Available online at www.sciencedirect.com

ScienceDirect

Mechanical Systems and Signal Processing 21 (2007) 2280–2294

Mechanical Systems and Signal Processing

www.elsevier.com/locate/jnlabr/ymssp

Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs

Yaguo Lei[a,*], Zhengjia He[a,b], Yanyang Zi[a], Qiao Hu[a]

[a]*Department of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, PR China*
[b]*State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, PR China*

Received 7 July 2006; received in revised form 15 November 2006; accepted 18 November 2006
Available online 2 January 2007

Abstract

This paper presents a novel method for fault diagnosis based on empirical mode decomposition (EMD), an improved distance evaluation technique and the combination of multiple adaptive neuro-fuzzy inference systems (ANFISs). The method consists of three stages. First, prior to feature extraction, some preprocessing techniques, like filtration, demodulation and EMD are performed on vibration signals to acquire more fault characteristic information. Then, six feature sets, including time- and frequency-domain statistical features of both the raw and preprocessed signals, are extracted. Second, an improved distance evaluation technique is proposed, and with it, six salient feature sets are selected from the six original feature sets, respectively. Finally, the six salient feature sets are input into the multiple ANFIS combination with genetic algorithms (GAs) to identify different abnormal cases. The proposed method is applied to the fault diagnosis of rolling element bearings, and testing results show that the multiple ANFIS combination can reliably recognise different fault categories and severities, which has a better classification performance compared to the individual classifiers based on ANFIS. Moreover, the effectiveness of the proposed feature selection method based on the improved distance evaluation technique is also demonstrated by the testing results.

Keywords: Empirical mode decomposition; Feature extraction; Feature selection; Improved distance evaluation technique; Multiple adaptive neuro-fuzzy inference system combination; Fault diagnosis

1. Introduction

Most of the machinery used in the modern manufacturing industry operates by means of bearings and other rotating parts which may develop faults. These faults may cause the machine to break down and decrease its level of performance. Different methods of fault diagnosis have been developed and used effectively to detect the machine faults at an early stage. One of the principal tools for diagnosing rotating machinery problems is the vibration analysis [1–4]. Through the use of some processing techniques of vibration signals, it is possible to obtain vital diagnosis information from the vibration signals. However, many techniques available presently require a good deal of expertise to apply them successfully. Simpler approaches are needed which

*Corresponding author. Tel.: +86 29 82667963.
E-mail address: leiyaguo@163.com (Y. Lei).

0888-3270/$ - see front matter
doi:10.1016/j.ymssp.2006.11.003

allow relatively unskilled operators to make reliable decisions without the need for a diagnosis specialist to examine data and diagnose problems. Therefore, there is a demand for techniques that can make decision on the running health of the machine automatically and reliably [4–6].

Artificial intelligent techniques, such as artificial neural networks (ANNs), fuzzy logic and genetic algorithms (GAs), etc., have been successfully applied to automated detection and diagnosis of machine conditions [1,5,7–9]. They largely increase the reliability of fault detection and diagnosis systems. The adaptive neuro-fuzzy inference system (ANFIS) is a hybrid model which combines the ANNs adaptive capability and the fuzzy logic qualitative approach [10]. ANFIS harnesses the power of the two paradigms: ANNs and fuzzy logic, and overcomes their own shortcomings simultaneously.

Recently the combination of multiple classifiers has been intensively studied to overcome the limitations of individual classifiers and achieve higher performance [11–15]. Classifiers differing in feature set usually exhibit complementary classification behaviour. Thus, if the classification results of multiple classifiers, which employ the same classification engine but different input feature sets, are combined by integration techniques to yield the final classification result, the final performance may be superior to the best performance of a single classifier on one feature set [16–19]. The weighted averaging technique is the simplest and most widely used technique combining multiple classifiers, which assigns a nonnegative weight to each individual classifier. By optimising an objective function, the classifier weights can be estimated using various techniques. Here, GAs can be used to search the optimal weights of multiple classifiers.

Different feature sets presented to multiple classifiers may be produced by employing different preprocessing and/or feature extraction methods [11]. In this study, vibration signals are preprocessed with band-pass filtration, high-pass filtration, demodulation and empirical mode decomposition (EMD). EMD is based on the local characteristic time scales of a signal and could decompose the complicated signal into a set of complete and almost orthogonal components named intrinsic mode function (IMF) [20]. Frequency components contained in each IMF not only relates to the sampling frequency, but also changes with the signal itself. Thus, it is a self-adaptive signal processing method that can be applied to non-linear and non-stationary process perfectly. To acquire more faulty information from the vibration signals and enhance the competence of the diagnosis systems, both time- and frequency-domain statistical characteristics are extracted, respectively, from not only the raw vibration signals but also the preprocessed signals. So, through adopting filtration, EMD, demodulation, time- and frequency-domain feature extraction methods, we can obtain several different feature sets. Each feature set contains irrelevant or redundant features as well as salient features. If all features in each feature set are input into a classifier directly, they will make the classification process slower and the classification accuracy lower. Thus, to further improve the classification accuracy and reduce the computational burden of the classifier, a few features which obviously characterise the machine conditions need to be selected from the whole feature set. Due to the simpleness and reliability of the distance evaluation technique [9,21], it is generally adopted in fault diagnosis. Here, an improved distance evaluation technique is presented and applied to feature selection.

In view of the above principles, we utilise a number of techniques, such as statistics approach, EMD, the improved distance evaluation technique, ANFIS and GAs, and propose a novel fault diagnosis method for rotating machinery. The proposed method is shown in Fig. 1. It includes the following several procedures. First, vibration signals are filtered and at the same time, they are decomposed by the EMD method and eight IMFs are acquired. The filtered signals and IMFs are further demodulated to calculate their Hilbert envelope spectrums. Second, six feature sets are obtained. They are, respectively, time- and frequency-domain statistical features of both the raw and preprocessed signals. Third, each feature set is evaluated and a few salient features are selected from it by applying the improved distance evaluation technique. Correspondingly, six salient feature sets are obtained. Forth, each salient feature set is presented to one classifier based on ANFIS for training and testing. There are altogether six different classifiers corresponding to the six salient feature sets. Finally, the weighted averaging technique with GAs is employed to combine the outputs of the six ANFISs and come up with the final diagnosis results. The proposed method is applied to fault diagnosis of rolling element bearings. The vibration signals were measured from the rolling element bearings under various operating loads, and different bearing conditions including different fault categories and severities. The results show the effectiveness of feature

