CrossMark

TECHNICAL PAPER

Efficient fault diagnosis of ball bearing using ReliefF and Random Forest classifier

V. Vakharia[1] · **V. K. Gupta[2]** · **P. K. Kankar[2]**

Abstract The present study focuses on identifying various faults present in ball bearing from the measured vibration signal. Features such as kurtosis, skewness, mean, and root mean square, and complexity measure such as Shannon Entropy are calculated from time domain and Discrete Wavelet Transform. To select the best wavelet function, Maximum Energy to Shannon Entropy ratio criterion is used. Information Gain and ReliefF ranking methods are used to assess the quality of features and features are ranked based on the weight gain obtained from the methods used. Support Vector Machine and Random Forest classifier are selected to identify bearing faults and comparison is made to diagnose faults on the ranked feature set. Experiments are conducted on Case Western Reserve University bearing data sets. Results show that ranking method is useful for identifying best feature set and to improve classification accuracy simultaneously. Cross-validation efficiency of 98.38% is obtained when ReliefF is used with Random Forest.

Keywords Fault diagnosis · Feature ranking · Bearings · Classification · Random Forest

✉ V. Vakharia
vinayvakharia4343@gmail.com

1 Department of Mechanical Engineering, PDPU, Gandhinagar 382007, India

2 Mechanical Engineering Discipline, PDPM IIITDM, Jabalpur 482005, India

1 Introduction

Most of the rotating equipment in industries has rolling element bearings as essential mechanical components to provide support and relative motion between parts. Safety and reliability are the major issues of concern for bearing applications. Though bearings are reliable rotary component, some failures are associated with it. Due to loading and high- speed rotation, wear occur in associated components of bearing, which lead to develop incipient faults. These incipient faults need to be detected at an early stage to avoid chances of catastrophic failure of machinery. Techniques such as vibration measurement, lubricant analysis, infrared thermography, and acoustic measurement are frequently used techniques to detect and diagnose the faults associated with bearing [1, 2]. Most of the researchers focused on vibration analysis of defective signals for detecting faults since vibration signals are related to the structural dynamics of the machine. It is observed that diagnostic information from vibration signal of faulty bearings can be obtained by applying signal processing techniques [3, 4]. Authors [5, 6] calculated the time domain statistical parameters from the healthy bearing and bearing with different faults. They used artificial intelligence (AI) techniques such as Artificial neural network (ANN) Support Vector Machine (SVM) and Random Forest for fault classification. Yang et al. [7] investigate the application of Random Forest as a classifier for machine fault diagnosis. They compared Random Forest with other classifiers, like SVM and ANN, and observed high accuracy from Random Forest classifier.

Nonlinearity due to varying stiffness and development of faults in bearings makes vibration signal to possess nonlinear characteristics [8, 9]. Therefore, features from time

⚙ Springer

domain signals are not efficient alone since these features calculated are masked by noise. In these circumstances, researchers focused their attention toward advanced signal processing techniques such as Wavelet Transform (WT) [10, 11]. To identify the type of faults, Nikolaou and Antoniadis [12] have used wavelet packet transform. Authors found that time–frequency localization capabilities of wavelet packet transform makes it an efficient method to identify the nature of rolling element bearing faults. Further, Discrete Wavelet Transform (DWT) was used by Prabhakar et al. [13] for detecting single and multiple faults in bearing races. While, using Wavelet Transform challenge is to choose most appropriate wavelet. Therefore, mother wavelet selection methodologies were proposed based on Maximum Energy to Shannon Entropy ratio, Maximum Relative wavelet Energy, and Multiscale Permutation Entropy [14–16].

The information contained by various features extracted from signals is an important issue in the area of machine learning. To enhance classification accuracy, feature selection procedure is applied by researchers in various applications. The goal is to select most informative features based on feature ranking and discard irrelevant features so that classification accuracy is enhanced with minimum feature subset. Fisher Score and Mahalanobis Distance technique were employed by Wu et al. [17] to select the top-ranked feature so that classification accuracy is improved. Zheng et al. [18] used another feature ranking technique, Laplacian score, for identifying informative features from

the various faults associated with the bearing. Kappaganthu and Nataraj [19] calculated statistical features from time domain, frequency domain, and time–frequency domain and, utilized mutual information (MI) technique to choose feature sets and realized that using feature ranking techniques classification accuracy can be enhanced considerably.

In the present study, a generalized approach has been proposed to select an optimal number of the feature set using Information Gain (IG) and ReliefF (RF) feature ranking techniques to select significant features calculated from time domain and Discrete Wavelet Transform. The advantage of feature ranking method is that it is independent of classifier used and the features are selected based on their ranking. Proposed feature ranking techniques are applied to the Case Western Reserve University (CWRU) bearing data set for testing the efficacy of proposed methodology. The combination of feature ranking methods and classifiers are used to select the optimum number of feature set so that maximum efficiency is obtained with reduced number of feature set. Considering the reduced scope of speeds and loading conditions, the optimum number of feature set and the features calculated may differ for a different data set. In most of the previous studies, features are divided into training and testing set; therefore, results reported may carry a statistical biasedness. In the present study, the robustness of proposed technique is evaluated using 10-fold cross-validation which is the standard method of testing classifiers and gives statistically unbiased results. The result shows

Fig. 1 Proposed methodology for fault diagnosis of bearing classes
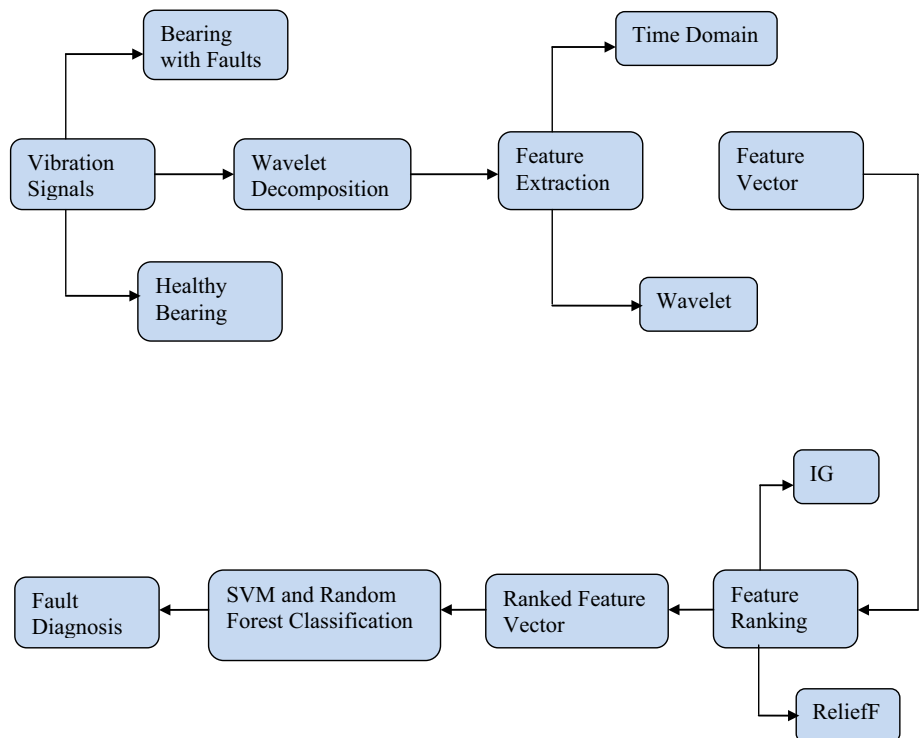
improvement in classification efficiency with reduced feature set. Figure 1 shows the flowchart of the methodology proposed.

