


Time–frequency envelope analysis-based sub-band selection and probabilistic support vector machines for multi-fault diagnosis of low-speed bearings

M. M. Manjurul Islam¹ · Jong-Myon Kim¹ 

Received: 18 June 2017 / Accepted: 23 September 2017 / Published online: 9 October 2017
© Springer-Verlag GmbH Germany 2017

Abstract This paper proposes a highly reliable multi-fault diagnosis scheme for low-speed rolling element bearings using an effective time–frequency envelope analysis and a Bayesian inference based one-against-all support vector machines (probabilistic-OAASVM) classifier. The proposed method first performs a wavelet packet transform based envelope analysis on an acoustic emission signal to select sub-bands of the signal that contain the most intrinsic and pertinent information about the defects. Frequency- and time-domain fault features are extracted only from selected sub-bands for fault classification. Traditional one-against-all SVMs (OAASVM), a widely used multi-class pattern recognition technique, employ an arbitrary combination of a series of binary classifiers yielding overlapped feature spaces, where a data sample can be unclassifiable. To address this limitation, we formulate the feature space of OAASVM as an appropriate Gaussian process prior (GPP) and interpret OAASVM results as a posterior probability estimation procedure using Bayesian inference under this GPP. The efficacy of the proposed probabilistic-OAASVM classifier is verified for low-speed rolling element bearings under various conditions. Experimental results indicate that the proposed method outperforms the state-of-the-art algorithms for multi-fault classification of low-speed bearings, yielding a 4.95–20.67% improvement in the average classification accuracy.

Keywords Acoustic emission · Wavelet packet transform · Envelope analysis · Fault detection and diagnosis

1 Introduction

The energy crisis as well as environmental pollution are both driving the development of renewable energy sources. Wind energy conversion systems (WECSs) are becoming more popular as an economically viable alternative to fossil-fuel based power generation. WECS consisting of thousands of units are now forming a major portion of devices with renewable electrical generating capacity and play a vital role in the future of global energy generation (de Bessa et al. 2016; Kandukuri et al. 2016). As the size of WECS continues to increase, their high maintenance costs and associated failure costs become increasingly important issues. A WECS usually includes four main components: the wind turbine (WT), generator, control systems, and an interconnection apparatus (Kandukuri et al. 2016). Among them, the WT is the element that fails most frequently (de Bessa et al. 2016). More specifically, bearing defects account for the highest percentage of all failures in wind turbines because of their contribution to coarse operating conditions and other external influences, such as the ratio of high torque to low-speed, vibration, improper loading, and misalignment (Kandukuri et al. 2016; Kang et al. 2015b). Because abrupt mechanical failures affected by the bearing faults in WECSs result in a substantial economic loss, the reliability of the bearing fault diagnosis framework is an increasingly vital concern (Kandukuri et al. 2016). The essence of a fault diagnosis scheme is the data (or signal) acquisition, feature extraction, and classification (Finogeev et al. 2017; Islam et al. 2017).

✉ Jong-Myon Kim
jongmyon.kim@gmail.com

M. M. Manjurul Islam
m.m.manjurul@gmail.com

¹ School of Electrical, Electronics, and Computer Engineering, University of Ulsan, Ulsan, South Korea

Motor current signature analysis (Hamadache et al. 2015), vibration techniques (Dong et al. 2015), temperature tests (Hamadache et al. 2015), and wear debris analysis have traditionally been used in the diagnosis of bearing faults and have shown improved performance over time. A wide range of research has been carried out considering vibration, especially for diagnosing high-speed machineries (Chen et al. 2014; Hamadache et al. 2015). Acoustic emissions (AE) detection is the latest technique in diagnosing faults in rolling element bearings (REBs) (He and Zhang 2012; Li and He 2012). The principal advantage of the AE technique over traditional vibration detection is that the former has a much better signal to noise ratio (SNR), even at very low frequencies, making it particularly suitable for detecting a possible failure at a very early stage. This study therefore records AE signals to classify single and multiple combined faults in the early stages of crack development under the low-speed operation of rolling element bearings.

Whenever any defect such as a crack or spoil occurs on any of the four different bearing elements (i.e., outer raceway, inner raceway, rollers or balls, and train or cage), it creates harmonics among the bearing characteristic (defect) frequencies [i.e., ball pass frequency inner raceway (BPFI), ball pass frequency outer raceway (BPFO), and ball spin frequency (BSF)] for each shaft rotation (Dong et al. 2015; Islam et al. 2017). It is important to note that a bearing defect symptom is hardly found around the raw and unfiltered harmonics of the defect frequencies in the original fault signal's power spectrum since it is an inherently nonstationary and nonlinear signal (Dong et al. 2015; Kang et al. 2015a). Frequency analysis and demodulation are further needed. This overall process, called envelope analysis (Dong et al. 2015), focuses on the transient, impact-type events (spikes on the time domain signal) such as BPFI, BPFO and BSF, while the fast Fourier transform (FFT) process otherwise misses such transient events because of the way the FFT processes inherently nonstationary and nonlinear fault signals. Several researchers, therefore, explored impacts in the envelope power spectrums acquired from various sub-band signals using either short-time Fourier transform (STFT) (Jeong et al. 2015; Lalani and Doye 2017) or multi-level band pass filters (Kang et al. 2015a; Wang et al. 2013).

Another important issue is to clarify pertinent and informative sub-band signals, *one of the key contributions of this study*, from a large input signal. The sub-band signals are further utilized to extract meaningful fault features for accurate classification of defects. In Wang et al. (2013), recently introduced a wavelet based kurtogram as a time–frequency analysis, which is broadly used to find useful sub-band signals since it can quantify the magnitudes of the rolling bearing defect frequencies BPFI, BPFO and BSF as well as their harmonics. However, this quantifying parameter is not precisely proportional to the degree of defectiveness of

bearing rolling elements. To solve this problem, we consider a Gaussian mixture model-based degree of defectiveness ratio (DDR) calculation, which is a ratio of defect-components to residual-components, instead of merely using a kurtosis value (Wang et al. 2013) in the envelope power spectrum of wavelet packet transform (WPT) nodes. The main concept of the DDR calculation is that it first generates a Gaussian window around the BPFI, BPFO, and BSF as well as their harmonics, and then calculates the DDRs about these defect frequencies. This evaluation metric provides a very efficient and meaningful way to accurately measure the degree of defectiveness. Further, the highest DDR values about BPFI, BPFO, and BSF of the 2D visualizations of WPT nodes are selected as the most informative sub-bands.

