proportion of true positive decisions in actually positive cases which is the sum of the amount of true positives (TP) and false negatives (FN). Specificity is the proportion of true negative decisions in actually negative cases which is the sum of the amount of true negatives (TN) and false positives (FP).  $F_1^*$  is another performance measure which can be computed according to formula (19). Receiver operating characteristic (ROC) curve can



**Fig. 21.** Frequency spectra of the signals in nine condition: (a) normal, (b) slight fault in outer races, (c) serious fault in outer race, (d) inner race fault, (e) roller fault, (f) compound faults in outer race and inner race, (g) compound faults in outer race and rollers, (h) compound faults in inner race and rollers, (i) compound faults in outer race, inner race and rollers.



**Fig. 22.** Nineteen statistical features ( $F_1 \sim F_{19}$ ) of the vibration signals in different conditions (N: normal condition, O: slight fault in outer race, S: serious fault in outer race, I: inner race fault, R: roller fault, OI: compound faults in outer race and inner race, OR: compound faults in outer race and rollers condition, IR: compound faults in inner race and rollers, OIR: compound faults in outer race, inner race and rollers).

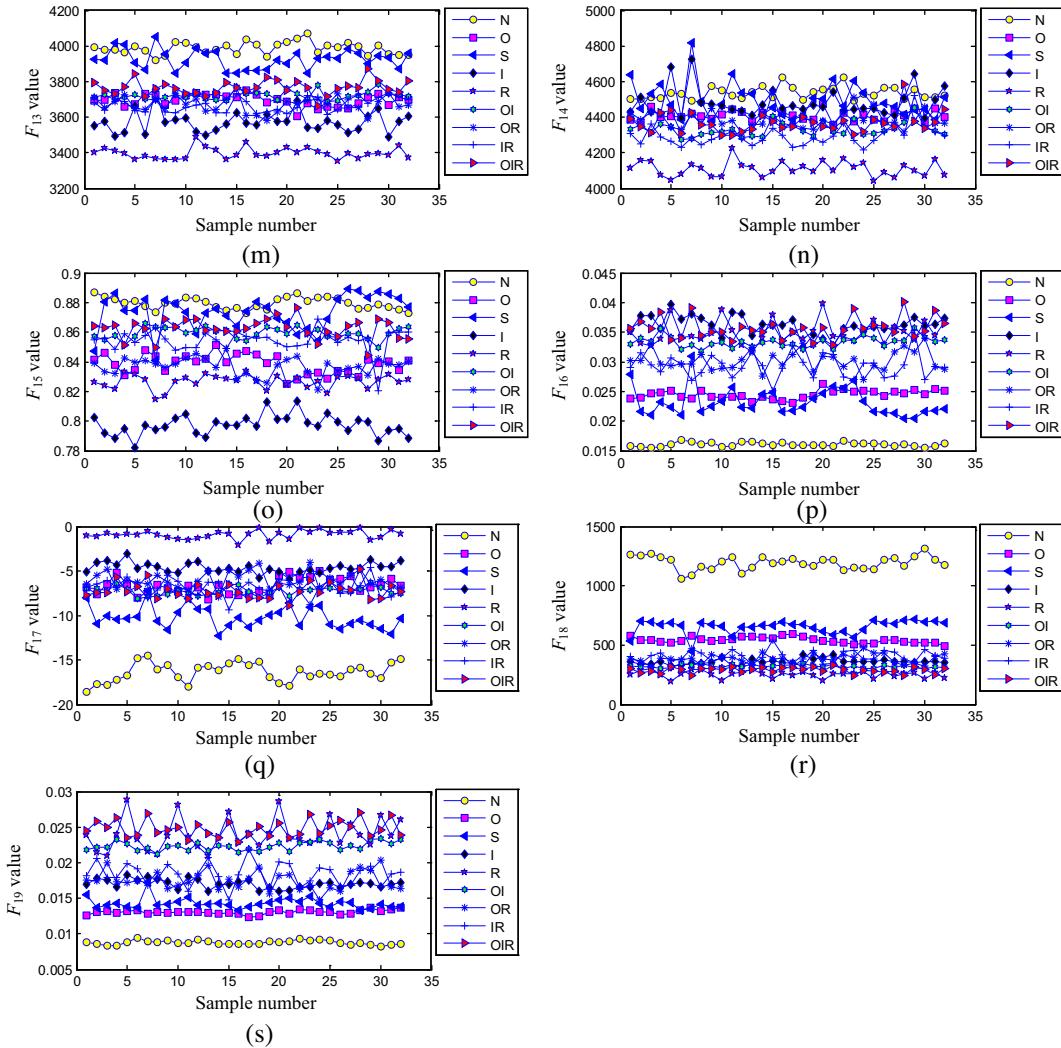


Fig. 22 (continued)

describe inherent detection characteristics of the test. It can handle unbalanced classes in which negative class samples are usually much more common than positive ones, which is a situation frequently encountered case in machinery fault diagnosis, where the normal operation condition is dominant [37]. A ROC curve can be constructed by plotting 1-specificity against the sensitivity over its range.