2 Machine learning techniques

2.1 Support Vector Machine

Support Vector Machine is a type of supervised learning algorithm mainly used for classification and regression. The theories of Support Vector Machines have been described by Vapnik [20]. Due to its high accuracy and good generalization capability even when the samples are few, some researchers [8, 15] had employed the SVM as a tool for the classification of mechanical faults in bearing. The formulation of SVM is based on the principle of structural risk minimization. For binary classification problem, the aim is to maximize margin between the separating planes. The maximum margin which separates the hyperplane (H_1) can be used to classify data sets into the classes consider. The equation for H_1 can be written as

$$x.w + b = 0, \tag{1}$$

here x is a point that lies on a separating plane (H_1) and w is a vector perpendicular to the plane. The normalization of parameter w for two classes can be represented as

$$x_i \cdot w + b \leq -1 + \xi_i \quad \text{for } y_i = -1 \tag{2a}$$

and

$$x_i \cdot w + b \geq +1 - \xi_i \quad \text{for } y_i = +1. \tag{2b}$$

On combining Eqs. (2a) and (2b) we get

$$y_i(x_i.w + b) \geq +1 - \xi_i, \tag{3}$$

here ξ_i represents slack parameter.

To evaluate the performance of SVM, Meyer et al. [21] conducted an extensive study and concluded that the SVM is best for fault classification problems. In the present study, the sequential minimal optimization (SMO) algorithm is used with ten-fold cross-validation method. Pearson VII kernel function (PUK) kernel is chosen, and penalty parameter C is set to 10. Due to better generalization capability, SVM as an algorithm for fault detection systems is of great interest for academic and industrial societies.

2.2 Random Forest

Random Forest is a type of Artificial intelligence technique to identify the state of machinery component. The Random Forest algorithm was developed by Breiman [22] and is based on building a decision tree. In the initial stage, the training set consisting of features is divided into the in-bag and out-bags set. The method of bootstrapping is repeated several times on feature set to produce several in-bag sets and out-bag set subsets. A Decision tree is modeled for each in-bag data set, and the out-of-bag set is used for evaluating the classification accuracy of each decision tree. The final outcomes based on algorithm are obtained from out-bag sets from the entire training dataset. Every decision tree casts a vote for one class, and this vote can be used to estimate the generalization capability of the classifier. The class from the feature set is recognized by gaining maximum vote [23]. The Random Forest error rate depends on the correlation between any two trees in forest and strength of each tree in the forest. Increasing the correlation increases the forest error rate. On the other hand, a tree with a low error rate is a strong classifier.

3 Feature selection

Feature selection is required to choose a small subset of features from original feature subset to reduce dimensionality without compromising information contained. To make a decision which feature has to retain and which feature has to discard depends solely on the technique which has to be applied. By eliminating redundant features, the performance of classifier improves. To eliminate an irrelevant feature, a feature selection criterion is desirable which can measure the significance of each feature in a feature set with respect to class labels. Feature-ranking techniques, such as Fisher Score [24], ReliefF [25], Information Gain [26], and dimensionality reduction technique as Principal Component Analysis (PCA) [27], are widely used techniques applied to a variety of problems. The advantage of features selected by applying feature-ranking techniques is to maintain the original physical structure without the space transformation as in PCA. Feature-ranking methods aim not only to reduce dimensionality but at the same time superior feature separability and conserving required information. In this study, two feature-ranking criteria are compared namely: Information Gain and ReliefF. These criteria are used for selecting features having high distinguishability with an optimum number of feature sets for obtaining maximum accuracy.

3.1 Information Gain

In Information Theory, Information Gain (IG) is associated to entropy which measures the unpredictability of system. When we apply Shannon Entropy to Random variable M [28] then

$$E(M) = -\sum_{i=\text{M}} p_i(M) \cdot \log_2(p_i(M)), \tag{4}$$

2972

J Braz. Soc. Mech. Sci. Eng. (2017) 39:2969–2982

Fig. 2 Bearing test rig
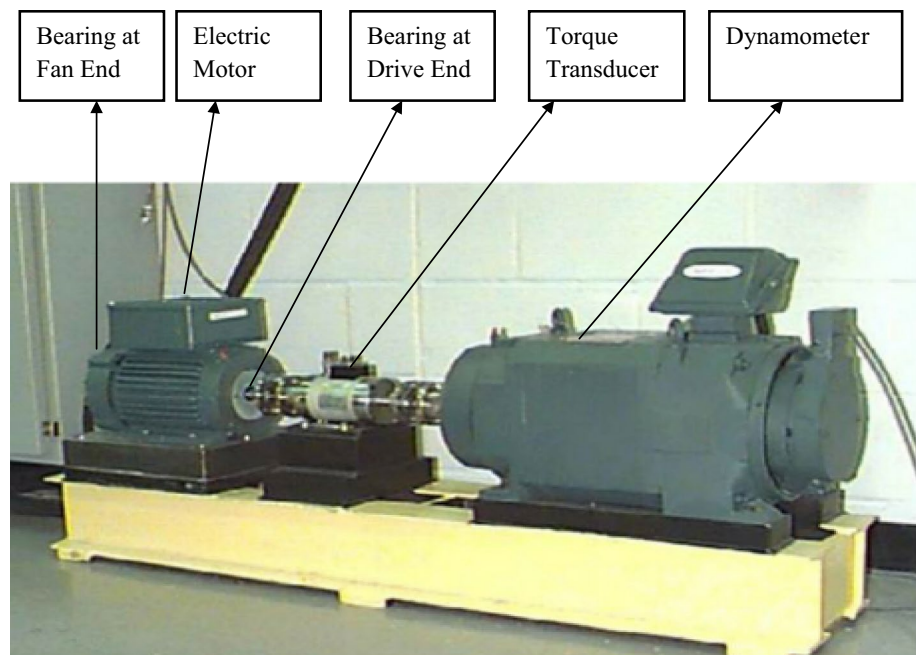
here $p_i(M)$ is probability density function of Random variable M; then, entropy of M after observing N is

$$E(M/N) = - \sum_{i=N} p_i(N) \cdot \sum_{i=M} p_i(M/N) \cdot \log_2(p_i(M/N)), \quad (5)$$

here $p_i(M/N)$ is conditional probability of M given N. Information Gain is given by

$$IG = E(M) - E(M/N). \quad (6)$$

Information Gain measures a reduction in uncertainty about M due to the knowledge on a feature N which is measured by entropy change.

3.2 ReliefF

Sikonja et al. [29] used ReliefF (RF) in their study as a feature subset selection method and found that it is a powerful attribute estimator method which can be applied to solve a variety of classification problems. RF computes the weight of feature W_i from the feature set X_i. Let *NH* and *NM* represent the nearest hit and nearest miss from the same class and opposite class, respectively. The instances which are closer within class are known as the nearest hit, and the closest different class instance is known as the nearest miss. Weight can be computed by [30]

$$W_i = W_i + \varepsilon_0 |X_i - NM_i| - \varepsilon_1 |X_i - NH_i|. \quad (7)$$

The weight of feature depends on the weight gain by that feature in nearby instances of the same class. ReliefF uses $\varepsilon_0 = \varepsilon 1 = 1$ which implies that within class conservation and inter-class divergence are weighted equally.