2-D visualization based sub-band selections are apparently effective for finding appropriate fault conditions, while most of the existing fault diagnosis studies for bearing elements (Dong et al. 2015; He and Zhang 2012; Kang et al. 2015a; Li and He 2012; Skolidis and Sanguinetti 2011; Wang et al. 2013) are confined only to the visual inspection of a fault trend in some form of spectrum view, with no classifier employed at all to identify fault types. This paper focuses on informative sub-band signals selection using wavelet packet transform based envelope analysis (WPT-EA) with DDR, as well as fault feature extraction from these informative sub-bands; this paper classifies faults using a Bayesian one-against-all support vector machine (probabilistic-OAASVM) classifier. To calculate meaningful fault feature components, this study searches the inherently nonstationary and nonlinear AE signals via WPT-EA, and selects the useful sub-band signals based on the highest DDR value in the 2-D visualization tool, and then features are extracted from these selected sub-bands. However, while WPT-EA with DDR prepares reliable fault features, the ultimate diagnostic performance strongly depends on the fault classification accuracy when these features are further utilized with classifier methods, for example, Naïve Bayesian (Hyun-Chul and Ghahramani 2006), k-NN (Yala et al. 2017), artificial neural networks (ANNs) (Chen et al. 2014; Khoobjou and Mazinan 2017), and support vector machines (SVMs) (Aydin et al. 2011; Islam et al. 2015). The SVM method is the most extensively used classifier technique in many real-world applications because of its high generalization performance. However, extending the SVM methodology, which was originally designed for binary class classification, into multi-class classification still remains a fundamental research issue (Abe 2015; Chih-Wei and Chih-Jen 2002). The main inherent problem in the traditional one-against-all SVM (traditional-OAASVM) is that arbitrary combinations of binary SVMs yield overlapped regions where a data point might be unclassifiable, implying that the point either gets rejected (negative response) by all classes or accepted (positive response) by more than one class; such a drawback can

severely degrade classification performance and render the diagnosis method ineffective.

To address these limitations, several methods have been proposed to compensate for issues regarding the unreliability of the traditional-OAASVM (Abe 2015; Chakrabarty and Cauwenberghs 2007; Islam et al. 2015; Nasiri et al. 2015). Recently, the Dempster–Shafer (D–S) theory-based evidence reasoning technique (Islam et al. 2015) introduced a static reliability measure to each binary classifier for an OAASVM to improve the classification performance, but this method does not consider stochastic information handling based on statistical inference. Especially, Abe recently proposed “fuzzy support vector machines (FSVM)” to improve the reliability issues of OAASVMs, in which he introduced a membership value calculation associated with overlapped regions during the training phase (Abe 2015). One of the major shortcomings of a training phase membership calculation is that this reliability measure is an offline manner and provides a mere static and binary value about the class competence, regardless of the location of a test sample.

Though the previous studies show progress, their shortcomings still motivate us to further improve the traditional-OAASVM for superior classification performance. As addressed, OAASVM classifications generally consider arbitrary combinations among classes that leave undecided large feature spaces where many samples are unaccounted for, which is a situation that results in no probabilistic interpretations of class outputs. This is further complicated by the fact that reliable diagnosis entails a large grain of uncertainty, especially related to unusual failures. Quantifying and managing this uncertainty incurs substantial overhead. In particular, artificial neural network (ANN) based approaches have often been used as fault classifiers both for binary fault classification and for multi-class fault classification (Chen et al. 2014). The shortcomings of ANNs is that they are black-box devices where the solution of ANN schemes are not globally optimal and the reasons for the solutions are impossible to ascertain (Chen et al. 2014). Therefore, a novel Bayesian inference-based on the one-against-all support vector machines (probabilistic-OAASVM) classifier, another major contribution of this study, is suggested that interprets the OAASVM as a maximum a posterior (MAP) based evidence function using an appropriate formulation of feature spaces in a Gaussian process prior (GPP), and then Bayesian inference is principally applied to estimate class probability for the unknown sample using this evidence function.

The remaining part of this paper is organized as follows. Section 2 presents support vector machines and multiclass schemes along with their shortcomings. Section 3 describes the proposed reliable fault diagnosis methodology with data acquisition system, sub-band analysis, and the proposed probabilistic-OAASVM classifier. Section 4 gives results and discussions, and Sect. 5 concludes this paper.

2 Support vector machines (SVM) and the multiclass approach

SVMs provide an efficient classifier approach and have shown substantial success in the diagnosis of many real-world applications because of their capability for generalization and robust control over unknown data distributions. The earliest and most widely used multi-class extension is the traditional-OAASVM [for example, (Chih-Wei and Chih-Jen 2002)].

Thus, to define traditional-OAASVM, let’s consider an m -class classification problem with dataset Q , having n data samples in the form of $Q = \{(x_i, y_i) | x_i \in R^d, y_i \in \{1, -1\}\}_{i=1}^n$, where $x_i \in R^d$ is the feature vector dimension. Hence, the traditional-OAASVM classifier creates m binary SVM classifiers, each of which separates one class from the rest. Thus, mathematically the k th SVM solves the optimization problem in (1), to find the minimum value of weight vector ω and bias b , so that a linearly separable hyperplane can be found (Aydin et al. 2011; Chih-Wei and Chih-Jen 2002):

$$\min_{\omega_i, b_i, \zeta_i} \left\{ \frac{1}{2} \|\omega_i\|^2 + C \sum_{j=1}^n (y_j [\omega \cdot \varphi(x_i) + b_i]) \right\} \quad (1)$$

subject to, $y_j(\omega_i \cdot \varphi(x_i) + b_i) \geq 1 - \zeta_j^i$,
and, $\zeta_j^i \geq 0, \forall_i = 1, 2, \dots, n$.

Here, ω is the norm vector to the hyperplane and b is the constant bias, such that the margin width, $2/\|\omega\|$, between the hyperplane is at a maximum. Equation (1) includes a training error ζ_j , while an optimal hyperplane can be found by adjusting the training error by a penalty parameter C . According to Aydin et al. (2011), φ can be defined as a mapping function that can map the original feature to a high dimensional space so that a linearly separable hyperplane can be obtained, and then it forms the dual optimization in (2), where the Lagrange multiplier is α_i .

$$\max_{\alpha_i} \left\{ \sum_i \alpha_i - 1/2 \sum_j \alpha_i \alpha_j y_i y_j \varphi(x_i)^T \varphi(x_j) \right\}_{i=1, j=1}^n \quad (2)$$

For any two different samples x and x_j , the inner product (\cdot) of the two samples vector in Eq. (2) can be used to generate a kernel function as below (Aydin et al. 2011):

$$\varphi(x) \cdot \varphi(x_j) = K(x, x_j) \quad (3)$$

This $K(x, x_j)$ defines the kernel function of the SVM and SVM decision function can be found in (4) after solving the dual optimization problem. The respective real-valued decision functions of Eq. (4) can be defined in Eq. (5). An

unknown test sample x can then be classified in class i for which the decision function $f_i(x)$ has the highest value in (4) for the sign value (1 or 0), or in Eq. (5) for the real-valued output; the corresponding class label is given in Eq. (6).