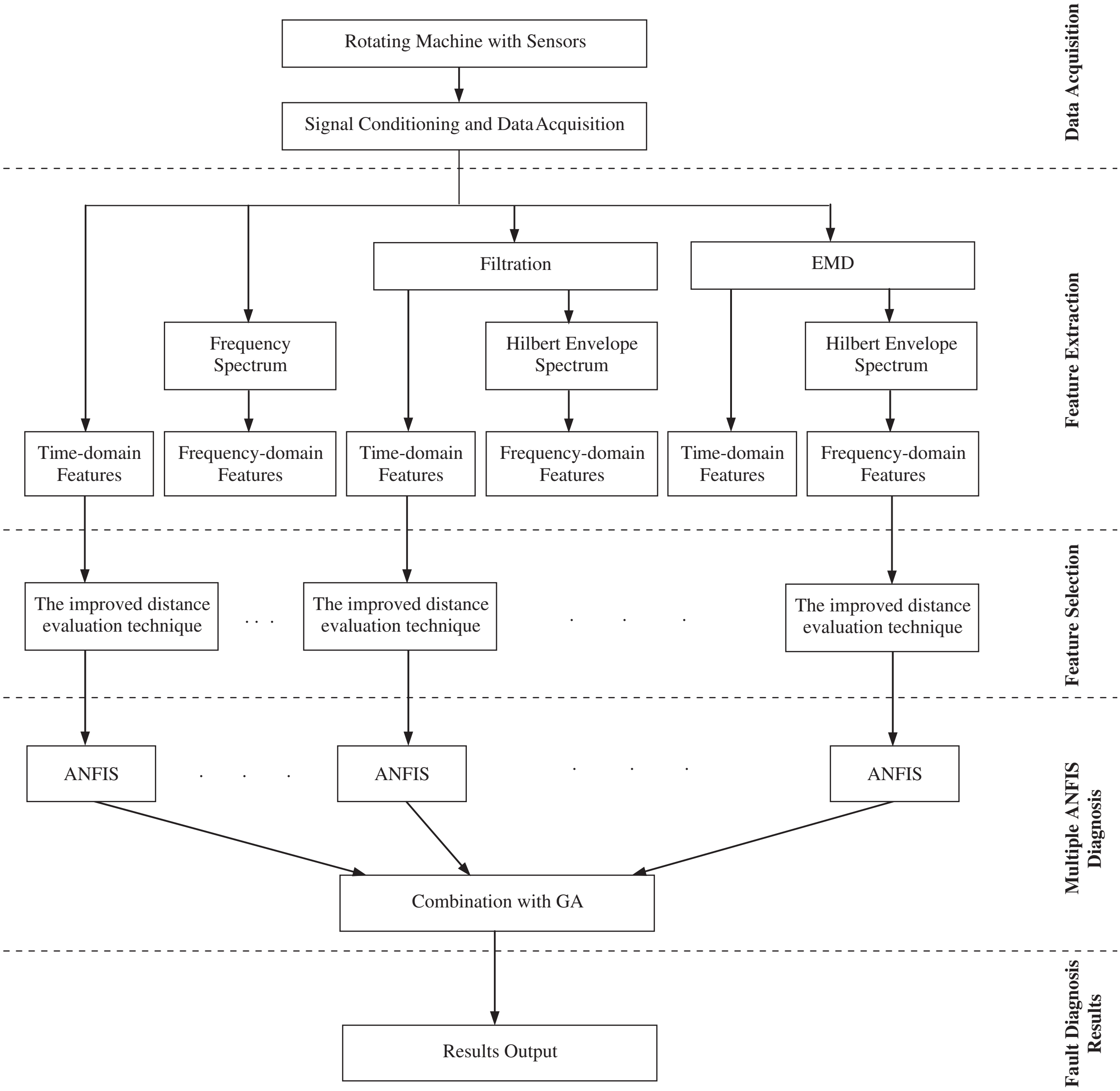

Fig. 1. Flow chart of diagnosis procedure.

extraction and feature selection. And a desired effect has been obtained adopting the combination of multiple ANFISs via GAs.

2. Vibration data

Rolling element bearings, as important components, are widely used in modern rotating machinery; faults occurring in the bearings may lead to fatal breakdowns of machines. Therefore, it is significant to be able to accurately and automatically detect and diagnosis the existence and severity of the faults occurring in the bearings.

The vibration data used in the paper have been obtained from the data set of the rolling element bearings under different operating loads and bearing conditions [22]. The ball bearings are installed in a motor driven

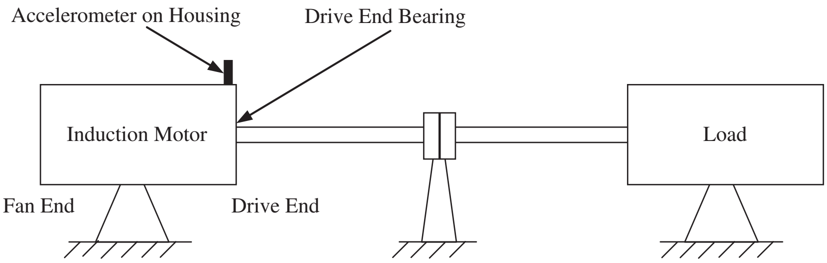

Fig. 2. A schematic of the experimental system.

mechanical system, as shown in Fig. 2. A 2 hp, three-phase induction motor is connected to a dynamometer and a torque sensor by a self-aligning coupling. The dynamometer is controlled so that desired torque load levels can be achieved. An accelerometer is mounted on the motor housing at the drive end of the motor to acquire the vibration signals from the bearing. The data collection system consists of a high bandwidth amplifier particularly designed for the vibration signals and a data recorder with a sampling frequency of 12,000 Hz per channel. The data recorder is equipped with low-pass filters at the input stage for anti-aliasing. On the other hand, the frequency content of interest in the vibration signals of the system under study did not exceed 5000 Hz, and therefore the sampling rate is ample.

The bearings used in this work are deep groove ball bearings manufactured by SKF. Some parameters are listed as follows: inside diameter: 0.9843 inches; outside diameter: 2.0472 inches; ball diameter: 0.3126 inches; pitch diameter: 1.537 inches; and rpm $\approx$ 1800.

The faults were introduced into the drive-end bearing of the motor using the EDM method. The defect sizes (diameter, depth) of the three faults were the same: 0.007, 0.014 or 0.021 inches. Each bearing was tested under the four different loads (0, 1, 2 and 3 hp). The bearing data set was obtained from the experimental system under the four different operating conditions: (1) normal condition; (2) with outer race fault; (3) with inner race fault; and (4) with ball fault.

3. EMD and feature extraction

3.1. EMD method

Huang et al. [20] presented the EMD method, which is able to decompose a signal into some IMFs. An IMF is a function that satisfies the two following conditions: (1) in the whole data set, the number of extrema and the number of zero-crossings must either equal or differ at most by one, and (2) at any point, the mean value of the envelope defined by local maxima and the envelope defined by the local minima is zero. An IMF represents simple oscillatory mode imbedded in the signal. With the simple assumption that any signal consists of different simple IMFs, the EMD method is developed to decompose a signal into IMF components. The EMD process of a signal $x(t)$ can be summarised in Table 1.

At the end of the procedure we have a residue r_I and a collection of I IMFs $c_i (i = 1,2,\ldots,I)$. Summing up all IMFs and the final residue r_I, we obtain

$$x(t) = \sum_{i=1}^{I} c_i + r_I. \tag{1}$$

Thus, we can achieve a decomposition of the signal into I IMFs and a residue r_I, which is the mean trend of $x(t)$. The IMFs $c_1,c_2,\ldots,c_I$ include different frequency bands ranging from high to low. The frequency components contained in each frequency band are different and they change with the variation of signal $x(t)$, while r_I represents the central tendency of signal $x(t)$. A more detailed explanation of EMD can be found in Ref. [20].