Table 1 Specification of bearing 6205 (drive end)

Bearing type	Inside diameter (mm)	Outside diameter (mm)	Ball diameter (mm)	Pitch diameter (mm)
6205	25	52	7.94	39

4 Experimental procedure

The experimental data set used in present study is made available by Case Western Reserve University Bearing Data Center [31]. As shown in Fig. 2, setup consists of 2HP three-phase induction motor, a dynamometer connected with the coupling device, and a torque transducer. 7 Ball bearings type 6205 is used as test bearing whose specification is shown in Table 1. The data set belongs to 12 K drive end bearing due to the broader variety of fault sets. Faults are introduced with the help of electric discharge machining with defect diameter 0.1778, 0.3556, 0.5334, and 0.7112 mm. Vibration signals are collected with four different conditions: Healthy bearing (HB), Inner race fault (IRF), Outer race fault (ORF), and Ball fault (BF). To record signal, an accelerometer was mounted at the drive end. Vibration signals are recorded at various rotational speed as 1725, 1748, 1772, and 1796 rpm, and the sampling frequency was set to 12 kHz.

4.1 Feature extraction/calculation

Vibration signals acquired with the help of sensors from bearings are highly nonlinear in nature. It is, therefore,

necessary to use proper signal processing techniques to extract useful information about the healthy or faulty status of the component. Generally, features are extracted from vibration signal using time domain, frequency domain (FFT), and time–frequency domain (Wavelet) [32]. Time domain signal can be directly analyzed by looking at its pattern and has the advantage of simple calculations. Time domain features are directly calculated from time waveform of the signal. Kurtosis, Skewness, Root Mean Square, Mean Value, and Shannon Entropy are frequently used features which are calculated from vibration signals. The drawback of features calculated by time domain method is that it is unable to detect faults at an early stage. Further, the statistical parameters calculated from time domain signal could not identify faults in bearing since vibration signals were masked by noise. Frequency domain method is another technique for fault diagnosis of bearing. Vibration signal can be of stationary or non-stationary nature. When signal statistics do not change with respect to time then it will be categorized as a stationary signal. In case of non-stationary signal, statistical properties change with time. Frequency domain method is suitable for analyzing stationary signal. Due to a narrow range of features, available FFT-based features are reported less.

Methods such as Wavelet Transform (WT), Short-Time Fourier Transform (STFT), and Wigner-Ville Distribution (WVD) emerged as a potential technique for analyzing non-stationary signals [32]. As compared to FFT, Wavelet Transform decomposes a signal into both time and frequency simultaneously. STFT is similar to WT in operation, but the difference lies in using window function. In STFT, fixed window function is used as a result both time and frequency resolution will be fixed. Whereas in WT, varying window functions are used which make it useful for detecting the impulses present in the signal. Discrete Wavelet Transform (DWT) is useful to obtain time–frequency information from the signal. With the help of DWT, the impulses occuring due to the presence of defect can be detected efficiently [13].

Kim et al. [33] validated the importance of calculating features from DWT by a comparative study. Methods compared were Short-Time Fourier Transform (STFT) and Wigner-Ville distribution (WVD). A detailed description of applications of wavelet for fault diagnosis can be found in [8, 10]. Detail list of time domain and DWT-based features calculated in the present study are listed in Table 2. The feature vector constructed consists of 62 instances and 35 attributes. The overall size of feature vector is 2170.

4.2 Wavelet selection

Wavelet Transform-based features are found useful in detecting abrupt changes from the measured vibration signal. The advantages of wavelet transform for bearing fault detection is the availability of several base wavelets functions developed over the past decades. To extract the fault feature of signals, an appropriate wavelet-base function should be selected. To calculate DWT-based features from detail and approximation coefficients from bearings signals, Maximum Energy to Shannon Entropy ratio criterion (MESE) [14, 15] is used for selecting the base wavelet in the present study.

A wavelet is selected as the base wavelet, when it extracts maximum amount of energy from the measured signal and simultaneously minimizing the Shannon Entropy of the corresponding wavelet coefficients [14]. The wavelets compared in this study are Daubechies (db1), Symlet (sym2), Coiflet (coif1), and reverse Biorthogonal (rbio1.1). Bearings with faults in outer race, inner race, and ball are considered in the present study to select wavelet at speed 1730, 1750, 1772, and 1797 rpm. To convert bearing signals into the wavelet coefficients, these raw signals are decomposed into detail and approximation coefficients using 'dwt'function. Signal is decomposed up to three levels, and statistical features are calculated. It is clear from Fig. 3 that Coiflet Wavelet gives Maximum Energy to Shannon Entropy ratio and thus it is chosen to calculate DWT-based statistical features. Figure 3 depicts MESE value with inner race fault (IRF), outer race fault (ORF), and ball fault (BF) with corresponding rotational speed. It is evident that for all the wavelet considered, MESE value for IRF is maximum, in almost all the cases considered, and for BF,

Table 2 Features considered in study

Domain	Features				
	Kurtosis	Skewness	RMS	Mean	Shannon entropy
Time domain	KT	ST	RT	MT	SET
DWT detail level 1	KD1	SD1	RD1	MD1	SED1
DWT approximation level 1	KA1	SA1	RA1	MA1	SEA1
DWT detail level 2	KD2	SD2	RD2	MD2	SED2
DWT approximation level 2	KA2	SA2	RA2	MA2	SEA2
DWT detail level 3	KD3	SD3	RD3	MD3	SED3
DWT approximation level 3	KA3	SA3	RA3	MA3	SEA3

2974

J Braz. Soc. Mech. Sci. Eng. (2017) 39:2969–2982

Fig. 3 Maximum Energy to Shannon Entropy ratio (MESE) for wavelets at rotational speed
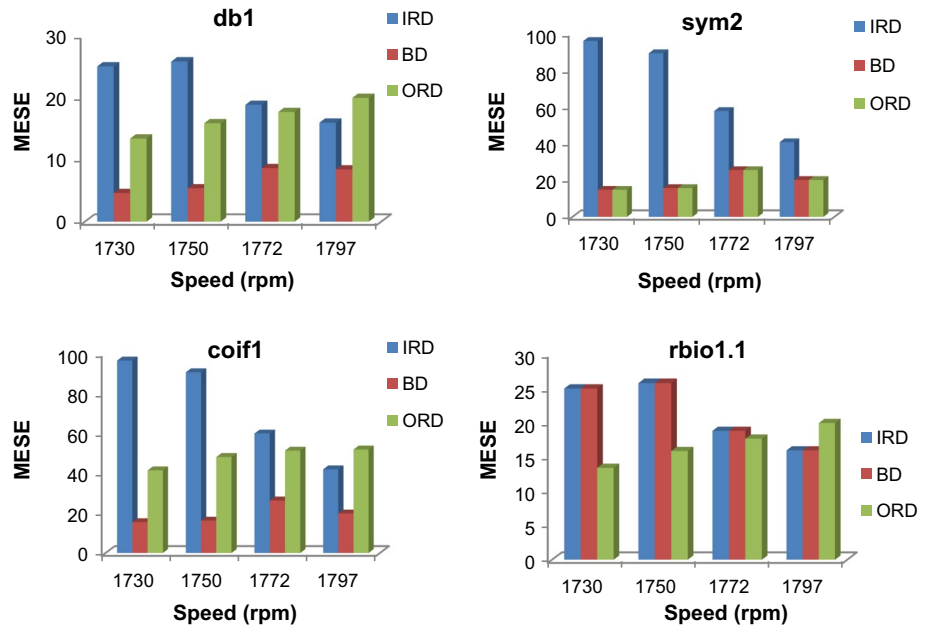
Table 3 Average MESE values for 0.7112 mm depth fault for wavelets