$$f_i(x) = \text{sgn}(\omega \cdot \varphi(x) + b) = \text{sgn}\left(\sum_i y_i \alpha_i K(x, x_i) + b\right). \tag{4}$$

$$f_i(x) = (\omega_i \cdot \varphi(x) + b) = \left(\sum_i y_i \alpha_i K(x, x_i) + b\right). \tag{5}$$

$$k = \underset{i = 1, 2, \dots, m}{\text{arg max}} f_i(x). \tag{6}$$

To show a representative example contributing to the unreliability problem, the obtained decision boundaries of traditional-OAASVM for three example classes are $f_1, f_2,$ and f_3 and are shown in Fig. 1. In Fig. 1, overlapping feature regions can be seen ($R_1, R_2, R_3,$ and R_4 in the shaded lines) between these decision boundaries in the left figure. Figure 1 (in the right figure) depicts the actual problem, in which many test samples in the overlapped region are misclassified due to their vicinity to the decision boundary of the opposite class decision boundary, shown by dotted arrows in the same figure. That means a sample can have more than one positive decision value (i.e., in the $R_1, R_2,$ and R_3 regions) or even all negative decision values (i.e., in the R_4 region) as in the traditional-OAASVM. However, the traditional-OAASVM requires that a sample can be classified within a certain class if and only if one SVM accepts it and rest of the SVMs reject the sample (Abe 2015; Chih-Wei and Chih-Jen

2002; Islam et al. 2015). Thus, this unreliability issue can severely impact the overall classification accuracy for many practical applications, including multi-fault diagnosis in rolling element bearings. Therefore, improving the classification performance and defining this uncertainty problem of the OAA SVM requires having the appropriate prior knowledge about the class distribution and further requires defining the OAASVM as a maximum a posteriori probability (MAP) estimation procedure using Bayesian inference according to this prior knowledge. Comprehensive and relevant derivations regarding this process are defined in Sect. 3.

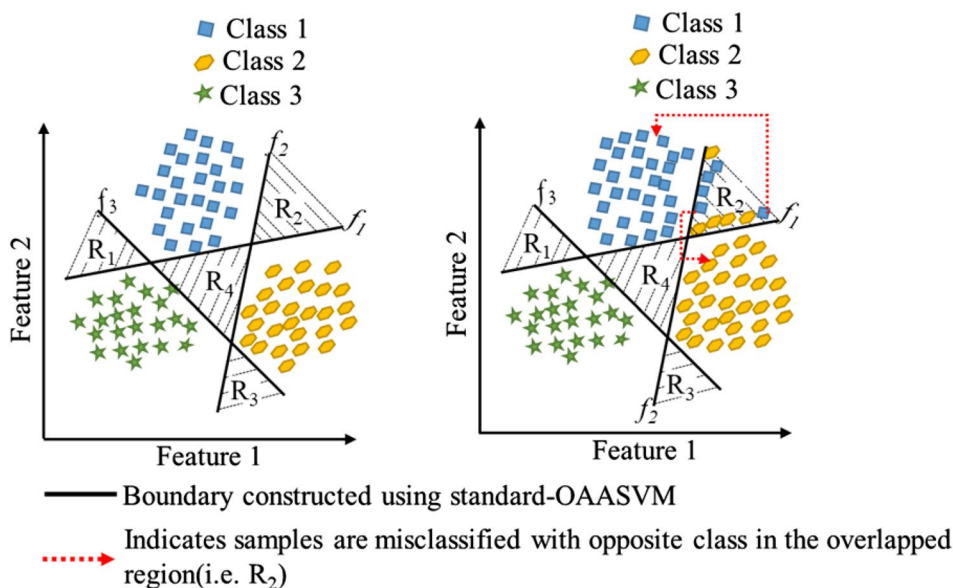
3 Proposed methodology for fault diagnosis

The detailed methodology of the developed fault diagnosis scheme is presented in Fig. 2, which consists of acoustic emission signals (or data) acquisition system, an effective envelope analysis to find informative sub-bands correlated to bearing defect conditions, feature extraction from the selected sub-bands, and a probabilistic-OAASVM classifier for the classification faults.

3.1 Experiment setup and AE data acquisition

In this paper, we present a robust bearing fault diagnosis scheme for the low-speed bearing for verifying whether the proposed fault diagnosis scheme is useful. In the experiment, widely used sensors and equipment are used. Figure 3 illustrates the designed experiment setup. In the figure, a three-phase induction motor is placed in the drive end shaft (DES) and the bearing house is attached to the motor shaft at a gear reduction ratio of 1.52:1, and a WS α AE sensor is

Fig. 1 Indecisive regions (shaded area marked by $R_1, R_2, R_3,$ and R_4) using traditional-OAASVM (left figure), and the problem of indecisive regions where samples are misclassified (right figure)



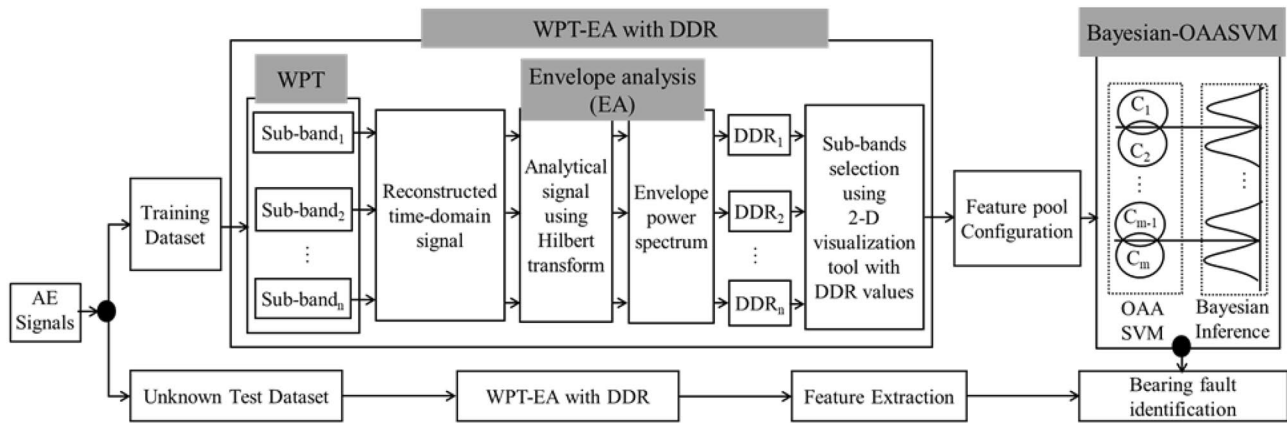


Fig. 2 The overall methodology of a reliable low-speed bearing fault diagnosis scheme

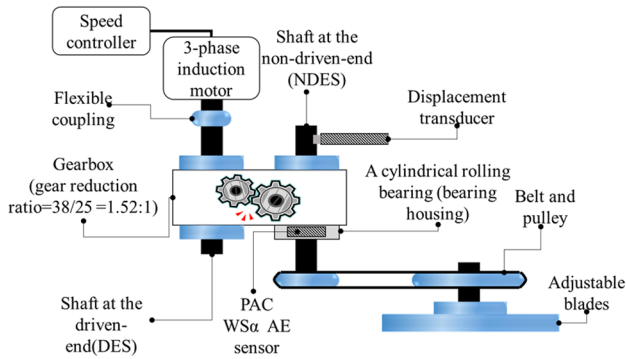


Fig. 3 An experimental setup for low-speed bearing fault diagnosis

placed over the bearing house in the shaft at the non-driven-end (NDES).