$$TPR = \frac{TP}{TP + FN} \quad (\text{Sensitivity, Recall}) \quad (16)$$

$$TNR = \frac{TN}{TN + FP} \quad (\text{Specificity}) \quad (17)$$

$$Precision = \frac{TP}{TP + FP} \quad (18)$$

$$F_1^* = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (19)$$

Therefore, the objective is to indicate the presence or absence of a bearing defect, regardless its type and size. The faultless class corresponds to normal bearing without any defect, which is regarded as negative class. Faulty class contains three types of bearing faults (outer race fault, inner race fault, ball fault), which is regarded as positive class. Table 16 describes the samples used

for performance evaluation. The results of performance evaluation criterions (sensitivity, specificity,  $F_1^*$  value and accuracy) are shown in Table 17. The ROC curves are plotted in Fig. 18. The results show that the proposed multivariable ensemble-based incremental support vector machine produces better classification results, compared to standard support vector machine. In summary, the experimental results indicate that the proposed method is promising for intelligent fault diagnosis of roller bearings.

**Table 18**  
Description of the fault conditions of the locomotive roller bearings.

Condition description	Abbreviation of condition type	Label
Normal	N	1
Slight fault in outer race	O	2
Serious fault in outer race	S	3
Inner race fault	I	4
Roller fault	R	5
Compound faults in outer race and inner race	OI	6
Compound faults in outer race and rollers	OR	7
Compound faults in inner race and rollers	IR	8
Compound faults in outer race, inner race and rollers	OIR	9

**Table 19**

Intelligent fault diagnosis results of the locomotive roller bearings.

Name	Accuracy of each fault (%)									Accuracy (%)	CPU time (s)
	N	O	S	I	R	OI	OR	IR	OIR		
SVM	100	100	87.5	100	100	75	75	75	93.75	89.58	187.25
MEISVM	100	100	93.75	100	100	87.5	93.75	93.75	100	96.53	201.36

## 5. Application in the intelligent fault diagnosis of locomotive roller bearings with multivariable ensemble-based incremental support vector machine

In the real application, the vibration signals are often submerged with much noise and one type of fault coupled with another type of fault named as compound fault (such as roller defect coupled with outer race fault) usually occurs. However, it is difficult to identify compound faults in terms of conventional methods, which is a big challenge in the community of monitoring and maintenance of rotating machinery. Moreover, locomotive roller bearings often suffer from high stress and bad environment in railway, so detecting incipient faults including complex compound faults of locomotive roller bearings plays an important role in assuring railway traffic safety. Therefore, the proposed intelligent fault diagnosis with multivariable ensemble-based incremental support vector machine is applied to detect multiple faults in locomotive roller bearings including complex compound faults and a same fault with different severe degrees.

The sketch map of the experimental system is described in Fig. 19. The test bench contains a hydraulic motor, two supporting pillow blocks (mounting with normal bearing), test bearings which are loaded on the outer race by a hydraulic cylinder, a hydraulic radial load application system, and a tachometer for shaft speed measurement. 608A11-type ICP accelerometers are mounted on the load module near the outer race of the test bearing, which measure the vibration data of the test bearings with sampling frequency of 12,800 Hz.

The vibration signals are measured respectively from nine fault conditions of locomotive roller bearings including normal state, slight fault in outer race, serious fault in outer race, inner race fault, roller fault, compound faults in outer race and inner race, compound faults in outer race and rollers, compound faults in inner race and rollers, compound faults in outer race and inner race and rollers, which are described in Table 18. The vibration waveforms and FFT spectrums of the nine fault conditions are respectively shown in Figs. 20 and 21, from which is hard to distinguish different fault types. Sixteen samples of each condition are acquired for training and sixteen samples are independently acquired for testing. Each sample includes 2048 points. Multiple variables of each sample are respectively computed according to formulas in Table 1 and then plotted in Fig. 22, which are fluctuated and overlapped. So the proposed intelligent fault diagnosis method with multivariable ensemble-based incremental support vector machine is performed. The diagnosis results are shown in Table 19. Parameters ( $C, \sigma$ ) in support vector machine are selected by ant colony algorithm from [0.5, 50] and Gaussian kernel is used. Typical support vector machine attains diagnosis accuracy of 89.58%. The proposed multivariable ensemble-based incremental support vector machine improves the recognition accuracy of serious fault in outer race and all kinds of compound faults, so the final diagnosis accuracy of 96.53% is attained.