3.2. Feature definition

3.2.1. Raw signal statistical characteristics

In this work, twenty-four feature parameters (p_1–p_{24}) were selected. The eleven parameters (p_1–p_{11}) are time-domain statistical characteristics and the thirteen parameters (p_{12}–p_{24}) are frequency-domain statistical characteristics. The twenty-four feature parameters are shown in Table 2.

When faults occur in rotating machinery, the time-domain signal may change. Both its amplitude and distribution may be different from those of a time-domain signal under normal condition. Also, the frequency spectrum and its distribution may change, which signifies that new frequency components may appear and a change of the convergence of the frequency spectrum may take place. Parameter p_1 and p_3–p_5 may reflect the vibration amplitude and energy in time domain. Parameter p_2 and p_6–p_{11} may represent the time series distribution of the signal in time domain. Parameter p_{12} may indicate the vibration energy in frequency domain. Parameter p_{13}–p_{15}, p_{17} and p_{21}–p_{24} may describe the convergence of the spectrum power. Parameter p_{16} and p_{18}–p_{20} may show the position change of main frequencies.

Table 1
The EMD algorithm

(1) Initialise: $r_0 = x(t)$, and $i = 1$.
(2) Extract the ith IMF.
  (a) Initialise: $h_{i(k-1)} = r_i$, $k = 1$.
  (b) Extract the local maxima and minima of $h_{i(k-1)}$.
  (c) Interpolate the local maxima and the minima by cubic spline lines to form upper and lower envelops of $h_{i(k-1)}$.
  (d) Calculate the mean $m_{i(k-1)}$ of the upper and lower envelops of $h_{i(k-1)}$.
  (e) Let $h_{ik} = h_{i(k-1)} - m_{i(k-1)}$.
  (f) If h_{ik} is a IMF then set $\text{IMF}_i = h_{ik}$, else go to step (b) with $k = k+1$.

(3) Define $r_{i+1} = r_i = \text{IMF}_i$.
(4) If r_{i+1} still has least 2 extrema then go to step (2) else decomposition process is finished and r_{i+1} is the residue of the signal.

Table 2
The feature parameters

Time-domain feature parameters		Frequency-domain feature parameters	
$p_1 = \frac{\sum_{n=1}^{N} x(n)}{N}$	$p_7 = \frac{\sum_{n=1}^{N} (x(n)-p_1)^4}{(N-1)p_2^4}$	$p_{12} = \frac{\sum_{k=1}^{K} s(k)}{K}$	$p_{19} = \sqrt{\frac{\sum_{k=1}^{K} f_k^4 s(k)}{\sum_{k=1}^{K} f_k^2 s(k)}}$
$P_2 = \sqrt{\frac{\sum_{n=1}^{N} (x(n)-p_1)^2}{N-1}}$	$p_8 = \frac{p_5}{p_4}$	$p_{13} = \frac{\sum_{k=1}^{K} (s(k)-p_{12})^2}{K-1}$	$p_{20} = \frac{\sum_{k=1}^{K} f_k^2 s(k)}{\sqrt{\sum_{k=1}^{K} s(k) \sum_{k=1}^{K} f_k^4 s(k)}}$
$p_3 = \left(\frac{\sum_{n=1}^{N} \sqrt{\lvert x(n) \rvert}}{N}\right)^2$	$p_9 = \frac{p_5}{p_3}$	$p_{14} = \frac{\sum_{k=1}^{K} (s(k)-p_{12})^3}{K(\sqrt{p_{13}})^3}$	$p_{21} = \frac{p_{17}}{p_{16}}$
$p_4 = \sqrt{\frac{\sum_{n=1}^{N} (x(n))^2}{N}}$	$p_{10} = \frac{p_4}{\frac{1}{N}\sum_{n=1}^{N} \lvert x(n) \rvert}$	$p_{15} = \frac{\sum_{k=1}^{K} (s(k)-p_{12})^4}{K p_{13}^2}$	$p_{22} = \frac{\sum_{k=1}^{K} (f_k-p_{16})^3 s(k)}{K p_{17}^3}$
$p_5 = \max \lvert x(n) \rvert$	$p_{11} = \frac{p_5}{\frac{1}{N}\sum_{n=1}^{N} \lvert x(n) \rvert}$	$p_{16} = \frac{\sum_{k=1}^{K} f_k s(k)}{\sum_{k=1}^{K} s(k)}$	$p_{23} = \frac{\sum_{k=1}^{K} (f_k-p_{16})^4 s(k)}{K p_{17}^4}$
$p_6 = \frac{\sum_{n=1}^{N} (x(n)-p_1)^3}{(N-1)p_2^3}$		$p_{17} = \sqrt{\frac{\sum_{k=1}^{K} (f_k-P_{16})^2 s(k)}{K}}$	$p_{24} = \frac{\sum_{k=1}^{K} (f_k-p_{16})^{1/2} s(k)}{K\sqrt{p_{17}}}$
where $x(n)$is a signal series for $n = 1,2,\ldots,N$, N is the number of data points.		$p_{18} = \sqrt{\frac{\sum_{k=1}^{K} f_k^2 s(k)}{\sum_{k=1}^{K} s(k)}}$	
		where $s(k)$ is a spectrum for $k = 1,2,\ldots,K$, K is the number of spectrum lines; f_k is the frequency value of the kth spectrum line.	

The original data sets were divided into some signals of 4096 data points. Each of these signals was processed to extract eleven time-domain features from it and thirteen frequency-domain features from its FFT spectrum, respectively. The time- and frequency-domain features extracted here were hereafter referred as feature set 1 and feature set 2, respectively. Therefore, feature set 1 and 2 contain 11 and 13 feature values, respectively.

3.2.2. Statistical characteristics of filtered signals

The examination of the acquired vibration signals indicated the presence of low-frequency interference. The signals were subjected to either high- or band-pass filtration to remove the low-frequency interference components. Three band-pass (BP1–BP3) and one high-pass (HP) filters were adopted. The band-pass frequencies (in kHz) of the BP1–BP3 filters were chosen as: BP1 (2.2–3.8), BP2 (3.0–3.8), and BP3 (3.0–4.5), respectively. The cut-off frequency of the HP filter was chosen as 2.2 kHz. These frequencies were selected to cover the signal components containing the majority of the rolling element bearing energy.

The eleven time-domain features shown in Table 2 were extracted from each of these filtered signals. 11×4 time-domain features were obtained and defined as feature set 3. Therefore, feature set 3 contains 44 feature values.

The effects of interfering signals within the selected frequency band can be minimised by demodulation. Demodulation detection makes the diagnosis process a little more independent of a particular machine since it focuses on the low-amplitude high-frequency broadband signals characterising bearing conditions [2,23].