To capture intrinsic information about seven defective-bearings (see Fig. 5) and one bearing-with-no-defect (BND) conditions, this study records AE signals at a 250-kHz sampling rate at two different rotational speeds (in rpm) and two different crack sizes (in mm) using a PCI-2 system. Table 1 present the detailed discretion of recorded dataset. The experimental hardware setup and PCI-based data acquisition module are presented in Fig. 4. Figure 4b depicts the PCI-based data acquisition that is utilized to record the AE signal from this setup. The effectiveness of our data acquisition scheme have been studied in (Islam et al. 2015).

Each dataset has signals of eight bearing conditions (single and multiple-combined-faults), which corresponding to the location of cracks, i.e., normal condition (BND), outer raceway crack (BCO), inner raceway crack (BCI), roller crack (BCR), inner and outer raceway cracks (BCIO), outer and roller cracks (BCOR), inner and roller cracks (BCIR), and inner, outer, and roller cracks (BCIOR), as shown in Fig. 5. Additionally, Fig. 6 presents the recorded AE signal

Table 1 Acoustic emission (AE) data acquisition using two different operational conditions at two crack sizes

Dataset ^a	Average rotational speed (rpm)	Crack sizes in the position of bearing's outer, inner, roller raceways.		
		Length (mm)	Width (mm)	Depth (mm)
Dataset 1	300	3	0.35	0.30
Dataset 2	500			
Dataset 3	300	12	0.49	0.50
Dataset 4	500			

^a90 AE signals (or data sample) are recorded for each condition (i.e., normal bearing, single and multiple-combined fault) in each dataset. Signal sampling rate, $f_s = 250$ kHz with a 10-s signal length

of each bearing condition for dataset 1 (see Table 1). Each bearing condition represents a unique pattern, as shown in Fig. 6.

3.2 WPT-EA with DDR to select informative sub-bands regarding bearing defects

As described in Sect. 1, bearing characteristic frequencies (for defects) are more observable in the envelope signal than in the fast Fourier transform (FFT) of the original AE signal. However, it is still an important issue to determine which sub-bands have pertinent information in the 10 s AE signal regarding bearing defects. This is accomplished with a wavelet packet transform based envelope analysis (WPT-EA), in which sub-bands with useful signals are searched to determine the DDR. The flowchart of the proposed method is illustrated in Fig. 7, with detailed steps. First, the input AE signal bearing defect signal is decomposed into a series of sub-bands by applying the wavelet packet transform (WPT) with five decomposition levels using a Daubechies 2 (or db2) filter (Jeong et al. 2015;

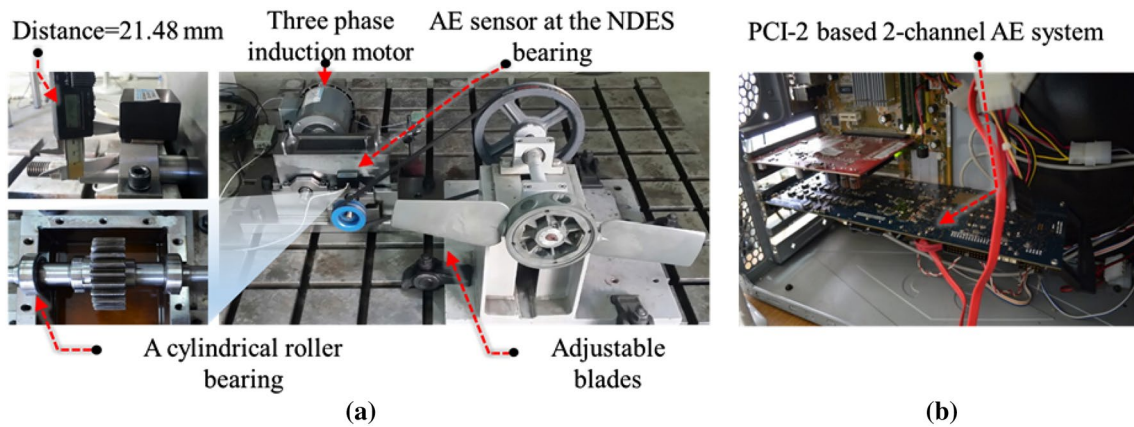


Fig. 4 **a** A screenshot of the self-designed experimental setup of the reliable low-speed rolling element bearings fault diagnosis system. **b** PCI based AE signal acquisition system

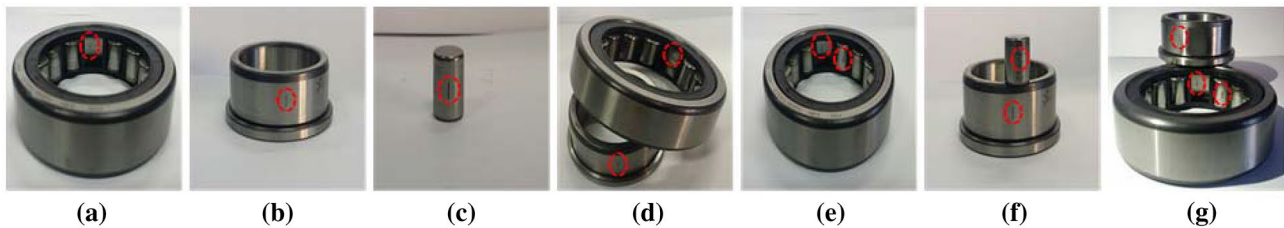


Fig. 5 Examples of bearing defects: **a** BCO, **b** BCI, **c** BCR, **d** BCOI, **e** BCOR, **f** BCIR, and **g** BCOIR in dataset 1