	db1	sym2	coif1	rbio 1.1
MESE	25.05	95.95	96.61	25.05
	25.83	89.24	90.74	25.83
	18.84	57.90	60.11	18.84
	15.97	40.86	42.28	16.0
	4.62	14.90	15.54	25.05
	5.39	15.86	16.27	25.83
	8.69	25.52	26.53	18.84
	8.48	20.31	19.9	16.0
	13.47	40.23	41.73	13.47
	15.88	47.96	48.45	15.88
	17.69	50.32	51.52	17.69
	19.98	51.79	52.04	20.0
Average value	14.997	45.909	46.812	19.873

MESE value is lowest except for biorthogonal wavelet. This can be inferred that IRF is severe as compared to other faults consider in the present study. The average MESE values for wavelets considered for 0.7112 mm depth fault are shown in Table 3.

5 Results and discussion

In the present study, statistical features are obtained from time domain and DWT. A sample time response and spectrum (FFT) of vibration signals for healthy bearing, bearing with a fault in an inner race (0.1778 mm), bearing with ball fault (0.1778 mm), and bearing with fault at the outer race (0.1778 mm) are shown in Fig. 4a–d. In Fig. 4a, the time response and frequency response in terms of FFT, for healthy bearing at 1797 rpm are shown. From FFT, it is observed that the peak amplitude was observed at Varying Compliance Frequency (VC) and their multiples. The Varying Compliance Frequency measured for healthy bearing at 1797 rpm is 87.51 Hz. Figure 4b shows the response of inner race fault at 0.1778 mm defect condition at a rotating speed of 1797 rpm. It is observed that the wave passage frequency on inner race (ω_{bpfi}) was 154 Hz for bearing with inner race defect. The peak amplitude appears at super-harmonic of wave passage frequency ($4\omega_{bpfi} = 616$ Hz). It is observed that time waveform is aperiodic in nature. When ball fault is considered at 0.1778 mm defect, aperiodic time waveform appears in Fig. 4c. The ball passage frequency (ω_{bpfs}) corresponding to defect in a ball of the roller bearing is 98 Hz. The peak amplitude appears at ($7\omega_{bpfs} = 686$ Hz) in frequency spectra. The measured vibration response for outer race defect is shown in Fig. 4d. The peak in amplitude appears at ($6\omega_{bpfo} = 684$ Hz). After analyzing the frequency spectrum and peak amplitude, it is observed that the impact of a fault in an outer race is more severe as compared to a fault in an inner race which is followed by a fault in a ball.

In total, thirty-five features based on the time domain and Discrete Wavelet Transform are measured from vibration signals and used as an input to machine learning algorithms for classification of faults [34]. To optimize the feature set two feature ranking algorithm viz. Information Gain and ReliefF are used to reduce the order of feature set. Tables 4 and 5 show the feature ranking of the
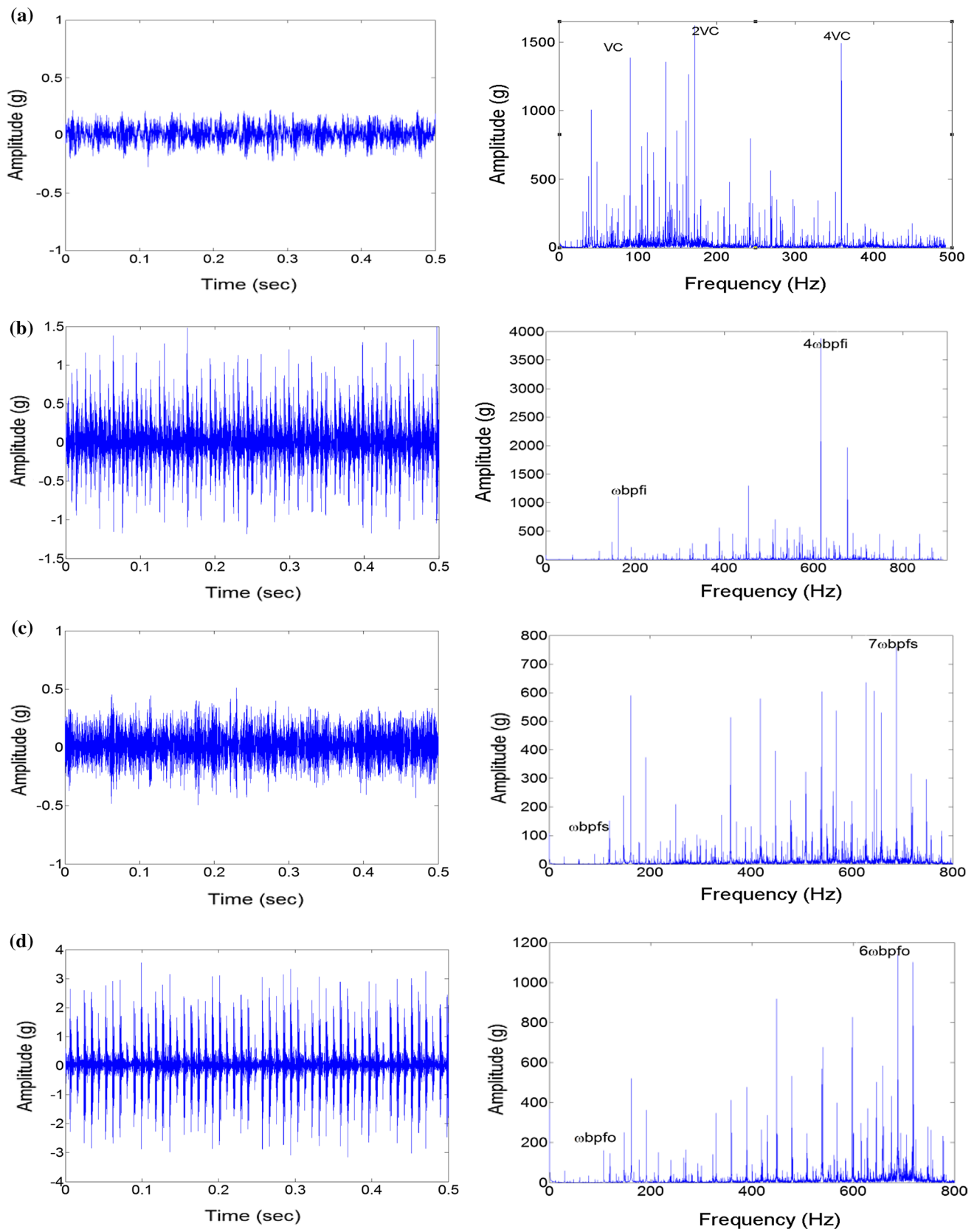
Fig. 4 Time response and frequency response for different bearing conditions **a** healthy bearing, **b** inner race fault, **c** ball fault, **d** outer race fault