Defining a signal $x(t)$, we can have its Hilbert transform as follows:

$$H[x(t)] = \frac{\int_{-\infty}^{\infty} \frac{x(t)}{t-\tau}\,\mathrm{d}\tau}{\pi}. \tag{2}$$

Then the Hilbert envelope spectrum can be given as follows:

$$h(f) = \int_{-\infty}^{\infty} \sqrt{x^2(t) + H^2[x(t)]}\mathrm{e}^{-j2\pi ft}\,\mathrm{d}t. \tag{3}$$

The Hilbert envelope spectrums of the filtered signals were further processed to extract another set of 13×4 frequency-domain features. This feature set was referred as feature set 4 and it contains 52 feature values.

3.2.3. Statistical characteristics of IMFs

To extract more information, each of these raw signals was decomposed via the EMD method. The first eight IMFs containing almost all valid information were selected. Similar to the feature extraction method of the raw signals, the eleven features in time-domain were extracted from each IMF. Then, we obtained an additional set of 11×8 time-domain features referred as feature set 5. Therefore, feature set 5 contains 88 feature values.

We can utilise the Hilbert transform to demodulate each IMF, and compute its Hilbert envelope spectrum. Having gotten the Hilbert envelope spectrum of each of the IMFs, we extracted the thirteen frequency-domain features from it and finally derived another set of 13×8 frequency-domain features defined as feature set 6. Thus, feature set 6 contains 104 feature values.

3.3. Feature selection

Although the above features may identify faults occurring in rotating machinery from different aspects, they have different importance degrees to identify the different faults. Some features are salient and closely related to the fault, but others are not. Thus, before a feature set is fed into a classifier, salient features providing bearing fault-related information must be selected from the feature set and irrelevant or redundant features must be discarded to improve the classifier performance and avoid the curse of dimensionality. Here, an improved distance evaluation technique is presented and it is used to select the salient features from the whole feature set.

Suppose that a feature set of C conditions is

$$\left\{q_{m,c,j}, \quad m=1,2,\ldots,M_c; \quad c=1,2,\ldots,C; \quad j=1,2,\ldots,J\right\}, \tag{4}$$

where $q_{m,c,j}$ is the jth eigenvalue of the mth sample under the cth condition, M_c is the sample number of the cth condition, and J is the feature number of each sample. We may collect M_c samples under the cth condition. Therefore, for C classes, we got $M_c \times C$ samples. For each sample, J features may be extracted to represent the sample. Thus, $M_c \times C \times J$ features may be obtained. And these features were defined as a feature set $\{q_{m,c,j}\}$.

Then the feature selection method based on the improved distance evaluation technique can be depicted as follows:

Step 1: Calculating the average distance of the same condition samples

$$d_{c,j} = \frac{1}{M_c(M_c-1)} \sum_{l,m=1}^{M_c} \left|q_{m,c,j} - q_{l,c,j}\right|, \quad l,m=1,2,\ldots,M_c, \quad l \neq m, \tag{5}$$

then getting the average distance of C conditions

$$d_j^{(w)} = \frac{1}{C}\sum_{c=1}^{C} d_{c,j}. \tag{6}$$

Step 2: Defining and calculating the variance factor of $d_j^{(w)}$ as follows:

$$v_j^{(w)} = \frac{\max(d_{c,j})}{\min(d_{c,j})}. \tag{7}$$

Step 3: Calculating the average eigenvalue of all samples under the same condition

$$u_{c,j} = \frac{1}{M_c}\sum_{m=1}^{M_c} q_{m,c,j}, \tag{8}$$

then obtaining the average distance between different condition samples

$$d_j^{(b)} = \frac{1}{C(C-1)} \sum_{c,e=1}^{C} \left|u_{e,j} - u_{c,j}\right|, \quad c,e=1,2,\ldots,C, \quad c \neq e. \tag{9}$$

Step 4: Defining and calculating the variance factor of $d_j^{(b)}$ as follows:

$$v_j^{(b)} = \frac{\max(|u_{e,j} - u_{c,j}|)}{\min(|u_{e,j} - u_{c,j}|)}, \quad c,e=1,2,\ldots,C, \quad c \neq e. \tag{10}$$

Step 5: Defining and calculating the compensation factor as follows:

$$\lambda_j = \frac{1}{\frac{v_j^{(w)}}{\max(v_j^{(w)})} + \frac{v_j^{(b)}}{\max(v_j^{(b)})}}. \tag{11}$$

Step 6: Calculating the ratio $d_j^{(b)}$ and $d_j^{(w)}$ and assigning the compensation factor

$$\alpha_j = \lambda_j \frac{d_j^{(b)}}{d_j^{(w)}} \tag{12}$$

then normalising α_j by its maximum value and getting the distance evaluation criteria

$$\bar{\alpha}_j = \frac{\alpha_j}{\max(\alpha_j)}. \tag{13}$$

Clearly, bigger $\overline{\alpha}_j (j=1,2,\ldots,J)$ signifies that the corresponding feature is better to separate the C conditions. Thus, the salient features can be selected from the feature set $q_{m,c,j}$ according to the distance evaluation criteria $\overline{\alpha}_j$ from large value to small value.

3.4. Normalisation

The importance of normalisation to both the speed and success of training a classifier has been proved. Prior to training the classifier, all features have been normalised. The eigenvalue f_t of the tth feature parameter was normalised by the following formula:

$$f'_t = \frac{f_t}{\max(|f_t|)}, \quad t = 1, 2, \ldots, T, \tag{14}$$

where T is the feature number.

Although the normalisation we used does not convert all features into the same range, it converts all of them in the interval [−1, 1]. Through several experiments, we finally decided to use this normalisation because the classification accuracy using it is higher than that of using the others.

4. Multiple ANFIS combination with GAs

4.1. Review of ANFIS

The ANFIS is a fuzzy Sugeno model of integration where the final fuzzy inference system is optimised via the ANNs training. It maps inputs through input membership functions and associated parameters, and then through output membership functions to outputs. The initial membership functions and rules for the fuzzy inference system can be designed by employing human expertise about the target system to be modeled. Then ANFIS can refine the fuzzy if–then rules and membership functions to describe the input/output behaviour of a complex system [3]. Jang [10] showed that even if human expertise is not available it is possible to intuitively set up reasonable membership functions and then employ the ANNs training process to generate a set of fuzzy if–then rules that approximate a desired data set.

In order to improve the training efficiency and eliminate the possible trapping due to local minima, a hybrid learning algorithm is employed to tune the parameters of the membership functions. It is a combination of the gradient descent approach and least-squares estimate. During the forward pass, the node outputs advance until the output membership function layer, where the consequent parameters are identified by the least-squares estimate. The backward pass uses the back propagation gradient descent method to update the premise parameters, based on the error signals that propagate backward. More detailed description about ANFIS can be referred to Ref. [10].

4.2. Combining multiple ANFISs with GAs

Combining multiple classifiers has recently been attracting considerable attention because of its potential to ameliorate the performance of individual classifiers. It is a collection of several classifiers whose individual decisions are combined in some way to classify testing examples. The idea of combining multiple classifiers in a committee is based on the expectation that the committee can outperform its members.