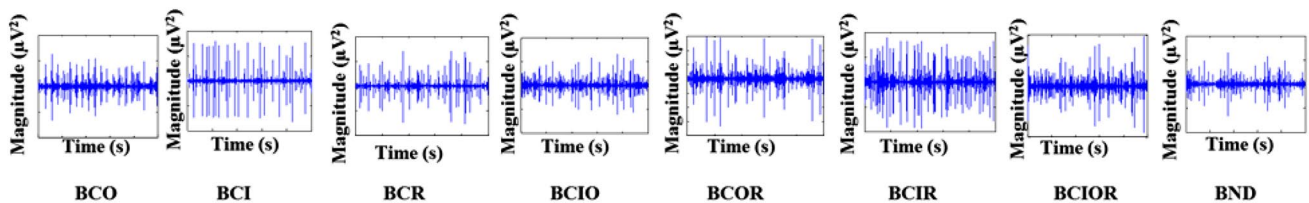


Fig. 6 Acquired original AE signals of different bearing conditions in dataset 1

Wang et al. 2013). This five-level WPT decomposition yields a total of 63 sub-band signals. Further, we calculate the envelope spectrum for each sub-band since defect symptoms are most observable in the envelope power spectrum. In order to quantify the degree to which each sub-band is informative, we compute the DDR (the ratio of the defect components to the residual components) value, instead of a mere kurtosis value in (Wang et al. 2013), to measure the degree of defectiveness in the envelope spectrum to characterize the hidden bearing defect signatures. In order to do this effectively, we first determine the bearing defect frequencies, which include the BPFO, BPF1, and BSF, and the first H harmonics (H is a value up to 3 in this paper) of each of these frequencies. These defect frequencies are defined as follows:

$$\begin{aligned}
 BPFO &= \frac{N_r \cdot S_{shaft}}{2} \left(1 - \frac{B_d}{P_d} \cos a \right), \\
 BPF1 &= \frac{N_r \cdot S_{shaft}}{2} \left(1 + \frac{B_d}{P_d} \cos a \right), \text{ and,} \\
 BSF &= \frac{P_d \cdot S_{shaft}}{2 \cdot B_d} \left(1 - \left(\frac{B_d}{P_d} \cos a \right)^2 \right),
 \end{aligned}
 \tag{7}$$

where *BPFO*, *BPF1*, and *BSF* define the bearing characteristic (defect) frequencies depending on whether there is a crack on the inner, outer, and roller raceways, respectively. *N_r* defines the number of rollers, *S_{shaft}* is the shaft speed, *P_d* and *B_d* are the pitch and roller diameters, respectively, and *a* is the contact angle.

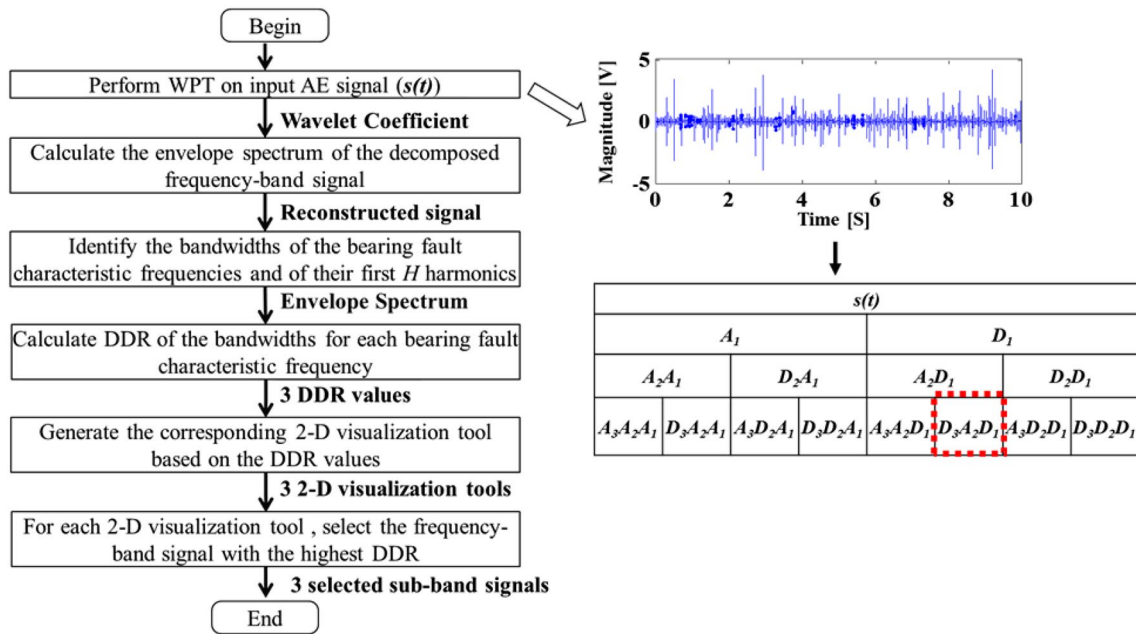


Fig. 7 Flowchart of the WPT-based envelope analysis (WPT-EA) with DDR to select informative sub-bands regarding bearing defects

The DDR calculation is applied to each node (i.e., red dotted $D_3A_2D_1$ in Fig. 7) of the WPT nodes, and then all DDR values are presented in the 2-D visualization tool. Further, the signal is analyzed, with an assumption that it could simultaneously contain all the possible bearing failures (i.e., BPFO, BPFI, and $2 \times$ BSF) since our purpose is also to diagnosis multiple-combined faults. Consequently, the WPT-EA outputs three informative sub-bands regarding BPFO, BPFI, $2 \times$ BSF, since these outputs are acquired from a 2-D visualization tool having the highest DDR values and their corresponding signals, which are used for fault features extraction. Further detailed steps of the DDR calculation are presented in Sect. 3.3.

3.3 DDR calculation

Figure 8 presents the general concepts and all the steps of the DDR calculation for all WPT nodes.

- *Step-1* We apply a Hilbert transform (HT) (Jeong et al. 2015; Kang et al. 2015a) on each segmented node of the input signal, obtained from WPT decomposition, to calculate the envelope signal. For a given a reconstructed time-domain signal $s(t)$ of a WPT node, its corresponding HT can be calculated as follows:

$$\hat{s}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{s(\tau)}{\tau} d\tau. \tag{8}$$

Here, t is the time and $\hat{s}(t)$ defines the HT of $s(t)$. By adding $s(t)$ and $\hat{s}(t)$, the analytical signal $a(t)$ can be computed as follows:

$$a(t) = s(t) + i\hat{s}(t), \text{ where } i = \sqrt{-1} \tag{9}$$

As the analytical signal is defined, the envelope signal of a given time domain signal $s(t)$ can be calculated by taking the absolute value of the analytical signal, denoted as $\|a\|$. Hence the envelope power spectrum is obtained by taking the square of the absolute value of the FFT of the envelope signal where defect frequencies of bearing failures are easily discerned. This envelope power spectrum reveals modulation in signals that are caused by bearing defects while it removes carrier signals, which may reduce the effects of unwanted information regarding bearing fault detection. This process is illustrated in Step 1 of Fig. 8.