Table 4 Information Gain-based feature ranking

Order	Ranking	Ranked feature	Selected feature	Cross-validation efficiency (%)	
				IG-SVM	IG-Random Forest
10	1	SEA1	1	48.38	69.35
3	2	RT	1–2	61.30	90.32
23	3	RD2	1–3	59.67	90.32
13	4	RD1	1–4	64.51	91.93
15	5	SED1	1–5	64.51	93.54
8	6	RA1	1–6	66.13	91.93
7	7	SA1	1–7	77.42	91.93
25	8	SED2	1–8	77.42	**96.77**
11	9	KD1	1–9	75.80	96.77
24	10	MD2	1–10	79.02	96.77
12	11	SD1	1-11	87.09	93.54
21	12	KD2	1–12	87.09	90.32
6	13	KA1	1–13	88.70	87.09
1	14	KT	1–14	**90.32**	88.70
26	15	KA3	1–15	90.32	91.93
30	16	SEA3	1–16	87.09	91.93
20	17	SEA2	1–17	87.09	91.93
35	18	SED3	1–18	88.02	87.09
9	19	MA1	1–19	85.48	88.70
5	20	SET	1–20	85.48	90.32
2	21	ST	1–21	83.87	91.93
4	22	MT	1–22	80.64	95.16
31	23	KD3	1–23	80.64	88.70
28	24	RA3	1–24	82.25	88.70
29	25	MA3	1–25	79.03	93.54
34	26	MD3	1–26	75.80	88.70
32	27	SD3	1–27	75.80	93.54
33	28	RD3	1–28	75.80	91.93
17	29	SA2	1–29	75.80	88.70
14	30	MD1	1–30	70.96	83.87
16	31	KA2	1–31	72.58	83.87
22	32	SD2	1–32	69.35	80.64
27	33	SA3	1–33	77.42	87.09
18	34	RA2	1–34	75.80	85.48
19	35	MA2	1–35	75.80	83.87

calculated features. For Information Gain-based feature ranking, Shannon Entropy at DWT approximation level 1 is found to be most informative features among the whole feature set. The reason lies in the fact that Shannon Entropy measures the disorders present in the signal. It also represents status of signal with considerable high information at DWT approximation level 1. RMS value at time domain is found to be the second most feature, since it represents the average power of measured signal and is weighted high as compared to the other feature. It is observed from Table 4 that features calculated from DWT secure top ranking based on IG selection criterion. In the present study, the feature ranking not only depends on the domain from which features are calculated but also depends on the weight assigned from the ranking method.

The 10-fold cross-validation efficiency of SVM classifier of ranked feature based on Information Gain (IG-SVM) is shown in Fig. 5. When SEA1 alone was used for classification than 49% cross-validation efficiency is obtained. As the numbers of features are increasing, the efficiency is increasing and the peak efficiency of 90.3226% is achieved when top fourteen ranked features are used. A very interesting fact is observed from Fig. 5 that as the number of features increases then efficiency goes on decreasing. This gives information that only fourteen top-ranked features are sufficient enough for achieving highest efficiency based on the classifier used. To compare the accuracy of Information Gain criterion one more classifier, Random Forest, is utilized. The cross-validation efficiency of Random Forest classifier based on Information Gain (IG-Random Forest) is shown in Fig. 6. When SEA1 alone was used as a feature than 69.3548% cross-validation efficiency is achieved. Further, when SEA1 together with RT was used then 90.3226% cross-validation efficiency is achieved. This suggests that only two features are sufficient to achieve accuracy of 90.3226% which is equivalent to the efficiency achieved when top fourteen ranked feature are used with SVM as a classifier. Therefore, it is concluded that feature ranking alone is not sufficient for fault classification but its combination with the classifier is also significant.

To further show the utilization of feature ranking method for fault classification, one more ranking algorithm ReliefF (RF) is used, and the ranked feature list is shown in Table 5. Mean value obtained from DWT detail level 2 becomes the most significant feature among thirty-five measured features from RF ranking method. The weight achieved by mean value from RF suggests that this feature carries most information content from all the classes considered in the present study, since mean value represents the average amplitude of the signal over time. The second-ranked feature is skewness at DWT detail level 1. It is observed from Table 5 that DWT-based measured features are more significant as compared to the time domain features. The same observation is also found in Table 4. Using the ranked feature set from RF, two classifiers are used for comparing classification accuracy with selected features. Figure 7 shows the cross-validation efficiency obtained from SVM as a classifier and ReliefF as ranking method (RF-SVM). It is observed that the maximum cross-validation efficiency 91.9355% is obtained when top twelve features are selected. The lowest classification accuracy obtained is 46.7742% when single top-ranked feature is used.

Table 5 ReliefF-based feature ranking

Order	Ranking	Ranked feature	Selected feature	Cross-validation efficiency	
				RF-SVM	RF-Random Forest
24	1	MD2	1	46.77	46.77
12	2	SD1	1–2	50.00	45.16
7	3	SA1	1–3	56.45	62.90
11	4	KD1	1–4	74.19	64.51
21	5	KD2	1–5	74.19	75.80
1	6	KT	1–6	75.80	77.41
6	7	KA1	1–7	75.80	69.35
13	8	RD1	1–8	80.64	79.03
3	9	RT	1–9	79.03	88.70
8	10	RA1	1–10	82.25	*98.38*
2	11	ST	1–11	88.70	95.16
23	12	RD2	1–12	*91.93*	96.77
16	13	KA2	1–13	91.93	96.77
18	14	RA2	1–14	90.32	90.32
31	15	KD3	1–15	90.32	93.54
26	16	KA3	1–16	87.09	93.54
28	17	RA3	1–17	85.48	91.93
33	18	RD3	1–18	85.48	93.54
10	19	SEA1	1–19	88.70	88.70
25	20	SED2	1–20	88.70	93.54
5	21	SET	1–21	88.70	95.16
15	22	SED1	1–22	88.70	93.54
20	23	SEA2	1–23	88.70	93.54
29	24	MA3	1–24	85.48	93.54
32	25	SD3	1–25	85.48	93.54
35	26	SED3	1–26	85.48	91.93
30	27	SEA3	1–27	83.87	93.54
22	28	SD2	1–28	80.64	96.77
4	29	MT	1–29	77.41	91.93
19	30	MA2	1–30	75.80	91.93
9	31	MA1	1–31	75.80	91.93
17	32	SA2	1–32	77.41	90.32
27	33	SA3	1–33	82.25	95.16
14	34	MD1	1–34	77.41	93.54
34	35	MD3	1–35	75.80	91.93