## 6. Discussion and conclusion

An intelligent fault diagnosis approach for roller bearing with multivariable ensemble-based incremental support vector

machine is proposed. Experiments of the fault diagnosis for roller bearings are conducted in the benchmark dataset of the Bearing Data Center of Case Western Reserve University to testify the efficiency of the proposed method in comparison with some other methods proposed in literature, which are conducted with the same dataset and the same experimental setup. The experimental results prove that the proposed intelligent fault diagnosis with multivariable ensemble-based incremental support vector machine can achieve desirable results than many other different methods. The proposed method is applied in intelligent fault diagnosis of locomotive roller bearings which contains multiple faults, compound faults and different severe degrees of a same fault. Since the final diagnosis result not only relies on building an effective intelligent model but also relies on inputting optimal features, it is worthy to point out that the intelligent fault diagnosis result may be further improved if an advanced feature extraction method or an optimal feature selection method is adopted. The proposed intelligent fault diagnosis with multivariable ensemble-based incremental support vector machine can also be applied to other machinery to achieve success.

## Acknowledgment

This work is supported by National Natural Science Foundation of China (No. 51405028), China Postdoctoral Science Foundation funded project (No. 2015M572553), National Natural Science Foundation of China (No. 51225501). The authors would like to thank Case Western Reserve University for their providing free download of the rolling element bearing fault datasets. The authors also would like to thank D. Parikh and R. Polikar for their contribution work on ensemble-based incremental learning approach.

## References

- [1] B.J. van Wyk, M.A. van Wyk, G. Qi, Difference histograms: a new tool for time series analysis applied to bearing fault diagnosis, Pattern Recogn. Lett. 30 (6) (2009) 595–599. April 15.
- [2] M. Singh, R. Kumar, Thrust bearing groove race defect measurement by wavelet decomposition of pre-processed vibration signal, Measurement 46 (9) (November 2013) 3508–3515.
- [3] X.T. Jiao, K. Ding, G.L. He, An algorithm for improving the coefficient accuracy of wavelet packet analysis, Measurement 47 (January 2014) 207–220.
- [4] S.J. Dong, B.P. Tang, R.X. Chen, Bearing running state recognition based on non-extensive wavelet feature scale entropy and support vector machine, Measurement 46 (10) (December 2013) 4189–4199.
- [5] R. Kumar, M. Singh, Outer race defect width measurement in taper roller bearing using discrete wavelet transform of vibration signal, Measurement 46 (1) (January 2013) 537–545.
- [6] J.H. Yan, L. Lu, Improved Hilbert–Huang transform based weak signal detection methodology and its application on incipient fault diagnosis and ECG signal analysis, Signal Process. 98 (May 2014) 74–87.
- [7] H.M. Liu, X. Wang, C. Lu, Rolling bearing fault diagnosis under variable conditions using Hilbert–Huang transform and singular value decomposition, Math. Probl. Eng. (2014).
- [8] S. Osman, W. Wang, An enhanced Hilbert–Huang transform technique for bearing condition monitoring, Meas. Sci. Technol. 24 (8) (August 2013).
- [9] Q. Sun, P. Chen, D.J. Zhang, et al., Pattern recognition for automatic machinery fault diagnosis, J. Vib. Acoust. Trans. Asme 126 (2) (April 2004) 307–316.
- [10] L. Mi, W. Tan, R. Chen, Multi-steps degradation process prediction for bearing based on improved back propagation neural network, Proc. Instit. Mech. Eng. Part C-J. Mech. Eng. Sci. 227 (7) (July 2013) 1544–1553.
- [11] G.S. Vijay, S.P. Pai, N.S. Sriram, et al., Radial basis function neural network based comparison of dimensionality reduction techniques for effective bearing

