

## Research Article

# Multifault Diagnosis of Rolling Element Bearings Using a Wavelet Kurtogram and Vector Median-Based Feature Analysis

**Phuong H. Nguyen and Jong-Myon Kim**

*School of Electrical, Electronics and Computer Engineering, University of Ulsan, Ulsan 680-749, Republic of Korea*

Correspondence should be addressed to Jong-Myon Kim; [jongmyon.kim@gmail.com](mailto:jongmyon.kim@gmail.com)

Received 3 July 2015; Accepted 18 August 2015

Academic Editor: Wahyu Caesarendra

Copyright © 2015 P. H. Nguyen and J.-M. Kim. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a comprehensive multifault diagnosis methodology for incipient rolling element bearing failures. This is done by combining a wavelet packet transform- (WPT-) based kurtogram and a new vector median-based feature analysis technique. The proposed approach first extracts useful features that are characteristic of the bearing health condition from the time domain, frequency domain, and envelope power spectrum of incoming acoustic emission (AE) signals by using a WPT-based kurtogram. Then, an enhanced feature analysis approach based on the linear discriminant analysis (LDA) technique is used to select the most discriminant bearing fault features from the original feature set. These selected fault features are used by a Naïve Bayes (NB) classifier to classify the bearing fault conditions. The performance of the proposed methodology is tested and validated under various bearing fault conditions on an experimental test rig and compared with conventional state-of-the-art approaches. The proposed bearing fault diagnosis methodology yields average classification accuracies of 91.11%, 96.67%, 98.89%, 99.44%, and 98.61% at rotational speeds of 300, 350, 400, 450, and 500 rpm, respectively.

## 1. Introduction

For the past several decades, the development of reliable fault diagnosis systems to accurately detect and classify various bearing faults has been at the heart of research in the field of machine condition monitoring for preventive and predictive maintenance. These fault diagnosis systems aim to accurately detect bearing faults at the early stages of their development in order to prevent potential breakdown of industrial machines, thereby improving their reliability and reducing maintenance costs.

Several signal processing-based methods have been developed for fault diagnosis of rolling element bearings. These methods extract characteristic fault features from the inherently nonstationary signals obtained from faulty bearings. Antoni [1] presented the kurtogram, which has proven to be a powerful method for characterizing and extracting hidden nonstationary features in bearing fault signals [2–4]. The kurtogram method relies on spectral kurtosis, which was originally proposed by Dwyer [5], to detect the presence of nonstationary transients and accurately indicate their

location in the frequency bands of bearing signals. Recently, in attempts to enhance the performance of the kurtogram method for fault diagnosis, many researchers have integrated the kurtogram either with the short-time Fourier transform (STFT) [6] or with multirate filter banks (MRFB) [1]. These approaches, however, have yielded little improvement in extracting transient characteristics and have rendered kurtogram analysis more sensitive to irrelevant impulsive components [7]. These shortcomings have been addressed by introducing enhanced variations of the kurtogram, such as the one proposed by Shi et al. [8], which uses the complex Morlet wavelet transform. The complex Morlet wavelet transform, however, has its own limitations. These include its high computational complexity and its inability to effectively separate the high frequency components, which usually carry the characteristic fault information. In this paper, we propose a way to improve the kurtogram-based method by using the envelope power spectrum and the wavelet packet transform (WPT). The wavelet packet transform (WPT) is computationally more efficient and highly effective in frequency decomposition. The proposed kurtogram is performed by computing

kurtosis values from the envelope power spectrum of each subband signal at different levels of decomposition. The subband signal that yields the highest spectral kurtosis value is selected for extracting the features that are characteristic of the bearing fault conditions.

In addition to feature extraction techniques, feature analysis methods have also been widely studied in machine fault diagnosis in order to prevent the degradation of classification performance that is caused by redundant information in the feature space [9–13]. Dimensionality reduction techniques have been used to identify the features that effectively represent the high-dimensional data in a lower-dimensional space while retaining the intrinsic information of the bearing defects; these eventually lead to improvements in the classification performance of the fault diagnosis system. Among various dimensionality reduction methods, principle component analysis (PCA) [14], independent component analysis (ICA) [15], and linear discriminant analysis (LDA) [16] are the most popular methods that have been effectively utilized in machinery fault diagnosis. For instance, Widodo and Yang [11] applied the PCA technique to machine health prognostics to extract population characteristics from the available condition monitoring data. Guo et al. [12] investigated the performance of an ICA algorithm in separating the envelopes of vibration signals to extract the weak impulsive features of incipient faults in rolling element bearings. In their study looking into fault diagnosis in ball bearings, Harmouche et al. [13] demonstrated the use of an LDA technique to effectively discriminate the rather weak and otherwise difficult to detect/identify spectral features extracted from the envelope spectra of vibration signals. The LDA feature space yields better separation between different classes of bearing faults than conventional feature analysis methods, thereby resulting in improved classification performance. PCA and ICA are unsupervised methods that ignore the label information during dimensionality reduction, whereas LDA is a supervised method that utilizes the class label information to find an optimal low-dimensional representation of the original feature set while also preserving discriminant information between classes. As a result, LDA yields better classification results as compared to PCA and ICA [9, 13, 14]. This paper presents a new LDA-based feature analysis method that uses a vector median-based discriminant criterion (VMDC) to characterize the intraclass compactness and interclass separability of the feature space in order to select the most discriminative subset of bearing fault features.

This paper proposes a reliable multifault diagnosis scheme for rolling element bearings that combines an improved WPT-based kurtogram and the VMDC-based feature analysis methods. The contributions discussed in this paper can be summarized as follows:

- (1) An improved WPT-based kurtogram is proposed for extracting the most characteristic fault features from nonstationary acoustic emission (AE) signals. The kurtogram is performed by computing the spectral kurtosis values for the envelope power spectrum of each subband signal at different levels of decomposition. The subband signal that yields the highest

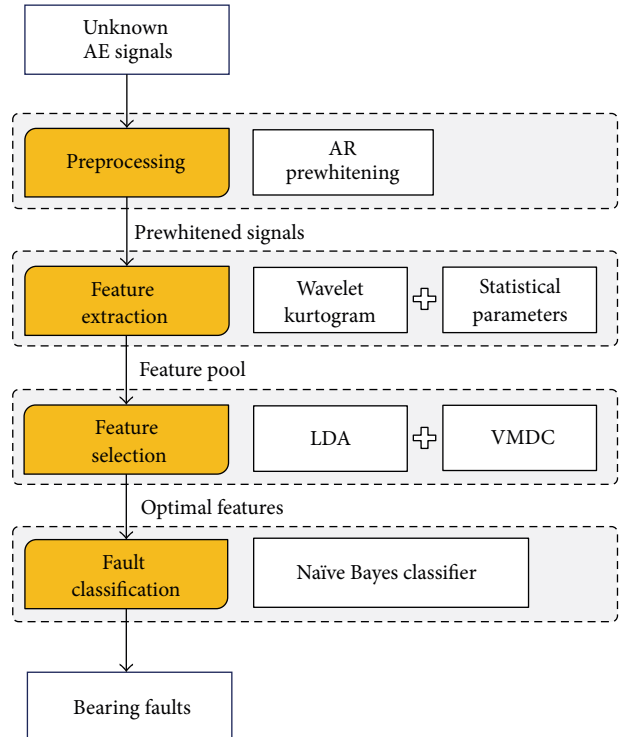


FIGURE 1: Flowchart of the proposed bearing fault diagnosis methodology.

spectral kurtosis value is selected for extracting the fault features.

- (2) An efficient VMDC-based discriminative fault feature analysis approach is proposed for selecting the most discriminative subset of fault features. A vector median-based discriminant criterion is presented to minimize the intraclass compactness and maximize the interclass separability of the feature space.
- (3) A comprehensive multiclass fault diagnosis methodology for rolling element bearings is presented, and its performance is validated using AE data for various single and compound bearing defects acquired under different simulated crack sizes and bearing rotational speeds.

The remainder of the paper is structured as follows. Section 2 describes the overall bearing fault diagnosis methodology, Section 3 introduces the self-designed fault simulator (including the data acquisition system and seeded bearing defects for collecting fault signals), and Section 4 shows the experimental results. Finally, Section 5 concludes this paper.

## 2. Proposed Bearing Fault Diagnosis Methodology

The proposed comprehensive bearing fault diagnosis scheme consists of four main processes including signal preprocessing, feature extraction, feature selection, and fault classification. Figure 1 illustrates the overall flowchart of the proposed methodology.