When KD1 is used along with previous top three ranked feature, there is a considerable increase in validation efficiency suggesting that the combination of ranked feature enhanced the classification accuracy. When numbers of features are increasing beyond twelve features, then efficiency decreases. It can be judged that only twelve features are sufficient enough to get the highest accuracy from SVM classifier and ReliefF ranking method (RF-SVM). The ReliefF ranking method is compared with Random Forest for fault classification. The highest cross-validation efficiency achieved is 98.3871% with top ten ranked features and the lowest cross-validation efficiency achieved

is 45.1613% with top two ranked features. It is observed from Fig. 8 that just top ten features are sufficient enough for achieving highest classification accuracy when ReliefF is combined with Random Forest classifier (RF-Random Forest). Table 6 shows the cross-validation efficiency-based confusion matrix obtained from Information Gain when SVM and Random Forest are used. From SVM 15, 15, 22, 4 instances are identified correctly for IRF, BF, ORF, and HB classes. The accuracy obtained is 90.3226% with fourteen features. Similarly, for Random Forest 16, 15, 26, 4 instances are identified correctly for IRF, BF, ORF and HB classes. The accuracy obtained is 96.7742% with eight

2978

J Braz. Soc. Mech. Sci. Eng. (2017) 39:2969–2982

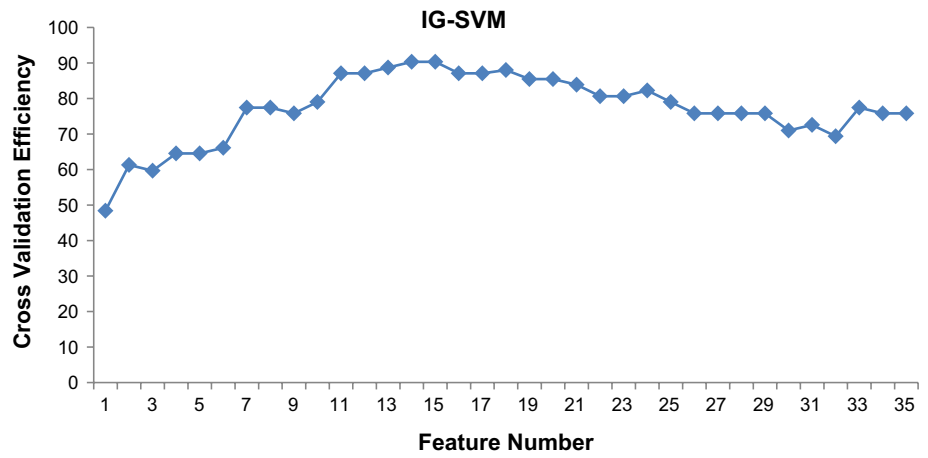**Fig. 5** Cross-validation efficiency of SVM classifier based on Information Gain

Fig. 6 Cross-validation efficiency of Random Forest classifier based on Information Gain
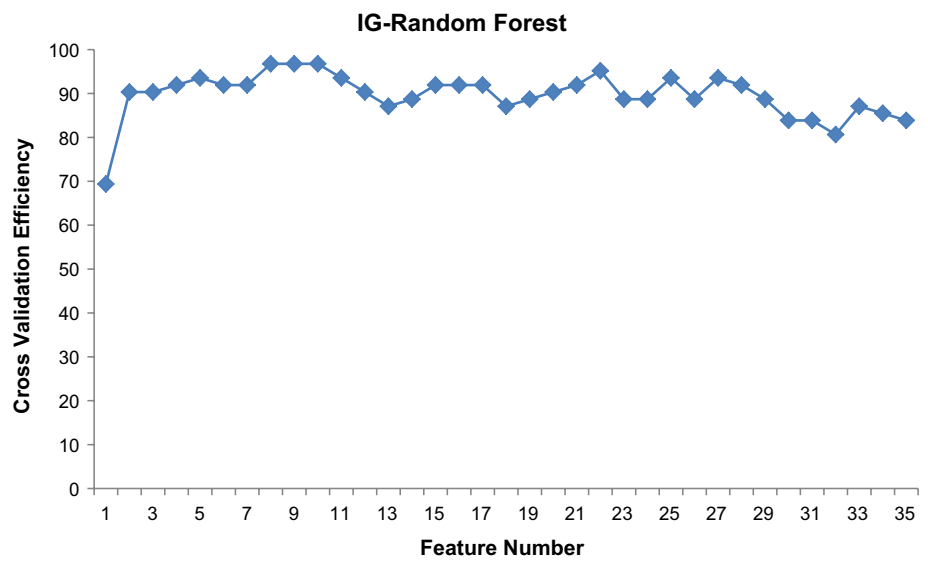

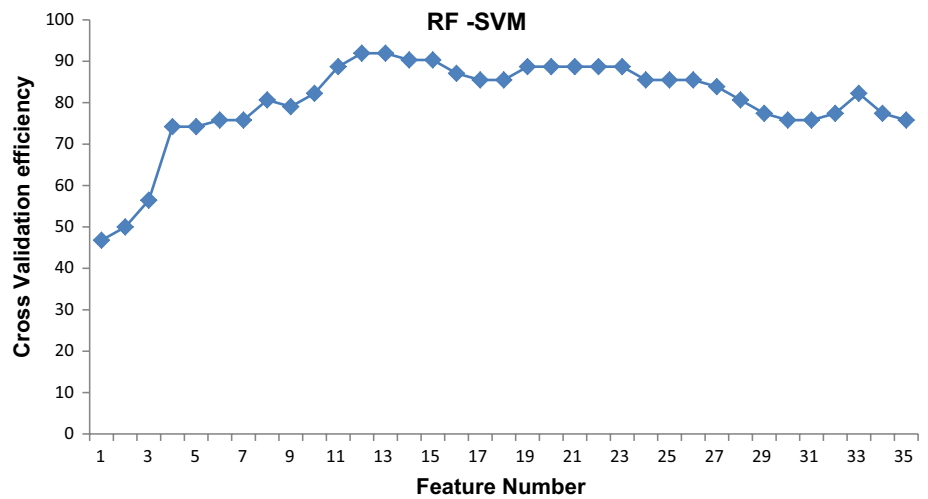**Fig. 7** Cross-validation efficiency of SVM classifier based on ReliefF

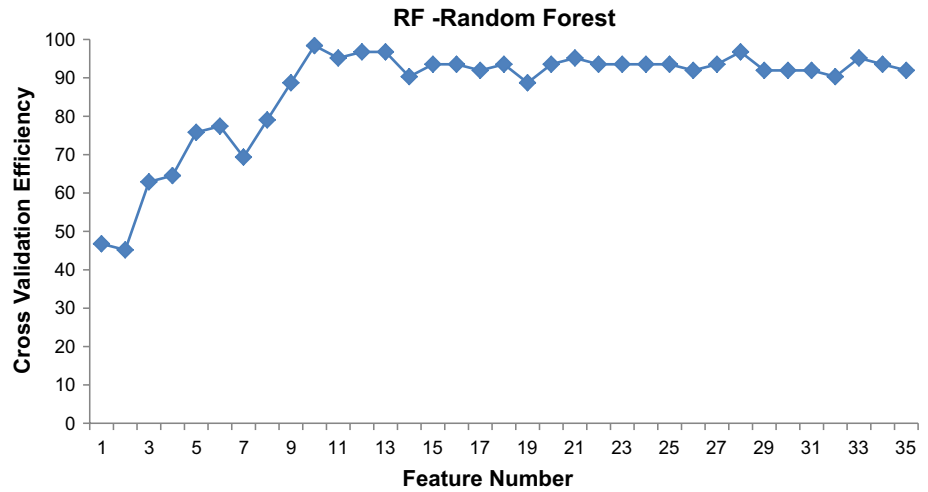**Fig. 8** Cross-validation efficiency of Random Forest classifier based on ReliefF