- diagnostics, *Proc. Instit. Mech. Eng. Part J-J. Eng. Tribol.* 227 (J6) (June 2013) 640–653.
- [12] D.Y. Dou, J.G. Yang, J.T. Liu, et al., A rule-based intelligent method for fault diagnosis of rotating machinery, *Knowl.-Based Syst.* 36 (December 2012) 1–8.
- [13] M. Boumahdi, J.P. Dron, S. Rechak, et al., On the extraction of rules in the identification of bearing defects in rotating machinery using decision tree, *Expert Syst. Appl.* 37 (8) (August 2010) 5887–5894.
- [14] V. Sugumaran, K.I. Ramachandran, Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing, *Mech. Syst. Signal Process.* 21 (5) (July 2007) 2237–2247.
- [15] Y.D. Zhang, S.H. Wang, P. Phillips, et al., Binary PSO with mutation operator for feature selection using decision tree applied to spam detection, *Knowl.-Based Syst.* 64 (July 2014) 22–31.
- [16] Y.J. Wang, S.Q. Kang, Y.C. Jiang, et al., Classification of fault location and the degree of performance degradation of a rolling bearing based on an improved hyper-sphere-structured multi-class support vector machine, *Mech. Syst. Signal Process.* 29 (May 2012) 404–414.
- [17] F.F. Chen, B.P. Tang, T. Song, et al., Multi-fault diagnosis study on roller bearing based on multi-kernel support vector machine with chaotic particle swarm optimization, *Measurement* 47 (January 2014) 576–590.
- [18] X.L. Tang, L. Zhuang, J. Cai, et al., Multi-fault classification based on support vector machine trained by chaos particle swarm optimization, *Knowl.-Based Syst.* 23 (5) (July 2010) 486–490.
- [19] V.N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1999.
- [20] J. Shawe-Taylor, N. Cristianini, *An Introduction to Support Vector Machine: And Other Kernel Based Learning Methods*, Cambridge University Press, 2000.
- [21] A. Shilton, M. Palaniswami, D. Ralph, et al., Incremental training of support vector machines, *IEEE Trans. Neural Networks* 16 (1) (January 2005) 114–131.
- [22] R. Polikar, L. Udpa, S.S. Udpa, et al., Learn++: an incremental learning algorithm for supervised neural networks, *IEEE Trans. Syst. Man Cybernetics Part C-Appl. Rev.* 31 (4) (November 2001) 497–508.
- [23] P. Laskov, C. Gehl, S. Kruger, et al., Incremental support vector learning: analysis, implementation and applications, *J. Mach. Learn. Res.* 7 (September 2006) 1909–1936.
- [24] D. Parikh, R. Polikar, An ensemble-based incremental learning approach to data fusion, *IEEE Trans. Syst. Man Cybern. Part B-Cybern.* 37 (2) (April 2007) 437–450.
- [25] Z. Erdem, R. Polikar, F. Gurgen, et al., Ensemble of SVMs for incremental learning, *Multiple Classifier Syst.* 3541 (2005) 246–256.
- [26] V.N. Vapnik, *Statistical Learning Theory*, Wiley & Sons Inc, 1998.
- [27] S. Kang, S. Cho, P. Kang, Constructing a multi-class classifier using one-against-one approach with different binary classifiers, *Neurocomputing* 149 (2015) 677–682. February 3.
- [28] C.W. Hsu, C.J. Lin, A comparison of methods for multiclass support vector machines, *IEEE Trans. Neural Networks* 13 (2) (March 2002) 415–425.
- [29] A.C. Lorena, A.C.P.L.F. de Carvalho, J.M.P. Gama, A review on the combination of binary classifiers in multiclass problems, *Artif. Intell. Rev.* 30 (1–4) (December 2008) 19–37.
- [30] M. Galar, A. Fernandez, E. Barrenechea, et al., An overview of ensemble methods for binary classifiers in multi-class problems: experimental study on one-vs-one and one-vs-all schemes, *Pattern Recogn.* 44 (8) (August 2011) 1761–1776.
- [31] Y. Lei, Z. He, Y. Zi, et al., Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAS, *Mech. Syst. Signal Process.* 21 (5) (July 2007) 2280–2294.
- [32] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, *J. Comput. Syst. Sci.* 55 (1) (August 1997) 119–139.
- [33] <http://www.eecs.case.edu/laboratory/bearing/>.
- [34] X.Q. Xiang, J.Z. Zhou, C.S. Li, et al., Fault diagnosis based on Walsh transform and rough sets, *Mech. Syst. Signal Process.* 23 (4) (May 2009) 1313–1326.
- [35] Z.B. Xu, J.P. Xuan, T.L. Shi, et al., A novel fault diagnosis method of bearing based on improved fuzzy ARTMAP and modified distance discriminant technique, *Expert Syst. Appl.* 36 (9) (November 2009) 11801–11807.
- [36] Z.B. Xu, J.P. Xuan, T.L. Shi, et al., Application of a modified fuzzy ARTMAP with feature-weight learning for the fault diagnosis of bearing, *Expert Syst. Appl.* 36 (6) (August 2009) 9961–9968.
- [37] T.W. Rauber, F.D. Boldt, F.M. Varejao, Heterogeneous feature models and feature selection applied to bearing fault diagnosis, *IEEE Trans. Industr. Electron.* 62 (1) (January 2015) 637–646.
- [38] X.L. Zhang, X.F. Chen, Z.J. He, An ACO-based algorithm for parameter optimization of support vector machines, *Expert Syst. Appl.* 37 (9) (September 2010) 6618–6628.
- [39] X.L. Zhang, X.F. Chen, Z.J. He, Fault diagnosis based on support vector machines with parameter optimization by an ant colony algorithm, *Proc. Instit. Mech. Eng. Part C-J. Mech. Eng. Sci.* 224 (C1) (2010) 217–229.
- [40] J. Demsar, Statistical comparisons of classifiers over multiple data sets, *J. Mach. Learn. Res.* 7 (January 2006) 1–30.