**2.1. Preprocessing.** The performance of wavelet kurtogram analysis can be improved if the variations between the peaks and flanks of the spectral energy are minimized as shown in Table 6; this can be achieved by preprocessing the AE signal before feature extraction [17]. In this study, the prewhitening technique is used to reduce the spectral energy variations in the incoming AE fault signal. To obtain the prewhitened signal, an autoregressive (AR) model  $y(n)$  is utilized, which is defined as follows:

$$y(n) = \sum_{i=1}^p a_i x(i+n) + e(n), \quad (1)$$

where  $x(n)$  is the incoming bearing signal,  $a_i$  are the AR coefficients,  $p$  denotes the order of the AR model, and  $e(n)$  is the residual signal representing a spectrum close to the white noise spectrum.

**2.2. Feature Extraction.** For the purpose of extracting the characteristic fault signatures from the AE signals, this paper presents a two-step feature extraction process that works as follows.

*Step 1.* In this step, the proposed WPT-based kurtogram method is performed on the prewhitened AE signal to determine its most informative subband. The flowchart of the proposed WPT-based kurtogram is illustrated in Figure 2, and its details are provided below.

First, the prewhitened bearing signal is decomposed into a series of subbands by performing the wavelet packet transform (WPT) with five decomposition levels ( $n = 5$ ) using a Daubechies 2 (or db2) filter. The five-level WPT decomposition results in a total of 63 subband signals. The envelope spectrum of each subband signal is then calculated using the fast Fourier transform (FFT) to identify the characteristic bearing fault frequencies. The envelope spectrum of an AE signal obtained from a defective bearing is usually flooded with the characteristic defect frequency and its harmonics. We utilize kurtosis to measure the degree of protrusion of the envelope spectrum and then characterize the hidden bearing defect signatures. In order to do this effectively, we first determine the bearing defect frequencies, which include the ball pass frequency of the bearing outer raceway (BPFO), the ball pass frequency of the bearing inner raceway (BPFI), the ball spin frequency of the bearing roller (BSF), and the first  $M$  harmonics ( $M = 3$  in this paper) of each of these frequencies. These defect frequencies are defined as follows:

$$\begin{aligned} \text{BPFO} &= \frac{N_r \cdot F_o}{2} \left( 1 - \frac{B_d}{P_d} \cos \alpha \right), \\ \text{BPFI} &= \frac{N_r \cdot F_o}{2} \left( 1 + \frac{B_d}{P_d} \cos \alpha \right), \\ \text{BSF} &= \frac{P_d \cdot F_o}{2 \cdot B_d} \left( 1 - \left( \frac{B_d}{P_d} \cos \alpha \right)^2 \right), \end{aligned} \quad (2)$$

where  $N_r$  denotes the number of cylindrical rollers,  $F_o$  is the shaft speed,  $\alpha$  is the contact angle, and  $B_d$  and  $P_d$  are

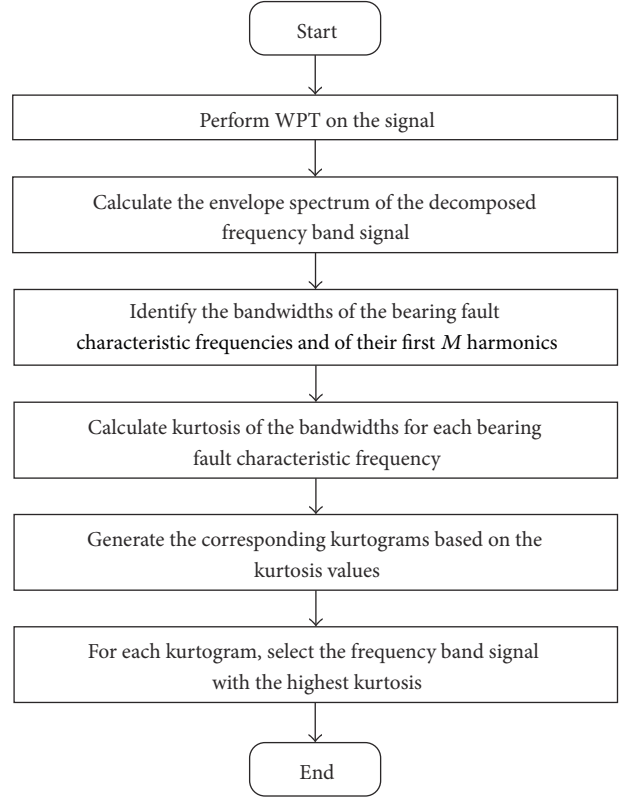


FIGURE 2: Flowchart of the wavelet kurtogram algorithm.

the roller and pitch diameters, respectively. Each defect-related bandwidth  $BW_i$  ranges from  $a_i$  to  $b_i$  ( $i = 1, 2, \dots, M$ ), which can be calculated as follows:

$$\begin{aligned} a_i &= \begin{cases} (i - \text{ER}) * \text{BPFO} & \text{if BPFO} \\ (i - \text{ER}) * \text{BPFI} - 2 * (1 + \text{ER}) * F_o & \text{if BPFI} \\ (i - \text{ER}) * \text{BSF} - 2 * (1 + \text{ER}) * F_C & \text{otherwise,} \end{cases} \\ b_i &= \begin{cases} (i + \text{ER}) * \text{BPFO} & \text{if BPFO} \\ (i + \text{ER}) * \text{BPFI} + 2 * (1 + \text{ER}) * F_o & \text{if BPFI} \\ (i + \text{ER}) * \text{BSF} + 2 * (1 + \text{ER}) * F_C & \text{otherwise.} \end{cases} \end{aligned} \quad (3)$$

Here, ER is the error rate of the bearing test rig and  $F_C$  is the cage frequency. Once the bearing fault bandwidths are identified, three different kurtosis values are calculated corresponding to the three bearing defect frequencies (i.e., BPFO, BPFI, and BSF). After obtaining the kurtosis values for the subband signals, the wavelet kurtogram is generated (Figure 3). Figure 3(a) shows the frequency-scale paving of the proposed wavelet kurtogram, while Figure 3(b) illustrates an example color kurtogram of the three bearing fault frequencies (i.e., BPFO, BPFI, and BSF), which represents the kurtosis values of all of the subbands.

Finally, for each kurtogram, the subband signal with the highest kurtosis value (i.e., the signal that has the most discriminative fault features) is selected for extracting the bearing fault signatures. In Figure 3(b), the selected subband