Table 6 Information Gain confusion matrix (SVM and Random Forest)

IG-SVM				Classified as	IG–Random Forest			
IRF	BF	ORF	HB		IRF	BF	ORF	HB
15	0	1	0	IRF	16	0	0	0
0	15	1	0	BF	0	15	1	0
0	4	22	0	ORF	0	0	26	0
0	0	0	4	HB	0	0	0	4
Accuracy–90.3226% (14 Features)					Accuracy–96.7742% (8 Features)			

Table 7 ReliefF confusion matrix (SVM and Random Forest)

SVM				Classified as	Random Forest			
IRF	BF	ORF	HB		IRF	BF	ORF	HB
15	0	1	0	IRF	15	0	1	0
0	16	0	0	BF	0	16	0	0
0	4	22	0	ORF	0	0	26	0
0	0	0	4	HB	0	0	0	4
Accuracy–91.935% (12 Features)					Accuracy–98.3871% (10 Features)			

features. It is clear from Table 6 that Random Forest is best classifier for Information Gain-based feature-ranking method. Table 7 shows cross-validation efficiency-based confusion matrix obtained from ReliefF when SVM and Random Forest are used. From SVM correctly identified instances are 15, 16, 22, and 4 for IRF, BF, ORF, and HB classes. Inner race fault and healthy bearings are identified exactly where as outer race fault is identified least. From Random Forest, correctly identified instances are 15, 16, 26, and 4 for IRF, BF, ORF and HB. Ball fault, outer race fault, and healthy bearing are identified exactly where as inner race fault is identified least.

Figure 9 shows the class identification rate through SVM and Random Forest using two ranking methods. Based on Information Gain ranking method shown in Fig. 9a the

classification accuracy of Random Forest (Eight features) is found to be efficient as compared to SVM (Fourteen features). Random Forest correctly predicted IRF, ORF, and HB classes, whereas SVM correctly predicted HB class only. For ReliefF-based ranking, Fig. 9b correctly identified classes are BF, ORF, and HB with Random Forest while SVM identified BF and HB classes correctly. Thus, the average class prediction rate with Random Forest classifier is higher than with SVM classifier. Table 8 shows the numeric prediction rate of SVM and Random Forest classifier. It can be concluded that for RF-Random Forest 98.3871% cross-validation efficiency is achieved which is a promising result based on the present methodology. Also, ten features are sufficient enough to gain the highest accuracy with ReliefF and Random Forest. Table 9 shows a
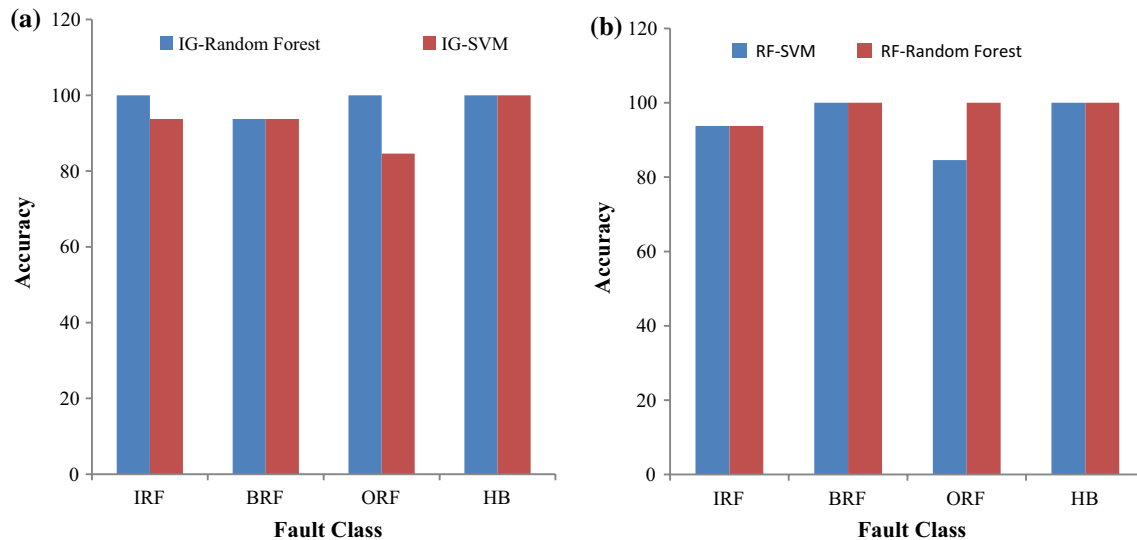
2980

J Braz. Soc. Mech. Sci. Eng. (2017) 39:2969–2982

Fig. 9 Cross-validation accuracy based on class

Table 8 Numeric prediction rate

Parameters	IG-SVM	IG-Random Forest	RF-SVM	RF–Random Forest
Instances which are correctly classified	56	60	57	61
Instances which are incorrectly classified	6	2	5	1
Kappa value	0.8135	0.9527	0.8839	0.9764
Time (s)	0.16	0.01	0.05	0.01
Cross-validation efficiency (%)	90.3226	96.7742	91.9355	98.3871

Table 9 Comparison table demonstrating significance of present study with published literature

Authors	Number of feature used	Feature selection technique	Classifier used	Peak efficiency	Remark
Kavathekar et al. [6]	Ten	Not applied	Seven classifier	75%	Training efficiency
Wu et al. [17]	Eighty (statistical)	Fisher score and Mahalanobis distance	SVM	98.5% (Ten features with Mahalanobis distance)	Training efficiency
Wang et al. [35]	Hundred	Statistical locally linear embedding	CART,KNN and SVM	97.26%	----
Li et al. [36]	One hundred eighty	Laplacian score	SVM-BT	100%	Training efficiency
Zhang and Li [37]	Twenty	PCA, LPP, LDA, and ISOP	Neighborhood preserving embedding and SOM	99%	Training efficiency
Present work	Thirty-five (statistical)	ReliefF and Information Gain	SVM and Random Forest	98.38% (Ten features with Random Forest classifier)	Ten-fold cross-validation efficiency

comparative study based on available literature demonstrating the effectiveness of present methodology.

In general, Random Forest achieves maximum accuracy when used with IG and RF due to less generalization error. Breiman [22] suggested that the generalization error in

Random Forest classifier is due to the correlation between any two trees in the forest and due to the strength of individual tree in the forest. It was observed that a tree with low error rate emerges as a strong classifier. On the other hand, decreasing correlation among different classes decreases

the forest generalization error. These can be a reason of better classification from Random Forest classifier. Another reason is only one variable, i.e., number of trees need to be set by user thus complexity achieved by classifier decreases. The disadvantage of SVM is that basically, SVM is a binary classifier and for conducting multi-class classification one against all procedure can be used. Thus, the average results obtained are computationally expensive and may contain biasedness. This can be a reason of the inferior performance of SVM as compare to Random Forest.