The classifiers exhibiting different behaviour will provide complementary information each other. When they are combined, performance improvement will be obtained. Thus, diversity between the classifiers is recognised to be one of the desired characteristics required to achieve this improvement [13,24–26]. This diversity can be achieved through using different feature sets, which means that the classifiers are constituted based on all feature sets separately and each classifier uses a unique feature set. Different feature sets may be created by adopting different preprocessing and/or feature extraction methods. Therefore, if the classification results of multiple classifiers using different feature sets are fused by integration techniques, the final recognition accuracy may be higher than that of any of the participating classifiers. Of the various integration techniques proposed in the literature, the weighted combination (weighted averaging technique) is the simplest and most frequently used, which is to assign a nonnegative weight to each individual classifier.

In our study, the six different feature sets have been extracted and the relevant six salient feature sets have been selected. ANFIS was adopted to implement the committee member. The weighted averaging technique was utilised to combine the six classifiers based on ANFIS, and the final classification result is given as follows:

$$\hat{y}_n = \sum_{k=1}^{6} w_k \hat{y}_{n,k}, \quad n = 1, 2 \ldots, N', \quad k = 1, 2, \ldots, 6, \tag{15}$$

subject to

$$\begin{cases} \sum_{k=1}^{6} w_k = 1, \\ w_k \geqslant 0, \end{cases} \tag{16}$$

where $\hat{y}_n$ and $\hat{y}_{n,k}$ represent the classification results of the nth sample using the classifier combination and the kth single classifier, respectively, w_k is the weight associated with the kth single classifier, N' is the number of all samples.

Here, the weights were estimated by using GAs to optimise the fitness function. Real-coded genomes were adopted and a population size of ten individuals was used starting with randomly generated genomes. The maximum number of generations 100 was chosen as the termination criterion for the solution process. Non-uniform-mutation function and arithmetic crossover operator [27] were used with the mutation probability of 0.01 and the crossover probability of 0.8, respectively. The fitness function can be defined as

$$f = \frac{1}{1+E}, \tag{17}$$

where E is root mean square training errors expressed as follows:

$$E = \left[\frac{1}{N''} \sum_{n=1}^{N''} (y_n - \hat{y}_n)^2 \right]^{1/2}, \quad n = 1, 2, \ldots, N'', \tag{18}$$

where y_n is the real result of the nth training sample, N'' is the number of the training samples.

5. Fault diagnosis

5.1. Classification performance comparison of different faulty data set

In order to evaluate the proposed method, we conducted four experiments over four different data subsets (A–D) from the whole data set of the rolling element bearings. The detailed descriptions of the four data sets are shown in Table 3.

Data set A consists of 240 data samples of four different operating conditions (normal condition, outer race fault, inner race fault and ball fault) with the fault defect size of 0.007 inches under four various loads (0, 1, 2 and 3 hp). Each of the four operating conditions includes 60 data samples. Data set A is split into two sets: 120 samples for training and 120 for testing. It is a four-class classification task corresponding to the four different operating conditions.

Data set B also contains 240 data samples. 120 samples with the fault detect size of 0.007 inches are identical with the 120 training samples of data set A and form the training set of data set B. The remaining 120 samples with the fault detect size of 0.021 inches are the testing samples of data set B. The experiment over this data set is carried out to further investigate the classification ability to the developing faults when the classifier is trained by the incipient faulty samples.

Data set C is similar to data set B. They have the same samples. But the training set of C is the testing set of B and the testing set of C is the training set of B. The purpose of using this data set is to test the classification performance to the incipient faults when the classifier is trained by the serious faulty samples.

Data set D comprises 600 data samples covering four different operating conditions and four different loads. Each fault condition includes three different defect sizes of 0.007, 0.014 and 0.021 inches, respectively. The 600 data samples are divided into 300 training and 300 testing instances. For data set D, in order to identify the severe grades of faults, we solved the ten-class classification problem.

Table 3
Description of four data sets

Data set	The number of training samples	The number of testing samples	Defect size of training/testing samples (inches)	Operating condition	Label of classification
A	30	30	0/0	Normal	1
	30	30	0.007/0.007	Outer race	2
	30	30	0.007/0.007	Inner race	3
	30	30	0.007/0.007	Ball	4
B	The training samples are the same to A above.	30	0/0	Normal	1
		30	0.007/0.021	Outer race	2
		30	0.007/0.021	Inner race	3
		30	0.007/0.021	Ball	4
C	The training samples are the testing samples of B.	The testing samples are the training samples of B.	0/0	Normal	1
			0.021/0.007	Outer race	2
			0.021/0.007	Inner race	3
			0.021/0.007	Ball	4
D	30	30	0/0	Normal	1
	30	30	0.007/0.007	Outer race	2
	30	30	0.007/0.007	Inner race	3
	30	30	0.007/0.007	Ball	4
	30	30	0.014/0.014	Outer race	5
	30	30	0.014/0.014	Inner race	6
	30	30	0.014/0.014	Ball	7
	30	30	0.021/0.021	Outer race	8
	30	30	0.021/0.021	Inner race	9
	30	30	0.021/0.021	Ball	10

Here, a single output was used to represent multiple classes. It is more likely to produce false results, and therefore may lead to lower classification accuracy. However, the single output may simplify the neural network architecture in a multiple class problem. Thus, the parameter number updated becomes less and the process of training and testing is accelerated.

In order to know how well the proposed method based on multiple ANFIS combination works, i.e. how significant the generalisation ability is improved by utilising the combination method, in our experiments we also tested the performance of the individual classifiers based on ANFIS.

Considering the computational burden, in each case, only four salient features have been selected from each of the six feature sets via the improved distance evaluation technique and then presented to the corresponding ANFIS classifier. For data set A, the distance evaluation criteria $\overline{\alpha}_j$ of the six feature sets are shown in Fig. 3. The classification results of the four data sets are shown in Table 4 and Fig. 4, respectively.

Examining the results from Table 4 and Fig. 4, a number of things can be seen. First, for data set A, since this classification problem is relatively simple, both the six individual classifiers and their combination achieve the high training and testing accuracy (100%). However, observing the classification errors of the individual classifiers and classifier combination shown in Fig. 5, we can see that the classification error of the classifier combination is the least.

Second, although 100% training success is obtained by both the individual classifiers and classifier combination on data set B, the testing success of the individual classifiers is obviously lower than that of the classifier combination. The testing success of the former is in the range of 60.83–83.33% (average 70%), whereas it is 92.5% for the latter. This experimental result indicates that the combination of the classifiers trained by the incipient faulty data can diagnose the developing faults better than every single classifier.