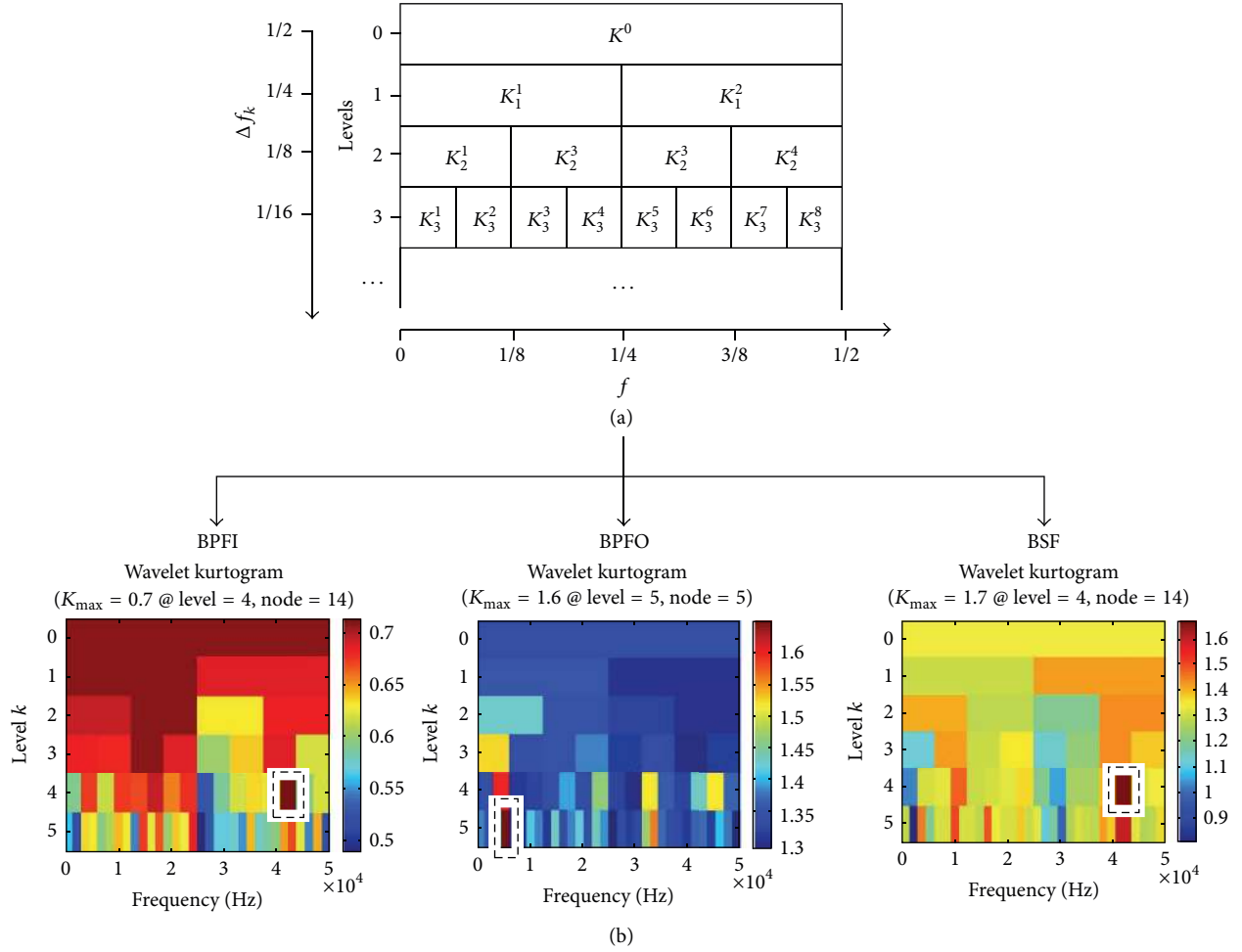


FIGURE 3: The proposed wavelet kurtogram. (a) The frequency-scale paving of the proposed wavelet kurtogram. (b) An example of a proposed wavelet kurtogram.

with the highest kurtosis value is highlighted using a dashed rectangle.

*Step 2.* In this step, 10 time domain ( $f_1, f_2, \dots, f_{10}$ ) and three frequency domain statistical features ( $f_{11}, f_{12}, f_{13}$ ) are extracted from each of the selected subband signals, as shown in Tables 1 and 2, respectively. Furthermore, we calculate three root mean square (RMS) features ( $f_{14}-f_{16}$ ) for the first three harmonics of the corresponding defect frequencies in the envelope spectrum of each of the selected subband signals, as described in Table 3. This results in a feature pool that consists of a total of  $N_c \times N_s \times N_f$  features, where  $N_c$  is the number of bearing fault classes,  $N_s$  is the number of AE signals obtained for each fault type, and  $N_f$  is the total number of features. In this case,  $N_f$  is  $10 + 3 + 3 = 16$  and includes 10 time domain statistical features, three frequency domain statistical features, and three envelope spectrum RMS features (calculated using the extracted wavelet kurtogram) for each of the three bearing defect frequencies (i.e., BPFO, BPF1, and BSF).

**2.3. Feature Selection.** In this paper, we propose a feature analysis method based on the linear discriminant analysis

(LDA) technique plus a vector median-based discriminant criterion (VMDC) to find the most discriminative subset of the extracted fault features for accurate fault diagnosis. LDA is a supervised dimensionality reduction algorithm that utilizes the class label information to find an optimal linear transformation matrix that projects the original high-dimensional feature space onto the lower-dimensional representation while also preserving the information that can help discriminate among different classes. Let  $X = \{x_1, x_2, \dots, x_N\} \in R^{D \times N}$  be a dataset, where  $D$  is the original dimensionality and  $N$  is the number of samples in  $X$ . Each sample  $x_i$  belongs to a class  $c_i = \{1, 2, \dots, C\}$ . Let  $N_i$  be the number of samples in class  $c_i$  and let  $N$  be the total number of samples in all of the classes. Then, the intraclass scatter matrix  $S_w$  and the interclass scatter matrix  $S_b$  can be evaluated with (4) and (5), respectively:

$$S_w = \sum_{i=1}^C \sum_{x_k \in C_i} (x_k - \mu_i)^T (x_k - \mu_i), \quad (4)$$

$$S_b = \sum_{i=1}^C (\mu_i - \mu)^T (\mu_i - \mu). \quad (5)$$

TABLE 1: Time domain statistical feature parameters.

Parameter	Definition
Root mean square ( $f_1$ )	$\sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$
Root mean square of the amplitude ( $f_2$ )	$\left( \frac{1}{N} \sum_{i=1}^N \sqrt{ x_i } \right)^2$
Kurtosis ( $f_3$ )	$\frac{1}{N} \sum_{i=1}^N \left( \frac{x_i - \bar{x}}{\sigma} \right)^4$
Skewness ( $f_4$ )	$\frac{1}{N} \sum_{i=1}^N \left( \frac{x_i - \bar{x}}{\sigma} \right)^3$
Peak-to-peak ( $f_5$ )	$\max(x_i) - \min(x_i)$
Crest factor ( $f_6$ )	$\frac{\max( x_i )}{\sqrt{(1/N) \sum_{i=1}^N x_i^2}}$
Impulse factor ( $f_7$ )	$\frac{\max( x_i )}{(1/N) \sum_{i=1}^N  x_i }$
Margin factor ( $f_8$ )	$\frac{\max( x_i )}{\left( (1/N) \sum_{i=1}^N \sqrt{ x_i } \right)^2}$
Shape factor ( $f_9$ )	$\frac{\sqrt{(1/N) \sum_{i=1}^N x_i^2}}{(1/N) \sum_{i=1}^N  x_i }$
Kurtosis factor ( $f_{10}$ )	$\frac{(1/N) \sum_{i=1}^N ((x_i - \bar{x})/\sigma)^4}{\left( (1/N) \sum_{i=1}^N x_i^2 \right)^2}$

TABLE 2: Frequency domain statistical feature parameters.

Parameter	Definition
Frequency center ( $f_{11}$ )	$\frac{1}{N} \sum_{i=1}^N f_i$
Root mean square frequency ( $f_{12}$ )	$\sqrt{\frac{1}{N} \sum_{i=1}^N f_i^2}$
Root variance frequency ( $f_{13}$ )	$\sqrt{\frac{1}{N} \sum_{i=1}^N (f_i - X_{fc})^2}$

TABLE 3: Envelope spectrum statistical feature parameters.

Parameter	Definition
RMS frequency of the first harmonic ( $f_{14}$ )	$\sqrt{\frac{1}{(b_1 - a_1)} \sum_{i=a_1}^{b_1} f_i^2}$
RMS frequency of the second harmonic ( $f_{15}$ )	$\sqrt{\frac{1}{(b_2 - a_2)} \sum_{i=a_2}^{b_2} f_i^2}$
RMS frequency of the third harmonic ( $f_{16}$ )	$\sqrt{\frac{1}{(b_3 - a_3)} \sum_{i=a_3}^{b_3} f_i^2}$

Here,  $\mu_i = (1/N_i) \sum_{x_i \in c_i} x_i$  is the mean of the samples labelled as class  $c_i$  and  $\mu = (1/N) \sum_{i=1}^N x_i$  is the mean of all of the samples. The LDA projects the space of the original variables

onto a  $(C - 1)$ -dimensional space by maximizing the Fisher discriminant rule [18], which is defined as follows:

$$W^* = \arg \max \frac{W^T S_b W}{W^T S_w W}. \quad (6)$$

This discriminant rule maximizes the ratio of inter-class separability to intraclass compactness by selecting a reduced number of the most discriminative components only. However, LDA-based feature analysis approaches are limited in their ability to efficiently select the optimal number of discriminative components to gain the highest classification performance. In the existing literature, there is no clear consensus on the optimal number of LDA components that yield the highest classification performance. To address this problem, we present a robust method based on the vector median approach to obtain the optimal number of LDA components that can maximize the classification performance. In order to identify the vector median of each class, this study utilizes a cumulative distance criterion using the Euclidean distance. The vector median  $VM_i$  of class  $c_i$  is defined as the data point that yields the minimum accumulated distance and is determined as follows:

$$VM_i = \arg \min_{k \in c_i} \sum_{j \in c_i, j \neq k} \|x_k - x_j\|_2. \quad (7)$$

Once the vector medians of all classes are determined, the modified intraclass compactness and interclass separability values are calculated using (8) and (9), respectively:

$$S_w = \frac{1}{C} \sum_{i=1}^C \frac{1}{N_i} \sum_{j=1}^{N_i} \|VM_i - x_j\|_2, \quad (8)$$

$$S_b = \frac{1}{C} \sum_{i=1}^C \arg \min_{j \in C, j \neq i} \|VM_i - VM_j\|_2. \quad (9)$$