- *Step-2* A Gaussian mixture model (GMM)-based window (w_{gmm}) is constructed around defect peaks and their integer multiples of the defect frequency to obtain residual components in the frequency domain of the envelope power spectrum. The coefficients of the GMM-based window are calculated as follows:

$$w_{gmm}(k, \delta) = \sum_{i=1}^n \exp \left(-\frac{1}{2} \left(\delta \frac{(k - H_n)^2}{\frac{N_{freq}}{2}} \right) \right), \text{ and, } H_n - f_{range} \leq k \leq H_n + f_{range}. \tag{10}$$

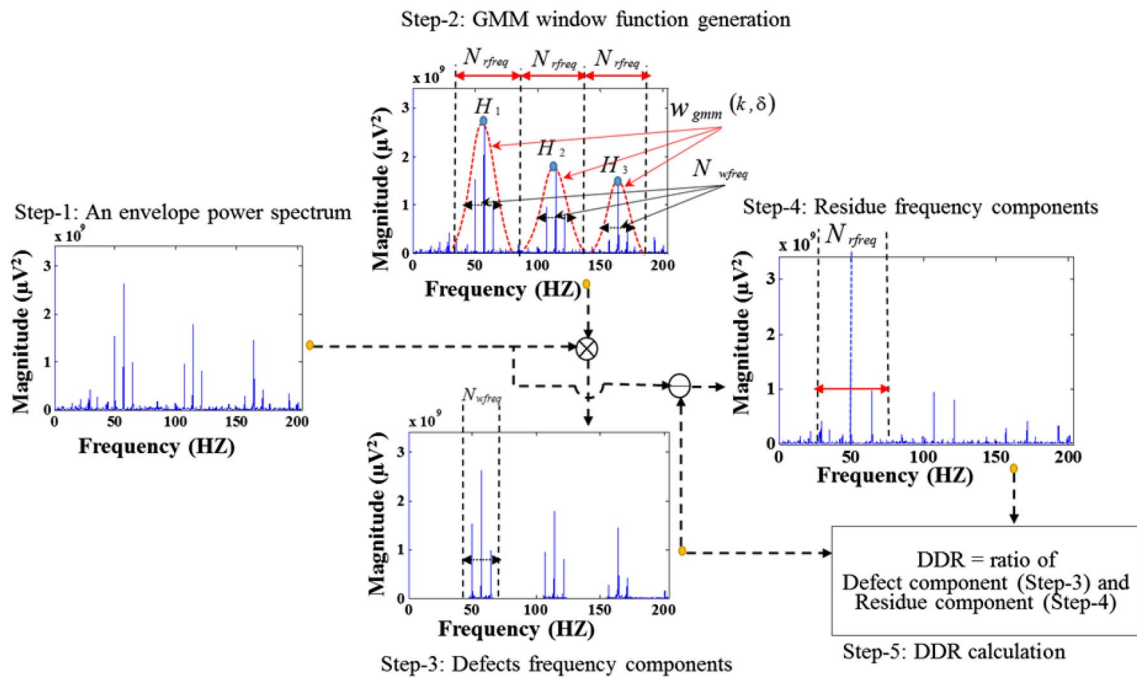


Fig. 8 Detailed framework of the DDR calculation for WPT-EA at each node

where H_n is n th harmonic of the defect frequency (or fixed valued integer multiples of the characteristic frequency) and n is the number (up to 3 in this study) of harmonics that are used to compute the DDR. N_{rfreq} defines the number of frequency bins from this range, $H_n - f_{range} \leq k \leq H_n + f_{range}$, and can be defined as:

$$N_{rfreq} = 2 \cdot f_{range} / f_{resolution} \tag{11}$$

Similarly, δ is a variable that defines the Gaussian random variables, which is inversely proportional to the traditional deviation, and it can be calculated as below:

$$\delta = (N_{rfreq} / N_{wfreq}) \sqrt{2 \ln \rho} \tag{12}$$

N_{wfreq} defines the size of the frequency bins around the defect frequency components and their harmonics (see Fig. 8 Step 2), and a fixed value of ρ , associated with the convergence of the Gaussian window in (12), is considered in the range $0 < \rho < 1$, ($\rho = 0.1$ for this study). The parameter f_{range} regulates the range of frequencies for computing the DDR value. So, it is important to find a proper range of frequencies, according to bearing dynamics, and a narrow band frequency range [i.e., $f_{range} = 1/4(\text{BPFO})$] is considered for calculating the outer race failure; on the other hand, a comparatively wider frequency range [i.e., $f_{range} = 1/2(\text{BPFO or BSF})$] is considered for both the roller and inner raceways defects.

- *Step-3* The defect frequency components are now calculated by multiplying the GMM-based window, $w_{gmm}(k, \delta)$,

around the BPFO, BPFI, or $2 \times \text{BSF}$ with their harmonics in the derived envelope spectrum, as can be seen in step 3 of Fig. 8.

- *Step-4* The residue frequency components are measured by deducting the defect frequency components (in step 3) from the envelope spectrum, as presented in step 4 of Fig. 8.
- *Step-5* Now, we have defect components and residue components, and the DDR is calculated as the ratio of defect frequencies and residue frequencies in the form below:

$$\text{DDR} = 10 \cdot \log \left\{ \sum_{n=1}^3 \left\{ \sum_{j=1}^{N_{wfreq}} M_{n,j}^2 / \sum_{j=1}^{N_{rfreq}} R_{n,j}^2 \right\} \right\} (\text{dB}) \tag{13}$$

Here, $M_{n,j}$ is the magnitude of each defect frequency around the harmonics in the GMM window, and likewise $R_{n,j}$ is the residue components of each harmonic in the range.

3.4 Feature pool configuration

It has been shown that WPT-EA is highly effective for finding informative sub-bands with information regarding bearing failures from a 10-s input signal. The three most informative sub-band signals regarding intrinsic fault symptoms are BPFO, BPFI, and, $2 \times \text{BSF}$, and this study then considers these three sub-bands for feature extraction rather than the

original 10 s AE signal. According to Kang et al. (2016), traditional statistical parameters from time- and frequency-domain signals are pertinent and useful for an intelligent fault diagnosis scheme. Table 2 displays fourteen extracted feature elements, including eight time-domain features [e.g., root means square (RMS), square root of amplitude (SRA), kurtosis values (KV), skewness factor (SK), margin factor (MF), skewness value (SV), impulse factor (IF) and peak-to-peak value (PPV)]; three frequency-domain features (e.g., RMS frequency (RMSF), frequency center (FC), and root variance frequency (RVF)); and three DDR values about outer raceway fault (DDR_{BPFO}), DDR values about inner raceway fault (DDR_{BPMFI}) and DDR values about roller raceway fault (DDR_{2×BSF}) for each sub-band signal. The dimension of the feature pool is $N_{class} \times N_{samples} \times N_{features}$, where N_{class} is the number of classes (8 in this study), $N_{samples}$ is the number of samples of each class (90 in this study), and $N_{features}$ defines the number of features (42 in this study). Thus, this 42-feature vector is considered for validating the proposed probabilistic-OAASVM classifier by accurately identifying faults.