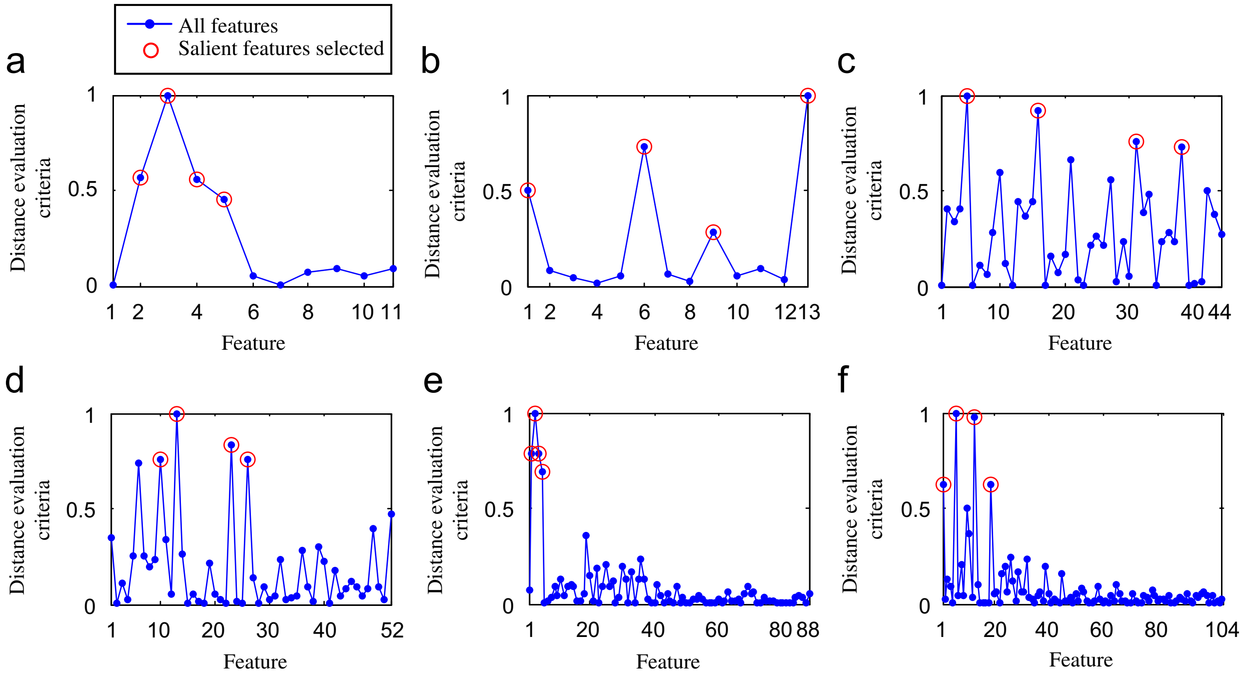

Fig. 3. Distance evaluation criteria of six feature sets of data set A: (a) feature set 1, (b) feature set 2, (c) feature set 3, (d) feature set 4, (e) feature set 5, and (f) feature set 6.

Table 4
Performance comparison between individual classifiers and classifier combination for four different data sets

Data set	Classifier 1		Classifier 2		Classifier 3		Classifier 4		Classifier 5		Classifier 6		Average of six classifiers		Classifier combination	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
A	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
B	100	66.67	100	75	100	61.67	100	83.33	100	60.83	100	72.5	100	70	100	92.5
C	100	62.5	100	79.17	100	74.17	100	80	100	55.83	100	81.67	100	72.22	100	90.83
D	65.67	61	90	87.67	72.67	68	80.33	77	67.67	67	87.33	81	77.28	73.61	93.67	91.33

Third, in case of data set C, we can see that the training accuracies of all the classifiers are 100%. However, the classifier combination produces better testing performance (90.83%) than any of the individual classifiers ranging from 55.83% to 81.67% (average 72.22%). This result shows that the combination of the classifiers trained by the serious faulty data can diagnose the incipient faults with a higher performance in comparison with the individual classifiers.

Finally, the training success rates of all the classifiers decrease and range from 65.67% to 93.67% for data set D because this is a ten-class classification problem and therefore relatively difficult. But the highest training accuracy (93.67%) is still achieved by the classifier combination. For testing, the classification success of the six individual classifiers is in the range of 61–87.67% (average 73.61%), whereas the classification success of the classifier combination is higher (91.33%). These imply that the classifier combination can identify not only the different fault categories but also the different fault severities better.

All the above result analyses prove that it would usually be better to combine the results of multiple classifiers for different choices of feature sets than to use a single classifier trained with one feature set. The proposed combination method have obtained significant achievements in recognition accuracy and provided a better generalisation capability compared to the individual classifiers.

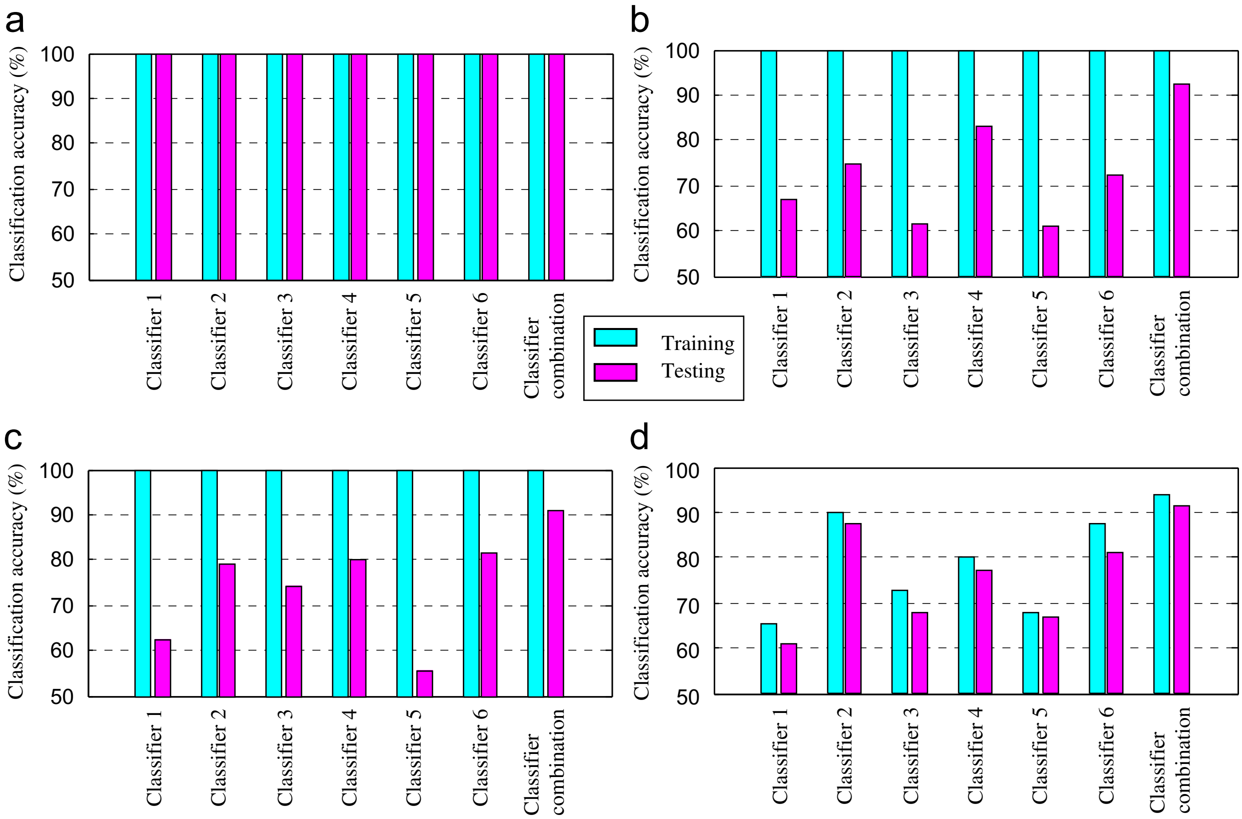

Fig. 4. Performance comparison between individual classifiers and classifier combination: (a) data set A, (b) data set B, (c) data set C, and (c) data set D.