In (8), the intraclass compactness of a class is reflected by the distance from its vector median to other samples in that class. Alternatively, in (9), the interclass separability of a class from other classes is represented by the minimum distance of the vector median of that class to the vector medians of other classes. The effect of outliers in the feature space is minimized by averaging these distance measurements, which improves the discriminatory power of the proposed feature analysis method. The vector median-based discriminant criterion that is used to select the optimal number of useful LDA components is defined as follows:

$$K^* = \arg \max \frac{S_b}{S_w}. \quad (10)$$

**2.4. Fault Classification.** In this study, a Naïve Bayes classifier [19] is used to classify various single and compound bearing faults. This classifier is selected due to its simplicity and high classification performance at a relatively low computational cost [19, 20]. The Naïve Bayes classifier has complexity  $O(Nd)$ , where  $N$  is the total number of training samples and  $d$  is the number of conditional features. Other classification



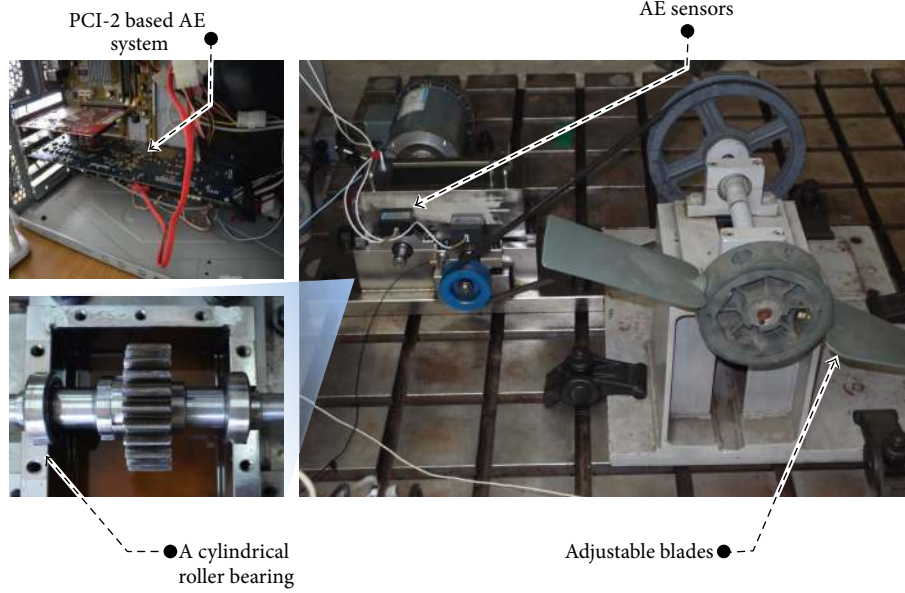


FIGURE 4: Experimental test rig for fault diagnosis of rolling element bearings [10].

algorithms, such as KNN, ANN, and SVM, have complexities  $O(N^2)$ ,  $O(N^2)$ , and  $O(N^3)$ , respectively [20]. Furthermore, the Naïve Bayes classifier is adaptive in nature and does not require any fixed parameters (unlike other classification algorithms).

In this study, in order to evaluate the generalized classification performance of the proposed comprehensive bearing fault diagnosis method,  $k$ -fold cross validation ( $k$ -cv) [21] is employed. In  $k$ -cv, the feature pool is randomly partitioned into  $k$  mutual subsets ( $k = 3$  in this study). Then, at the  $i$ th iteration of  $k$ -cv, one of the subsets is used as the testing dataset and the other ( $k - 1$ ) subsets are employed to train the Naïve Bayes classifier. The classification performance is averaged over  $k$  trials of  $k$ -cv. The advantage of this technique is that no matter how we divide the data every data sample is in the testing set once and is included in the training set ( $k - 1$ ) times. Thus, variance in the classification results is minimized. To validate the classification performance of the proposed method, this paper utilizes the average classification accuracy (ACA) and the true positive rate (TPR) as performance measures. These are defined in (11) and (12), respectively:

$$ACA = \frac{1}{k} \sum_{i=1}^k \frac{1}{N} \sum_{j=1}^C N_{TP}^{i,j} \times 100 (\%), \quad (11)$$

$$TPR_j = \frac{1}{k} \sum_{i=1}^k \frac{N_{TP}^{i,j}}{N_{TP}^{i,j} + N_{FN}^{i,j}} \times 100 (\%), \quad (12)$$

where  $k$  is the number of cross validation folds and  $N_{TP}^{i,j}$  and  $N_{FN}^{i,j}$  are the number of true positives and false negatives of class  $c_j$  resulting in the  $i$ th iteration of  $k$ -cv, respectively. In addition, the number of true positives is defined as the total number of samples in class  $c_j$  that are accurately classified as

class  $c_j$ . Alternatively, the number of false negatives indicates the number of samples in class  $c_j$  that are not classified as  $c_j$ .

### 3. Experimental Setup

In this study, we used an experimental test rig developed by Intelligence Dynamics Lab (Gyeongsang National University, Korea) to validate the performance of the proposed fault diagnosis methodology, as shown in Figure 4. An acoustic emission (AE) sensor (WS $\alpha$  type, from Physical Acoustic Cooperation) was employed to capture continuous AE signals. Table 3 describes the specifications of the data acquisition system used in this study.

The bearings used in this study were cylindrical rolling element type (FAG NJ206-3-TVP2). These bearings were engrained with different bearing defects, which are shown in Figure 5 and listed as follows: bearing-crack-on-outer-raceway (BCO), bearing-crack-on-inner-raceway (BCI), bearing-crack-on-roller (BCR), bearing-crack-on-outer-and-inner-raceways (BCOI), bearing-crack-on-outer-raceway-and-roller (BCOR), bearing-crack-on-inner-race-and-roller (BCIR), and bearing-crack-on-outer-and-inner-raceways-and-roller (BCOIR). A healthy or defect-free bearing (DFB) was used as a reference for baseline measurements. In total, eight types of bearing fault signals, including a normal defect-free bearing (as a baseline), were acquired under different rotational speeds (300, 350, 400, 450, and 500 rpm). A total of 360 AE fault signals, each 10 seconds in duration, were collected at a sampling rate of 250 KHz for each bearing defect in this study.