3.5 Proposed probabilistic-OAASVM classifier

Consider an l -class classification problem with the dataset $Q = \{(x_i, y_i) | x_i \in R^d\}_{i=1}^n$, where $x_i \in R^d$ is a d -dimensional feature vector, $y_i \in \{1, 2, \dots, l\}$ is the set of class labels, and n is the number of data points in the training dataset. In the traditional OAASVM, the following optimization problem is solved to distinguish a particular class $k = 1$ from the remaining $l - 1$ classes (Chih-Wei and Chih-Jen 2002).

$$\begin{aligned}
 & \underset{\omega, b}{\text{minimize}} \left\{ \frac{1}{2} \|\omega\|^2 + C \sum_{j=1}^n (y_j [\omega \cdot \varphi(x_j) + b]) \right\} \\
 & \text{subject to,} \quad y_j (\omega \cdot \varphi(x_j) + b) \geq 1 - \zeta_j^i, \quad \text{if } y_j = i \\
 & \text{and,} \quad \zeta_j^i \geq 0, \quad \forall i = 1, 2, \dots, n.
 \end{aligned} \tag{14}$$

Here, b is the bias; ω is the weight vector; $\varphi(x_j)$ is the kernel function that maps input feature vectors x_j to a high-dimensional space, where they are linearly separable by a hyperplane with a maximum margin of $b/\|\omega\|$; and C is the linearity constraint. During classification, the traditional OAASVM labels a data point x as i^* if the decision function f_{i^*} generates the highest value for i^* , as given in Eq. (15).

$$i^* = \arg \max_{i=1,2,\dots,l} f_i(x) = \arg \max_{i=1,2,\dots,l} (\omega_i^T \varphi(x) + b_i) \tag{15}$$

The optimization problem in Eq. (14) can be defined as a maximum a posteriori (MAP) evidence function under appropriate prior distributions, and then used to estimate the class probabilities of the data points in the ambiguously labeled regions by means of Bayesian inference (Murphy 2012). In Eq. (14), $\psi(x) = \omega \cdot \varphi(x) + b$ is the only data-dependent function, in which ω and b appear separately, and it is reasonable to define a joint prior distribution over ω and b . We assume this joint distribution over $\psi(x)$ to be Gaussian, with covariance $\psi(x) \psi(x') = \langle (\varphi(x) \cdot \omega) (\omega \cdot \varphi(x')) \rangle + v^2 = \varphi(x) \cdot \varphi(x') + v^2$, where v^2 and v are the variance and traditional deviation of b , respectively. Then, the support vector machine (SVM) can be defined as a Gaussian process prior (GPP) with zero mean, over the function ψ . The covariance of ψ is defined as the kernel function $K(x, x') = \hat{K}(x, x') + v^2$, where $\hat{K}(x, x') = \varphi(x) \cdot \varphi(x')$ has zero mean (Rasmussen and Williams 2006; Sollich 2000). The probability of obtaining the output y for a given data sample x is given as follows:

$$P(y = \pm 1 | x, \psi, b) = \kappa(C) e^{-C(y(\psi(x)+b))}. \tag{16}$$

The normalization constant $\kappa(C)$ is chosen in a way such that the probability for $y = \pm 1$ never sums to a value larger than one. The likelihood probability for all the data points of class j based on the prior probability $P(x_i)$ and the conditional probability $P(y_j | x_i, \psi)$ is given by Bayesian inference in $P(Q | \psi_j) = \prod_i P(x_i) p(y_j | x_i, \psi)$. The maximum a posteriori

Table 2 Extracted time- and frequency-domain features from the selected sub-bands

Time-domain statistical parameters			
$RMS = \left(\frac{1}{N} \sum_{i=1}^N x_i^2 \right)^{1/2}$	$SRA = \left(\frac{1}{N} \sum_{i=1}^N \sqrt{ x_i } \right)^2$	$KV = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma} \right)^4$	$SF = \left(\frac{1}{N} \sum_{i=1}^N x_i^2 \right)^{1/2} / \frac{1}{N} \sum_{i=1}^N x_i $
$MF = \max(s_i) / \frac{1}{N} \sum_{i=1}^N s_i $	$SV = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma} \right)^3$	$IF = \frac{\max(s_i)}{\frac{1}{N} \sum_{i=1}^N s_i }$	$PPV(x) = \max(x_i) - \min(x_i)$
Frequency-domain statistical parameters			
$RMSF = \left(\frac{1}{N} \sum_{i=1}^N f_i^2 \right)^{1/2}$	$FC = \frac{1}{N} \sum_{i=1}^N f_i$		$RVF = \left(\frac{1}{N} \sum_{i=1}^N (f_i - FC)^2 \right)^{1/2}$

In addition to above, DDR values for BPFO, BPMFI, and BSF are also included in the feature pool

(MAP) evidence function can be obtained using Eq. (17), which is analogous in the formulation to SVM regression (Hyun-Chul and Ghahramani 2006). This conception is depicted in Fig. 9 where feature spaces are defined by the appropriate prior. Further, this relation also bears a resemblance to SVMs and GPs. A similar relation can be found in the literature, where feature spaces are defined by the kernel function (Rasmussen and Williams 2006; Smola et al. 1998).

$$f(\psi, b) = \ln P(\psi_j | Q) = -\frac{1}{2} \sum_{x, x'} K^{-1}(x, x') \psi_j(x') - C \sum_{x \in X} (y_j [\psi_j(x) + b_j]) + \text{const} \quad (17)$$

Here, $K^{-1}(x, x')$ is the inverse of the covariance matrix $K(x, x')$. The SVM algorithm finds the maximum about $f(\psi, b)$ by differentiating Eq. (17) with respect to $\psi(x)$. The non-training input samples imply that $\sum_{x'} K^{-1}(x, x') \psi(x') = 0$

at the maximum. The derivative of Eq. (17) with respect to $\psi(x)$ can be simplified into Eq. (18), which defines the MAP evidence function of the proposed probabilistic OAASVM classifier.