6 Conclusion

This study presents the comparison of two ranking methods Information Gain and ReliefF for ranking features. Features are calculated from vibration signals obtained from CWRU-bearing data center. The feature set obtained consists of five time domain features and thirty DWT-based features. To obtain classification accuracy, cross-validation technique is used for SVM and Random Forest as a classifier. For fault classification, optimum feature selection is an important task which is reported less in the literature. In the experiments conducted, ReliefF is found to be efficient feature-ranking method when it is used with Random Forest classifier. An insight is obtained after conducting an experiment about a number of features necessary for getting highest accuracy from the feature set. Till now, few researchers focused on reducing feature set and improving classification accuracy for fault diagnosis of bearing which is shown in Table 9. The importance of present study lies in comparing ranking methods and using them with classifiers which are reported less in the literature. Further, the results obtain depends on the data sets used and the methodology adopted for calculating statistical features. Experimental results show that the RF–Random Forest method can choose an optimum set of features with high cross-validation efficiency of 98.3871% as compared to IG-SVM, IG-Random Forest, and RF-SVM. The result obtained shows the efficacy of proposed methodology.

References

1. Tandon N, Yadava GS, Ramakrishna KM (2007) A comparison of some condition monitoring techniques for the detection of defects in induction motor ball bearings. Mech Syst Signal Process 21:244–256

2. Dong S, Xu X, Chen R (2015) Application of fuzzy C-means method and classification model of optimized K-nearest neighbor for fault diagnosis of bearing. J Braz Soc Mech Sci Eng. doi:10.1007/s40430-015-0455-9

3. Singh S, Kumar N (2014) Combined rotor fault diagnosis in rotating machinery using empirical mode decomposition. J Mech Sci Technol 28:4869–4876

4. Katsifarakis N, Riga M, Voukantsis D, Karatzas K (2016) Computational intelligence methods for rolling bearing fault detection. J Braz Soc Mech Sci Eng. doi:10.1007/s40430-015-0458-6

5. Kankar PK, Sharma SC, Harsha SP (2011) Fault diagnosis of ball bearings using machine learning methods. Expert Syst Appl 38:1876–1886

6. Kavathekar S, Upadhyay N, Kankar PK (2016) Fault classification of ball bearing by rotation forest technique. Procedia Technol 23:187–192

7. Yang BS, Di X, Han T (2008) Random Forests classifier for machine fault diagnosis. J Mech Sci Technol 22:1716–1725

8. Vakharia V, Gupta VK, Kankar PK (2015) Ball bearing fault diagnosis using supervised and unsupervised machine learning methods. Int J Acoust Vib 20:244–250

9. Kankar PK, Sharma SC, Harsha SP (2011) Nonlinear vibration signature analysis of a rotor bearing system due to race imperfections. J Comput Nonlinear Dyn 7:011014-1–01101416

10. Kankar PK, Sharma SC, Harsha SP (2011) Rolling element bearing fault diagnosis using Wavelet Transform. Neurocomputing 74:1638–1645

11. Upadhyay R, Jharia S, Padhy PK, Kankar PK (2015) Application of wavelet-fractal features for the automated detection of epileptic seizure using electroencephalogram signals. Int J Biomed Eng Technol 19(4):355–372

12. Nikolaou NG, Antoniadis IA (2002) Rolling element bearing fault diagnosis using wavelet packets. NDT and E Int 35:197–205

13. Prabhakar S, Mohanty AR, Sekhar AS (2002) Application of discrete Wavelet Transform for detection of ball bearing race faults. Tribol Int 35:793–800

14. Yan R, Gao RX (2009) Base wavelet selection for bearing vibration signal analysis. Int J Wavelets Multiresolut Inf Process 7:411–426

15. Kankar PK, Sharma SC, Harsha SP (2011) Fault diagnosis of ball bearings using continuous Wavelet Transform. Appl Soft Comput 11:2300–2312

16. Vakharia V, Gupta VK, Kankar PK (2015) A multiscale entropy based approach to select wavelet for fault diagnosis of ball bearings. J Vib Control 21:3123–3131

17. Wu SD, Wu CW, Wu TY, Wang CC (2013) Multi-scale analysis based ball bearing defect diagnostics using Mahalanobis distance and support vector machine. Entropy 15:416–433

18. Zheng J, Cheng J, Yang Y (2014) Multiscale permutation entropy based rolling bearing fault diagnosis. Shock Vib 2014:1–8

19. Kappaganthu K, Nataraj C (2011) Feature selection for fault detection in rolling element bearings using mutual information. J Vib Acoust 133:1–11

20. Vapnik VN (1998) Statistical learning theory. Wiley, New York

21. Meyer D, Leisch F, Hornik K (2003) The support vector machine under test. Neurocomputing 55:169–186

22. Breiman L (2001) Random Forests. Mach Learn 45:5–32

23. Peng HW, Chiang PJ (2011) Ball bearing fault diagnosis using machine learning techniques. In: Proceedings of 2011 8th Asian control conference (ASCC) Kaohsiung, Taiwan, pp 175–180

24. Duda RO, Hart PE, Stork DG (2001) Pattern classification. Wiley, New York

25. Kononenko I (1994) Estimating attributes: analysis and extension of relief. In: Proceedings of the European conference on machine learning, Catania, Italy, April 6–8, pp 171–182

26. Naseriparsa M, Bidgoli AM, Varaee T (2013) A hybrid feature selection method to improve performance of a group of classification algorithms. Int J Comput Appl 69:28–35

27. Moura EP, Souto CR, Silva AA, Irmao MAS (2011) Evaluation of principal component analysis and neural network performance for bearing fault diagnosis from vibration signal processed by RS and DF analyses. Mech Syst Signal Process 25:1765–1772

28. Novakovic J (2009) Using information gain attribute evaluation to classify sonar targets 17th Telecommunications forum TELFOR 2009, Belgrade, pp 1351–1354

29. Sikonja MR, Kononenko I (2003) Theoretical and empirical analysis of ReliefF and RReliefF. Mach Learn J 53:23–69

30. Megchelenbrink W (2010) Relief-based feature selection in bioinformatics: detecting functional specificity residues from multiple sequence alignments. Master Thesis Dissertation, Radboud University Nijmegen. Thesis number: 132IK

31. Loparo KA (2016), Bearing vibration data set. Case Western Reserve University. Available at: www.eecs.cwru.edu/laboratory/bearing

32. Baydar N, Ball A (2001) A comparative study of acoustic and vibration signals in detection of gear failures using Wigner-Ville distribution. Mech Syst Signal Process 15:1091–1107

33. Kim BS, Lee SH, Lee MG, Ni J, Song JY, Lee CW (2007) A comparative study on damage detection in speed-up and coast-down process of grinding spindle-typed rotor-bearing system. J Mater Process Technol 187–188:30–36

34. Hall M, Frank E, Geoffrey H et al (2009) The WEKA data mining software: an update. SIGKDD Explor 11:10–18

35. Wang X, Zheng Y, Zhao Z, Wang J (2015) Bearing fault diagnosis based on statistical locally linear embedding. Sensors 15:16225–16247

36. Li Y, Xu M, Wei Y, Huang W (2016) A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree. Measurement 77:80–94

37. Zhang S, Li W (2014) Bearing condition recognition and degradation assessment under varying running conditions using NPE and SOM. Math Probl Eng. doi:10.1155/2014/781583