### 4. Experimental Results

**4.1. Data Preprocessing.** In this subsection, we investigate and justify the preprocessing of the AE fault signals to

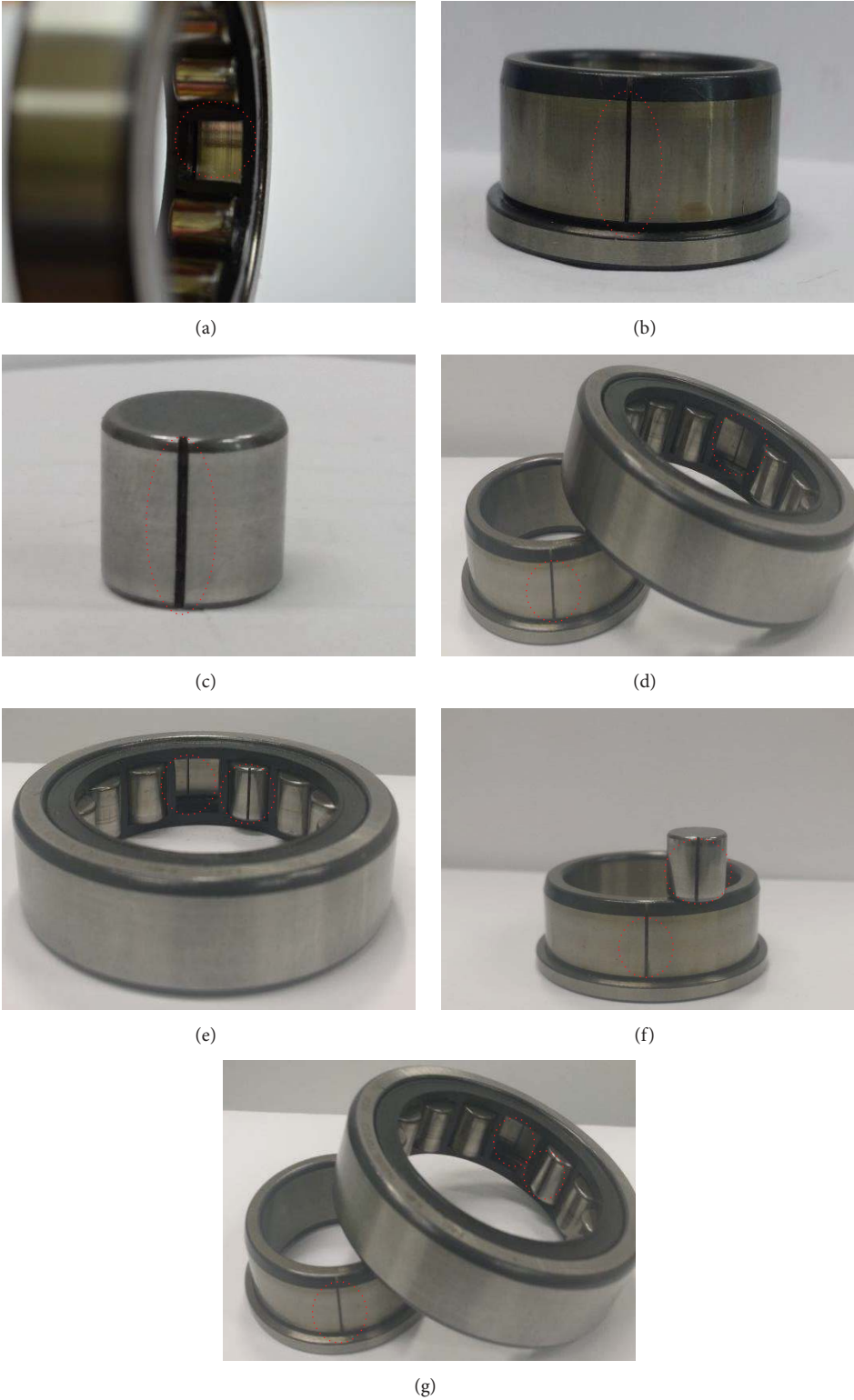


FIGURE 5: Single and multiple-combined bearing defects. (a) BCO, (b) BCI, (c) BCR, (d) BCOI, (e) BCOR, (f) BCIR, and (g) BCOIR.

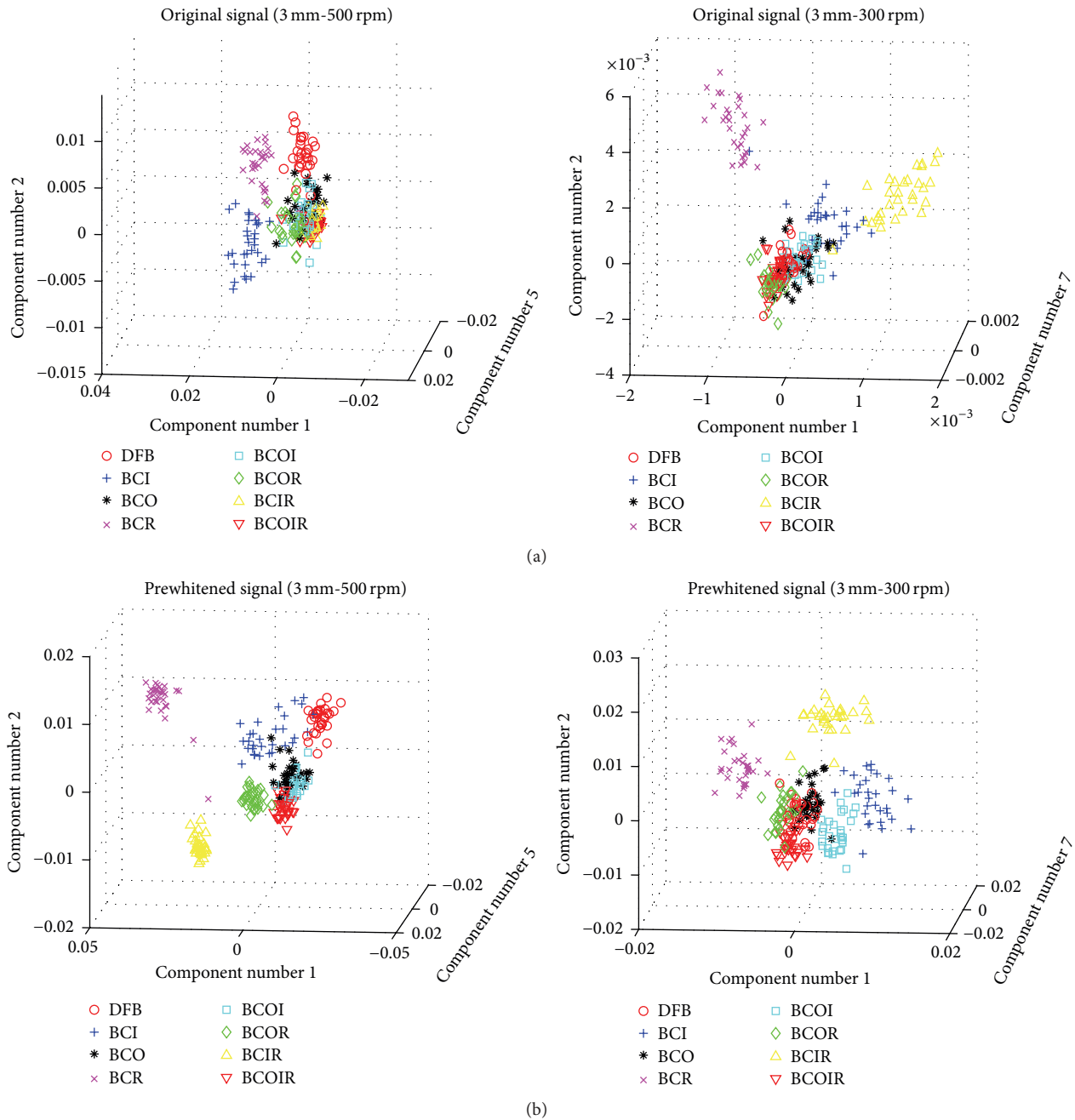


FIGURE 6: Feature distribution spaces with a bearing crack size of 3 mm. (a) Before prewhitening and (b) after prewhitening.

improve the classification performance of our methodology. As mentioned earlier, a total of eight single and compound bearing defects were investigated in this study. Figures 6 and 7 show the feature distribution of the optimal LDA discriminant components for four datasets. These datasets were acquired at two different rotational speeds (i.e., 500 rpm and 300 rpm) and for two different bearing crack sizes (i.e., 3 mm and 12 mm). From Figure 6, it is clear that the fault features are not well separated and discrimination between different classes is poor without preprocessing. For instance, the class clusters of the 3 mm crack length and 300 rpm rotational speed datasets are heavily overlapped,

which leads to higher error rate in classification. Conversely, preprocessing the original signals to obtain the prewhitened signals reduces overlapping between clusters and improves separation between their boundaries.

Likewise, Figure 7 shows that the boundaries between some class clusters of the original signals are blurred and unclear, whereas the prewhitened signals have clearer and more distinct boundaries. These results strongly suggest that prewhitened signals are able to strongly project the discriminative information that is hidden in the bearing signals. The variation between the peaks and flanks of the spectral energy in spectral kurtosis analysis is one reason why each class of