5.2. Effect of feature selection

To examine the effect of the proposed feature selection method based on the improved distance evaluation technique in this work, we took the above data set D for example and carried out another experiment on it. Here, four features, the same feature number as the above experiments, were selected from each of the six feature sets randomly instead of employing the feature selection method. We did not evaluate the performance of a feature selection method by comparing the classifier performance from selected features with that from all unselected features. If we input all unselected features into the classifier based on ANFIS, the computational time is too long, which cannot be accepted by us. Thus, we adopt the proposed method in our paper to evaluate the performance of a feature selection. Though the proposed comparison is not standard, we may attain the purpose of evaluating the performance of a feature selection method. Moreover, the time taken for carrying the proposed comparison method is acceptable.

The experiment was repeated ten times and the average results are shown in Table 5 and Fig. 6, respectively. For comparison, Table 5 and Fig. 6 also present the classification results of all the classifiers with the salient features selected using the proposed feature selection method.

From Table 5 and Fig. 6, we can see that, using the four features selected randomly, the training and testing accuracies of the six individual classifiers are in the range of 56.5–73.73% (average 62.89%) and 50.8–66.93% (average 58.02%), respectively. Their average training accuracy and average testing accuracy, respectively, decrease by 14.39% and 15.59% compared with those of using the salient features selected by the proposed feature selection method. Similarly, the training and testing success rates of the classifier combination with the features selected randomly fall from 93.67% to 81.19% and 91.33% to 77.15%, respectively. This might suggest that some of the features selected randomly contain too much fault-unrelated information and that

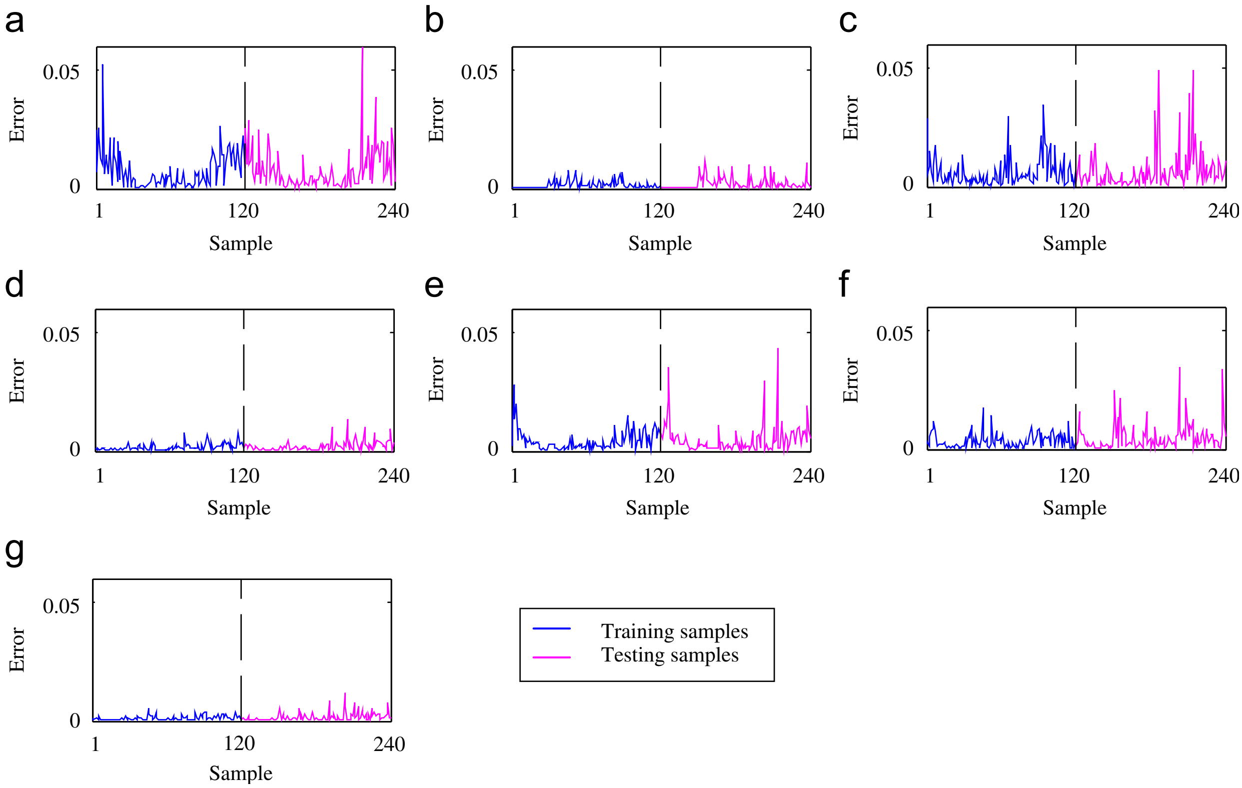

Fig. 5. Classification errors of individual classifiers and classifier combination: (a) classifier 1, (b) classifier 2, (c) classifier 3, (d) classifier 4, (e) classifier 5, (f) classifier 6, and (g) classifier combination.

Table 5
Performance comparison for different features

Feature	Classifier 1		Classifier 2		Classifier 3		Classifier 4		Classifier 5		Classifier 6		Average of six classifiers		Classifier combination	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
Salient features	65.67	61	90	87.67	72.67	68	80.33	77	67.67	67	87.33	81	77.28	73.61	93.67	91.33
Features selected randomly	62	58.03	69.27	66.93	58.5	55.1	73.73	66.3	56.5	50.93	57.37	50.8	62.89	58.02	81.19	77.15

there is a high degree of overlap between the values of these features between these ten classes. These features would confuse the classifiers and therefore cause the classification success to reduce clearly. These confirm our idea that the proposed feature selection method based on the improved distance evaluation technique can remove the redundant and irrelevant features from the original feature set, and therefore is a powerful feature selection method.

Besides, it can also be seen that in Table 5 and Fig. 6, although the six individual classifiers with the features selected randomly suffer on classification performance, 56.5–73.73% for training and 50.8–66.93% for testing, they are complementary. Thus, their combination also offers higher accuracies than any of them. Its training and testing accuracies increase to 81.19% and 77.15%, respectively. These highlight the proposed combination method is valid in the fault diagnosis of the rolling element bearings once again.