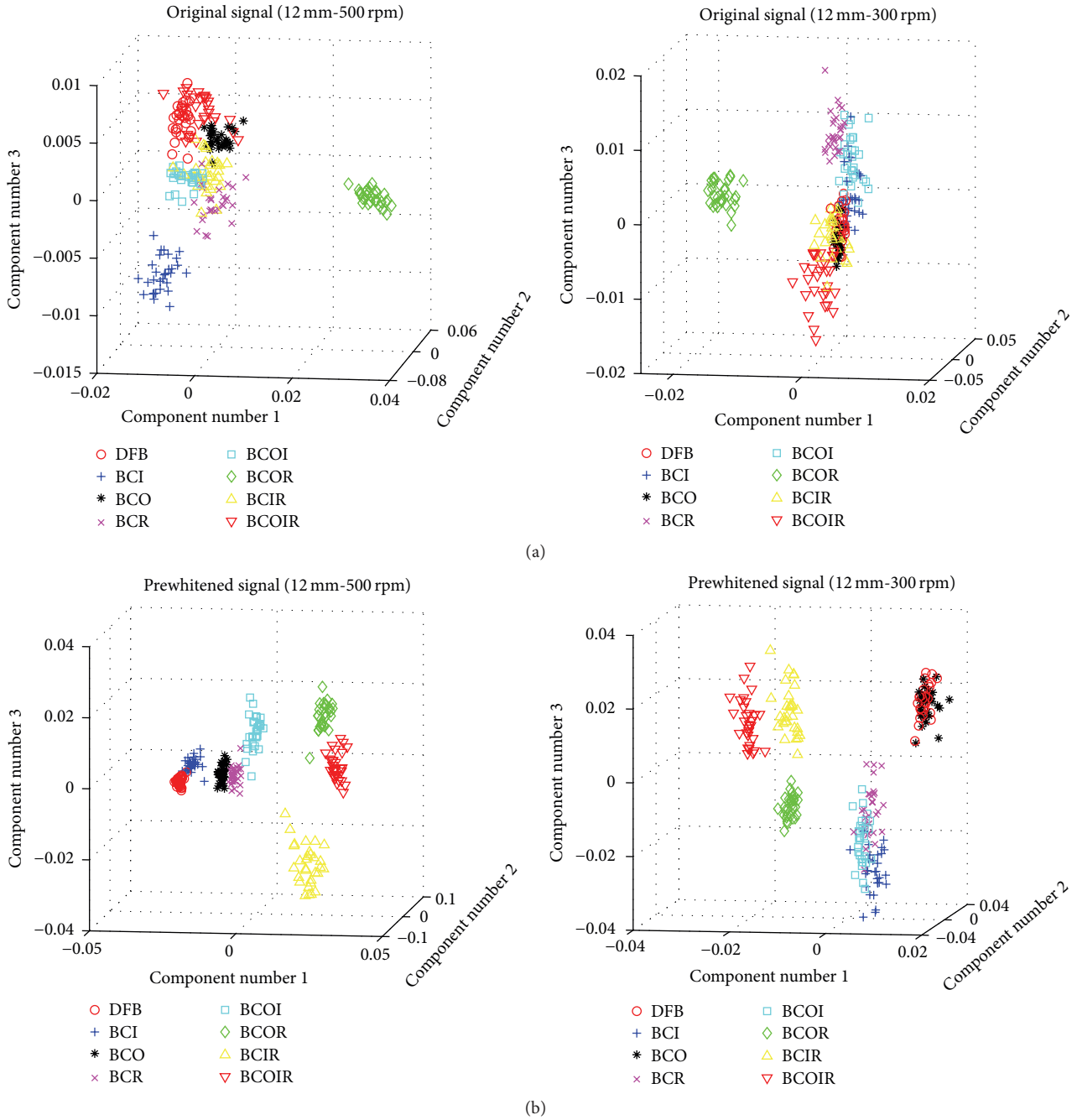


FIGURE 7: Feature distribution spaces with a bearing crack size of 12 mm. (a) Before prewhitening and (b) after prewhitening.

the original signals cannot be clustered accurately. Therefore, a prewhitening process is required to improve separation between different clusters, which makes classification easier and more accurate, even with a Naïve Bayes classifier.

**4.2. Proposed VMDC-Based Feature Analysis Method.** In order to validate the effectiveness of the proposed feature analysis method, this study compares the feature selection performance (in terms of the classification accuracy) of the proposed method and the original LDA for 10 different datasets, as shown in Figures 8 and 9. The vertical axis

represents the average classification accuracy calculated for the classification results of different fault types through  $k$  trials of  $k$ -cv, and the horizontal axis represents the number of LDA discriminant components that are selected. There is no general consensus about the appropriate number of LDA components that always guarantees the highest classification accuracy. As evident in Figures 8 and 9, the LDA method achieves its highest classification accuracy when the number of LDA discriminant components reaches seven. Nevertheless, there are some exceptions. For example, in Figure 8 (for the dataset of a 3 mm length crack and a rotational speed of

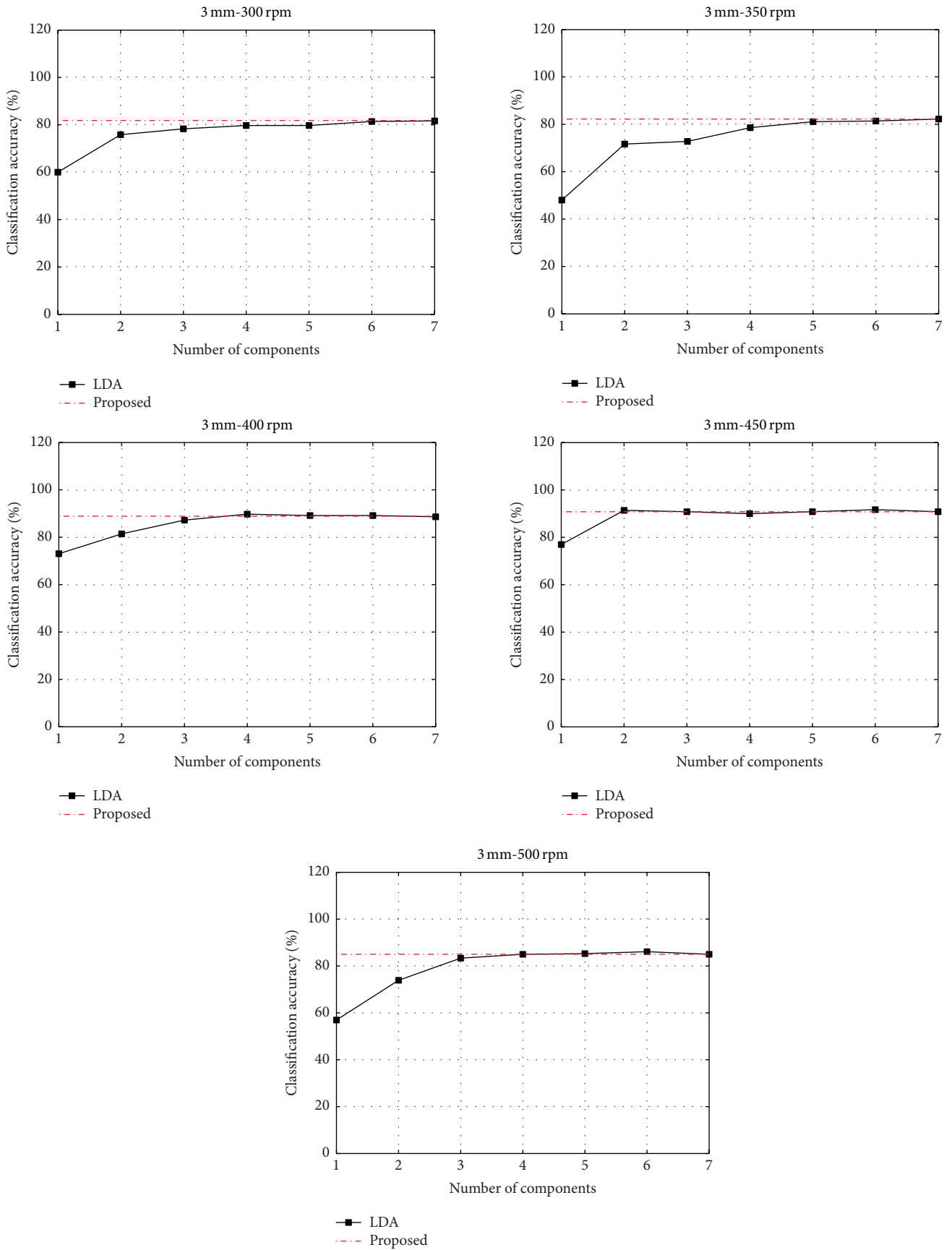


FIGURE 8: Performance of the proposed feature analysis at a bearing crack size of 3 mm.

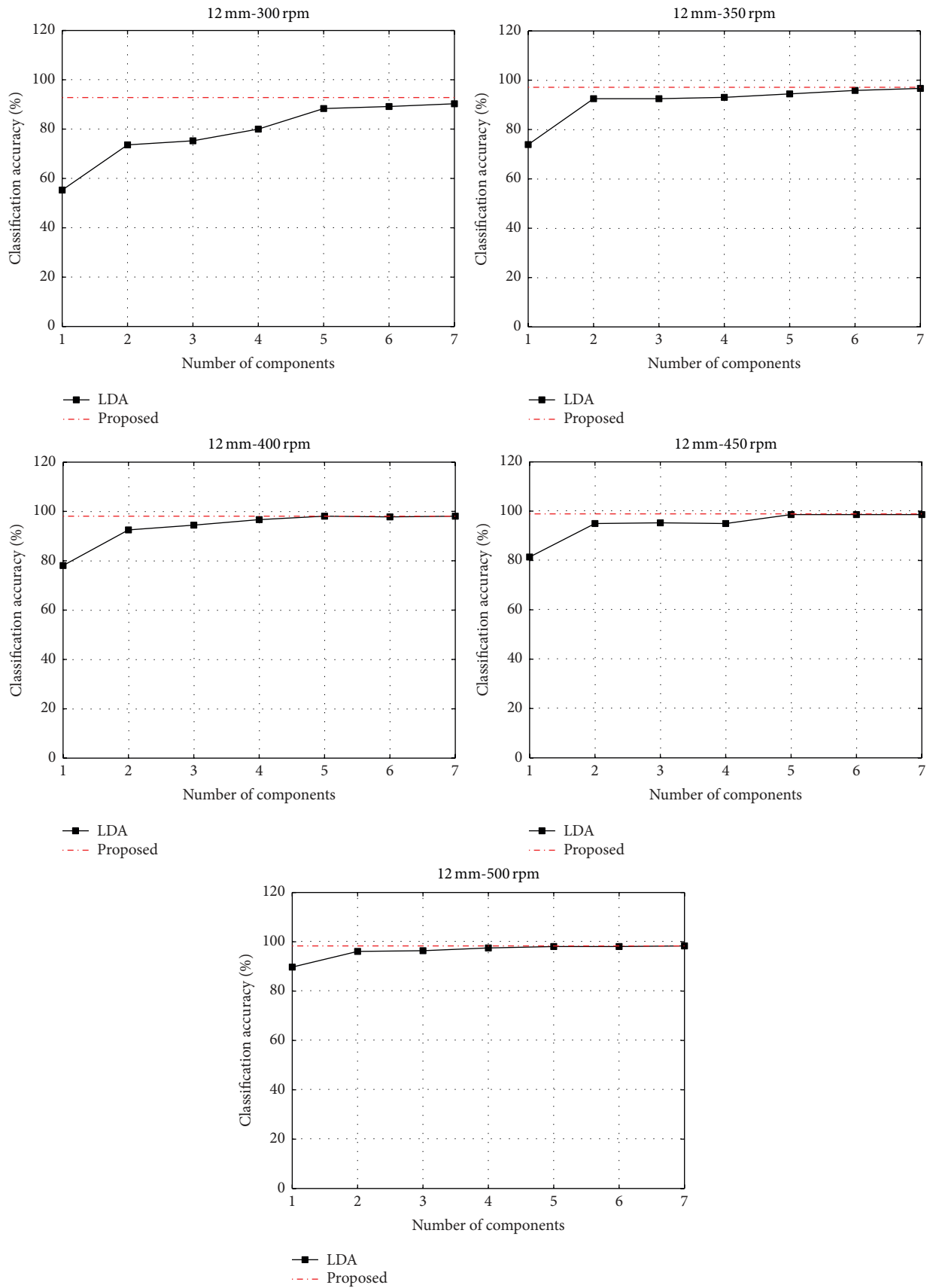


FIGURE 9: Performance of the proposed feature analysis at a bearing crack size of 12 mm.

TABLE 4: Specifications of the data acquisition system.

PCI-2 based AE system	(i) ADC: 18-bit 40 MS/s per channel maximum (ii) Frequency response: 1 kHz–3 MHz (at –3 dB points) (iii) Sample rate: 100 kS/s, 200 kS/s, 500 kS/s, 1 MS/s, 2 MS/s, 5 MS/s, 10 MS/s, 20 MS/s, and 40 MS/s are selectable
WS $\alpha$ sensor	(i) Peak sensitivity (V/ $\mu$ bar): –62 dB (ii) Operating frequency range: 100–900 kHz (iii) Resonant frequency: 650 kHz

kS/s: kSamples per second; MS/s: MSamples per second.

TABLE 5: Performance comparison between conventional feature analysis approaches and the proposed feature analysis method in terms of the average classification accuracy and the true positive rate at a bearing crack size of 3 mm (unit: %).

	Shaft speed	TPR per bearing fault condition under 3 mm crack size								ACA
		DFB	BCI	BCO	BCR	BCOI	BCOR	BCIR	BCOIR	
PCA	300 RPM	40.00	62.22	62.22	66.67	53.33	62.22	66.67	75.56	<b>61.11</b>
	350 RPM	57.78	55.56	28.89	91.11	44.44	68.89	31.11	55.56	<b>54.17</b>
	400 RPM	66.67	71.11	15.56	86.67	75.56	51.11	97.78	60.00	<b>65.56</b>
	450 RPM	86.67	86.67	26.67	91.11	33.33	75.56	95.56	80.00	<b>71.94</b>
	500 RPM	82.22	82.22	33.33	91.11	48.89	71.11	93.33	46.67	<b>68.61</b>
ICA	300 RPM	48.89	60.00	62.22	71.11	40.00	57.78	71.11	77.78	<b>61.11</b>
	350 RPM	42.22	51.11	13.33	86.67	35.56	73.33	57.78	40.00	<b>50.00</b>
	400 RPM	88.89	80.00	42.22	80.00	66.67	35.56	100.00	62.22	<b>69.44</b>
	450 RPM	88.89	82.22	28.89	88.89	53.33	82.22	95.56	73.33	<b>74.17</b>
	500 RPM	88.89	84.44	35.56	91.11	31.11	75.56	93.33	44.44	<b>68.06</b>
LDA	300 RPM	53.33	93.33	73.33	100.00	77.78	84.44	82.22	77.78	<b>80.28</b>
	350 RPM	82.22	91.11	71.11	86.67	77.78	86.67	80.00	88.89	<b>83.06</b>
	400 RPM	93.33	91.11	57.78	93.33	91.11	77.78	97.78	80.00	<b>85.28</b>
	450 RPM	88.89	91.11	77.78	100.00	86.67	95.56	100.00	71.11	<b>88.89</b>
	500 RPM	88.89	97.78	68.89	91.11	73.33	86.67	100.00	80.00	<b>85.83</b>
Proposed	300 RPM	53.33	80.00	62.22	100.00	75.56	80.00	82.22	80.00	<b>76.67</b>
	350 RPM	80.00	88.89	73.33	88.89	80.00	86.67	82.22	91.11	<b>83.89</b>
	400 RPM	91.11	91.11	60.00	93.33	93.33	82.22	100.00	82.22	<b>86.67</b>
	450 RPM	88.89	97.78	86.67	97.78	88.89	97.78	95.56	84.44	<b>92.22</b>
	500 RPM	88.89	97.78	64.44	91.11	75.56	91.11	100.00	80.00	<b>86.11</b>

450 rpm), the highest accuracy that can be achieved is 91.67% (with six LDA components); the highest accuracy obtained with seven LDA components is only 90.83%. Thus, the number of LDA components that can achieve the maximum classification accuracy must be determined empirically for each study. On the contrary, the proposed feature analysis approach always shows better average classification accuracies when compared with the original LDA algorithm. For instance, in Figure 9 (for the dataset of a 12 mm length crack and a rotational speed of 300 rpm), the proposed method yields a classification accuracy of up to 92.8%, clearly outperforming the LDA method with any number of components. This can be explained by the fact that our proposed method utilizes the vector-median discriminant criterion, which can adaptively select the optimal number of LDA components to maximize the ratio of interclass separability and intraclass compactness, thereby achieving the maximum discriminatory power in a low-dimensional feature space. This enables the proposed approach to always achieve the best classification performance. The analysis of

these results highlights the main contributions of this study in analyzing and selecting the optimal features from the original feature pool to achieve superior classification performance as compared to conventional feature analysis algorithms. Using the LDA method, the number of LDA components that yield the best classification results must be determined empirically on a dataset to dataset basis, whereas the proposed VMDC-based approach does this adaptively.

**4.3. Classification Performance.** In this section, we compare the classification performance of the proposed methodology with other state-of-the-art feature analysis approaches (i.e., PCA, ICA, and LDA). The optimal number of components for PCA, ICA, and LDA was determined by utilizing the training set of  $k$ -cv fold at each iteration and finding the number of components that yielded the highest accuracy for each of these methods. The final classification results, which have been calculated through  $k$  folds of  $k$ -cv, are summarized in Tables 4 and 5. We used two evaluation indexes: the true positive rate (TPR) and the average classification accuracy

TABLE 6: Performance comparison between conventional feature analysis approaches and the proposed feature analysis method in terms of the average classification accuracy and the true positive rate at a bearing crack size of 12 mm (unit: %).