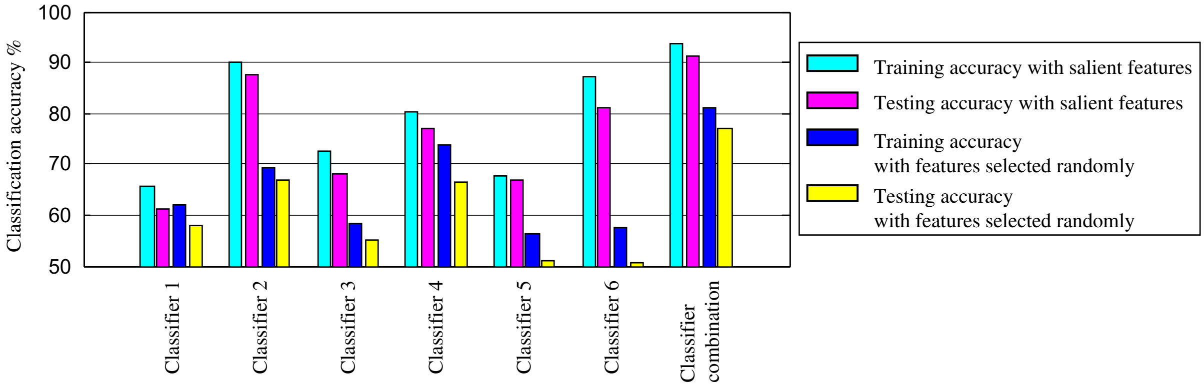

Fig. 6. Performance comparison for different features.

6. Conclusions

In this paper, a novel method for intelligent fault diagnosis of rotating machinery based on statistics analysis, empirical mode decomposition (EMD), the improved distance evaluation technique, adaptive neuro-fuzzy inference system (ANFIS) and genetic algorithms (GAs) is proposed. In order to derive more faulty information from vibration signals, several preprocessing methods, like EMD, filtration and demodulation, are performed.

In order to remove the redundant and irrelevant information and reduce the burden of the classification system, an improved distance evaluation technique is presented and used to select the salient features, and the application in rolling element bearings indicates the presented distance evaluation technique is very effective. Multiple ANFIS combination with GAs is adopted to construct a more reliable and intelligent fault diagnosis system of rotating machinery. The experimental results show that the proposed combination method based on multiple ANFISs enables the detection of abnormalities in bearings and at the same time identification of the category and severity of faults, and the fault diagnosis accuracy is higher than those of the individual ANFISs.

In this paper, a single output was used to represent multiple classes. It is more likely to produce false results. Presently, we are designing a multiple level fault diagnosis model to solve a large multiple class problem. We divide the large problem into several small multiple class problems, and in each small classification problem, the class encoding of a multiple output is adopted.

Acknowledgements

This work was supported by the key project of National Nature Science Foundation of China (no. 50335030) and National Basic Research Program of China (no. 2005CB724106).

References

[1] B.A. Paya, I.I. Esat, Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor, Mechanical Systems and Signal Processing 11 (1997) 751–765.
[2] B. Samanta, K.R. Al-Balushi, Artificial neural network based fault diagnostics of rolling element bearings using time-domain features, Mechanical Systems and Signal Processing 17 (2003) 317–328.
[3] X.S. Lou, K.A. Loparo, Bearing fault diagnosis based on wavelet transform and fuzzy inference, Mechanical Systems and Signal Processing 18 (2004) 1077–1095.
[4] M.L.D. Wong, L.B. Jack, A.K. Nandi, Modified self-organising map for automated novelty detection applied to vibration signal monitoring, Mechanical Systems and Signal Processing 20 (2006) 593–610.
[5] L.B. Jack, A.K. Nandi, Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms, Mechanical Systems and Signal Processing 16 (2002) 373–390.
[6] A.K.S. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mechanical Systems and Signal Processing 20 (2006) 1483–1510.

[7] D. Chen, W.J. Wang, Classification of wavelet map patterns using multi-layer neural networks for gear fault detection, Mechanical Systems and Signal Processing 16 (2002) 695–704.
[8] B. Samanta, Artificial neural networks and genetic algorithms for gear fault detection, Mechanical Systems and Signal Processing 18 (2004) 1273–1282.
[9] B.S. Yang, T. Han, J.L. An, ART-KOHONEN neural network for fault diagnosis of rotating machinery, Mechanical Systems and Signal Processing 18 (2004) 645–657.
[10] J.S.R. Jang, ANFIS: adaptive-network-based fuzzy inference system, IEEE Transactions on Systems, Man, and Cybernetics 23 (1993) 665–685.
[11] J. Kittler, M. Hatef, R.P.W. Duin, J. Matas, On combining classifiers, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (1998) 226–239.
[12] A. Rahman, M. Fairhurst, Decision combination of multiple classifiers for pattern classification: hybridization of majority voting and divide and conquer techniques, in: Proceedings of the Fifth IEEE Conference on Application of Computer Vision, Palm Springs, CA, USA, 2000, pp. 58–63.
[13] L.I. Kuncheva, A theoretical study on six classifier fusion strategies, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2002) 281–286.
[14] G. Fumera, F. Roli, A theoretical and experimental analysis of linear combiners for multiple classifier systems, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2005) 942–956.
[15] C. Dimitrakakis, S. Bengio, Online adaptive polices for ensemble classifiers, Neurocomputing 64 (2005) 211–221.
[16] Z.Q. Zhao, D.S. Huang, B.Y. Sun, Human face recognition based on multi-features using neural networks committee, Pattern Recognition Letters 25 (2004) 1351–1358.
[17] Ö. Toygar, A. Acan, Multiple classifier implementation of a divide-and-conquer approach using appearance-based statistical methods for face recognition, Pattern Recognition Letters 25 (2004) 1421–1430.
[18] D.J. Mashao, M. Skosan, Combining classifier decisions for robust speaker identification, Pattern Recognition 39 (2006) 147–155.
[19] X.C. Yin, C.P. Lin, Z. Han, Feature combination using boosting, Pattern Recognition Letters 26 (2005) 2195–2205.
[20] N.E. Huang, Z. Shen, S.R. Long, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proceedings of the Royal Society of London 454 (1998) 903–995.
[21] B.S. Yang, T. Han, W.W. Huang, Fault diagnosis of rotating machinery based on multi-class support vector machines, Journal of Mechanical Science and Technology 19 (2005) 846–859.
[22] K.A. Loparo, Bearings vibration data set, Case Western Reserve University, 〈http://www.eecs.cwru.edu/laboratory/bearing/download.htm〉.
[23] U. Benko, J. Petrovčič, D. Juričić, J. Tavčar, J. Rejec, An approach to fault diagnosis of vacuum cleaner motors based on sound analysis, Mechanical Systems and Signal Processing 19 (2005) 427–445.
[24] H. Zhu, X.L. Tang, Classifier geometrical characteristic comparison and its application in classifier selection, Pattern Recognition Letters 26 (2005) 829–842.
[25] H. Zouari, L. Heutte, Y. Lecourtier, Controlling the diversity in classifier ensembles though a measure of agreement, Pattern Recognition 39 (2006) 2195–2199.
[26] M. Aksela, J. Laaksonen, Using diversity of errors for selecting members of a committee classifier, Pattern Recognition 39 (2006) 608–623.
[27] Z. Michalewicz, Genetic Algorithms + Data Structures = Evolution Programs, third ed, Springer, New York, USA, 1999.