	Shaft speed	TPR per bearing fault condition under 12 mm crack size								ACA
		DFB	BCI	BCO	BCR	BCOI	BCOR	BCIR	BCOIR	
PCA	300 RPM	84.44	75.56	64.44	66.67	86.67	95.56	71.11	75.56	<b>77.50</b>
	350 RPM	95.56	53.33	75.56	71.11	77.78	88.89	80.00	95.56	<b>79.72</b>
	400 RPM	93.33	73.33	86.67	82.22	75.56	91.11	95.56	97.78	<b>86.94</b>
	450 RPM	91.11	77.78	88.89	73.33	91.11	95.56	91.11	91.11	<b>87.50</b>
	500 RPM	88.89	95.56	91.11	86.67	95.56	93.33	91.11	93.33	<b>91.94</b>
ICA	300 RPM	51.11	48.89	86.67	66.67	71.11	93.33	71.11	64.44	<b>69.17</b>
	350 RPM	64.44	42.22	88.89	71.11	86.67	91.11	77.78	91.11	<b>76.67</b>
	400 RPM	88.89	62.22	88.89	75.56	75.56	88.89	91.11	97.78	<b>83.61</b>
	450 RPM	100.00	48.89	93.33	68.89	88.89	97.78	93.33	88.89	<b>85.00</b>
	500 RPM	91.11	88.89	91.11	84.44	97.78	93.33	88.89	95.56	<b>91.39</b>
LDA	300 RPM	77.78	88.89	82.22	93.33	95.56	100.00	97.78	88.89	<b>90.56</b>
	350 RPM	91.11	100.00	91.11	93.33	100.00	100.00	97.78	100.00	<b>96.67</b>
	400 RPM	95.56	100.00	97.78	95.56	100.00	100.00	97.78	97.78	<b>98.06</b>
	450 RPM	100.00	100.00	97.78	93.33	97.78	100.00	100.00	97.78	<b>98.33</b>
	500 RPM	91.11	100.00	100.00	97.78	100.00	97.78	100.00	100.00	<b>98.33</b>
Proposed	300 RPM	80.00	86.67	88.89	93.33	93.33	100.00	97.78	88.89	<b>91.11</b>
	350 RPM	91.11	100.00	88.89	93.33	100.00	100.00	100.00	100.00	<b>96.67</b>
	400 RPM	95.56	100.00	97.78	97.78	100.00	100.00	100.00	100.00	<b>98.89</b>
	450 RPM	100.00	100.00	100.00	97.78	97.78	100.00	100.00	100.00	<b>99.44</b>
	500 RPM	93.33	100.00	100.00	97.78	100.00	97.78	100.00	100.00	<b>98.61</b>

(ACA). From the figures in Tables 4 and 5, it is clear that the proposed feature analysis scheme consistently outperforms the conventional approaches of PCA, ICA, and LDA for all of the datasets that were analyzed in this study. For instance, Table 5 reveals that the TPR of the proposed method is around 100% for almost all bearing faults (except for the BCR fault condition, where it ranges between 93 and 97% but is still higher than that of PCA and ICA). In Table 4, it can be observed that the proposed method achieves good classification performance even though the feature clustering of the 3 mm length crack datasets was poor and not well separated as compared to the 12 mm length crack datasets. This comparatively high classification performance demonstrates that the proposed approach is more effective in feature analysis as compared to other conventional methods. Our approach is also able to yield maximum discrimination in a lower-dimensional feature space.

## 5. Conclusion

This paper proposed a comprehensive multifault diagnosis methodology based on AE analysis to detect multiple localized bearing faults of a rolling element bearing. The method comprises a prewhitening step, a feature extraction based on wavelet kurtogram, a useful feature selection with the proposed VMDC-based feature analysis, and a fault classification by employing the NB classifier. One of the main advantages of the proposed methodology is that the subband of the nonstationary AE signal, which carries the most sensitive information related to fault characteristic,

is adaptively identified without any prior knowledge of the bearing fault type by using an enhanced WPT-based kurtogram algorithm. In addition, the proposed VMDC-based discriminative feature analysis approach efficiently identifies optimal features where their discriminatory power is maximized. The proposed methodology was evaluated by an experimental setup using a healthy bearing and seven damaged ones under different rotational speeds and different simulated crack sizes. Experimental results indicated that the proposed multifault diagnosis methodology achieves the highest classification accuracy up to 91.11%, 96.67%, 98.89%, 99.44%, and 98.61% under bearing rotational speeds of 300 rpm, 350 rpm, 400 rpm, 450 rpm, and 500 rpm, respectively.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgment

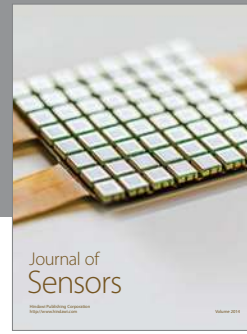
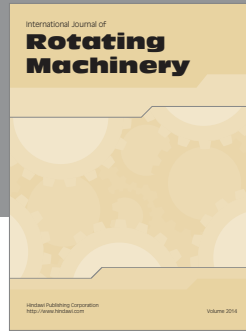
This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (no. NRF-2013R1A2A2A05004566).

## References

- [1] J. Antoni, "Fast computation of the kurtogram for the detection of transient faults," *Mechanical Systems and Signal Processing*, vol. 21, no. 1, pp. 108–124, 2007.



- [2] N. Sawalhi, R. B. Randall, and H. Endo, "The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis," *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2616–2633, 2007.
- [3] T. Barszcz and R. B. Randall, "Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine," *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 1352–1365, 2009.
- [4] P. Bošković, J. Petrovičić, B. Musizza, and D. Jurišić, "Detection of lubrication starved bearings in electrical motors by means of vibration analysis," *Tribology International*, vol. 43, no. 9, pp. 1683–1692, 2010.
- [5] R. F. Dwyer, "Detection of non-Gaussian signals by frequency domain kurtosis estimation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '83)*, pp. 607–610, Boston, Mass, USA, 1984.
- [6] J. Antoni and R. B. Randall, "The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines," *Mechanical Systems and Signal Processing*, vol. 20, no. 2, pp. 308–331, 2006.
- [7] B. Chen, Z. Zhang, Y. Zi, Z. He, and C. Sun, "Detecting of transient vibration signatures using an improved fast spatial-spectral ensemble kurtosis kurtogram and its applications to mechanical signature analysis of short duration data from rotating machinery," *Mechanical Systems and Signal Processing*, vol. 40, no. 1, pp. 1–37, 2013.
- [8] L. Shi, R. B. Randall, and J. Antoni, "Rolling element bearing fault detection using improved envelope analysis," in *Proceedings of the Eighth International Conference on Vibrations in Rotating Machinery*, pp. 301–311, Swansea, Wales, September 2004.
- [9] M. Kang, J. Kim, J.-M. Kim, A. C. C. Tan, E. Y. Kim, and B.-K. Choi, "Reliable fault diagnosis for low-speed bearings using individually trained support vector machines with kernel discriminative feature analysis," *IEEE Transactions on Power Electronics*, vol. 30, no. 5, pp. 2786–2797, 2015.
- [10] M. Kang, J. Kim, B.-K. Choi, and J.-M. Kim, "Envelop analysis with a genetic algorithm-based adaptive filter bank for bearing fault detection," *The Journal of the Acoustical Society of America*, vol. 138, no. 1, article EL65, 6 pages, 2015.
- [11] A. Widodo and B.-S. Yang, "Machine health prognostics using survival probability and support vector machine," *Expert Systems with Applications*, vol. 38, no. 7, pp. 8430–8437, 2011.
- [12] Y. Guo, J. Na, B. Li, and R.-F. Fung, "Envelope extraction based dimension reduction for independent component analysis in fault diagnosis of rolling element bearing," *Journal of Sound and Vibration*, vol. 333, no. 13, pp. 2983–2994, 2014.
- [13] J. Harmouche, C. Delpha, and D. Diallo, "Improved fault diagnosis of ball bearings based on the global spectrum of vibration signals," *IEEE Transactions on Energy Conversion*, vol. 30, no. 1, pp. 376–383, 2015.
- [14] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Computer Science and Scientific Computing, Academic Press, Boston, Mass, USA, 1990.
- [15] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.
- [16] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, Wiley, New York, NY, USA, 1992.
- [17] F. Cong, J. Chen, and G. Dong, "Spectral kurtosis based on AR model for fault diagnosis and condition monitoring of rolling bearing," *Journal of Mechanical Science and Technology*, vol. 26, no. 2, pp. 301–306, 2012.
- [18] K. Mardia, J. Kent, and J. Bibby, *Multivariate Analysis*, Academic Press, London, UK, 1979.
- [19] J. Vaidya, M. Kantarcioğlu, and C. Clifton, "Privacy-preserving Naïve Bayes classification," *The VLDB Journal*, vol. 17, no. 4, pp. 879–898, 2008.
- [20] Y.-L. He, R. Wang, S. Kwong, and X.-Z. Wang, "Bayesian classifiers based on probability density estimation and their applications to simultaneous fault diagnosis," *Information Sciences*, vol. 259, pp. 252–268, 2014.
- [21] J. Rodriguez, A. J. Perez, and J. Lozano, "Sensitivity analysis of k-fold cross validation in prediction error estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 569–575, 2010.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