$$\psi_j^*(x) = \sum \alpha_i y_i K(x, x_i) \quad (18)$$

Here, α_i is an optimum value for which ψ_j^* is maximized. Using the evidence function in Eq. (18), the class probability of an unknown test sample x for class j can be calculated as the average over the posterior distribution of the function $\psi_j(x)$ as follows:

$$P(y|x, Q) \approx P(y|x, \bar{\psi}_j^*(x)), \quad (19)$$

where $\bar{\psi}_j^*$ is the expectation of the evidence function in Eq. (19), which is determined using a sampling technique (Hyun-Chul and Ghahramani 2006). Thus, the posterior average of class j , as given in Eq. (20), can be written as a linear combination of the posterior expectation of the evidence function ψ_j^* as follows:

$$\bar{\psi}_j^*(x) = P(x) = \sum \bar{\alpha}_i y_i K(x, x_i). \quad (20)$$

Hence, in the probabilistic-OAASVM classifier, an unknown test sample x is labeled as j^* , where j^* is the value of j for which the corresponding classifier provides the highest probabilistic decision value of $\bar{\psi}_j^*(x)$, as given in Eq. (21).

$$j^* = \max_{j=1,2,\dots,l} \bar{\psi}_j^*(x) \quad (21)$$

Here, l (8 in this study) is the number of fault classes. Equation (21) is the probabilistic decision function, as

opposed to the decision function in Eq. (6) that is employed in the traditional-OAASVM.

3.6 Fault classification

As indicated in the fault diagnosis scheme in Fig. 2, the main goal of this study is to classify faults using a new decision function of the proposed probabilistic-OAASVM classifier in (21). Further, this study compares the classification performance with state-of-art classifiers, such as the traditional-OAASVM (Chih-Wei and Chih-Jen 2002) and FSVM (Abe 2015).

4 Results and discussions

The effects of two main components of the proposed reliable bearing fault diagnosis scheme—WPT-EA with DDR based on subband selection and a probabilistic-OAASVM classifier with a higher classification accuracy—are analyzed and discussed in this section.

4.1 Performance evaluation of WPT-EA with DDR

Though kurtogram analysis is widely used for finding informative sub-bands regarding abnormal fault symptoms, it is still important to have an appropriate degree of defectiveness measure. This study, therefore, improves spectral kurtosis value (SKV) based sub-band analysis in (Wang et al. 2013) by developing a new evaluation metric of the DDR for the proposed WPT-EA. Figure 10 compares the result between the proposed WPT-EA with DDR in Fig. 10b and SKV in Fig. 10a. According to the figure, the proposed evaluation metric is highly efficient for finding the three informative sub-band signals of BPFO, BPFI, and $2 \times$ BSF for the outer, inner, and roller raceway faults, respectively. Another important point to note is that SKV based analysis is incapable of selecting informative sub-band information since it misses the defect frequencies, BPFO, BPFI, $2 \times$ BSF, as well as their harmonics in the corresponding sub-bands spectrum views (i.e., the right of Fig. 10a), while the proposed WPT-EA with DDR method is highly capable of finding appropriate sub-bands, as can be seen in the spectrum views (i.e., the right of Fig. 10b).

WTP-EA with DDR yields three informative sub-bands that are utilized for feature extraction for fault classification. The effectiveness of the feature extraction process in the selected sub-bands is shown in Fig. 11. It is important to note that this feature extraction process clearly encodes the appropriate fault conditions since the separation among fault classes is augmented as the rpm and crack sizes are increased from dataset 1 to 4. These feature elements are

further utilized for the proposed probabilistic-OAASVM classifier for performance evaluation.

4.2 Performance evaluation of the proposed probabilistic-OAASVM classifier for identifying single and multiple-combined faults

The utilization of appropriate training and test dataset configurations is an important aspect of reliable classification performance. Thus, we randomly divide an initial set of 90 samples for each fault type into two subsets: one is for training and the other is for testing. The training dataset includes 40 randomly selected samples for each bearing condition and the remaining $(90 - 40) = 50$ for testing. The size of the training data is kept lower than the test data size to ensure the reliability of the diagnosis performance. Therefore, this section verifies the efficacy of the probabilistic-OAASVM classifier approach by comparing its performance with that of three state-of-the-art algorithms, as summarized below:

- *Methodology 1* This study improves the traditional-OAASVM, the most widely used multi-class classification technique (Chih-Wei and Chih-Jen 2002). Thus, traditional-OAASVM decision output with a sign function in (4) is considered as a potential candidate to make a comparison with the proposed probabilistic-OAASVM classifier.
- *Methodology 2* In addition, the traditional-OAASVM can generate real-valued decision output in (5). Therefore, the proposed method also compares its effectiveness with the traditional-OAASVM with a real-valued decision output.
- *Methodology 3* Abe has recently proposed a fuzzy support vector machine (FSVM) (Abe 2015) to solve the unreliability problem in traditional-OAASVM by introducing a membership function associated with the undefined region. Thus, this study considers an FSVM as a potential candidate to make a comparison with the proposed classifier method.

To validate the effectiveness of the proposed probabilistic-OAASVM classifier, a set of experiments was carried out with four datasets (see Table 1) under various operating conditions with all possible combinations of single and multiple-combined faults (i.e., BCO, BCI, BCR, BCIO, BCOR, BCIR, BCIOR, and BND). Additionally, k -fold cross validation (k -cv) (Kang et al. 2016), a popular method to estimate generalized classification accuracy, is deployed to evaluate the diagnostic performance of the proposed method relative to the other three methodologies, in terms of the sensitivity and average classification accuracy (ACA), which are given below (Kang et al. 2016):

$$\text{Sensitivity} = \frac{N_{TRP}}{N_{TRP} + N_{FNR}} \times 100\%. \quad (22)$$

$$\text{ACA} = \frac{\sum N_{TRP}}{N_{TS}} \times 100\%. \quad (23)$$

Here, N_{TRP} defines the number of fault samples of a particular class j that are accurately classified as *class j*; N_{FNR} defines the number of fault samples of class j that are (not accurately) classified as *class i*; N_{TS} defines the number of test samples, and N_C is the number of fault classes (i.e. 8 for this study).

Figures 12, 13, 14 and 15 show comprehensive diagnostic performance results for dataset 1, dataset 2, dataset 3, and dataset 4, respectively. According to these results, each figure compares sensitivities of the proposed method with three conventional methods. As can be seen in Fig. 12, the proposed methodology shows an improved performance, in terms of sensitivity, for each fault type with a 90% or greater accuracy, noticeably outperforming the other three methods. This accuracy can be further validated by the fact that this dataset is recorded at an rpm that is as low as possible (i.e., 300) with a very tiny crack (i.e., 3 mm in length). Similarly, from Fig. 13, we find that the proposed methodology retains its superiority over the other three methodologies. In contrast, methodologies 1 and 3 suffer from degraded performance with regard to identifying several fault types relative to the proposed method, especially for datasets 2 through 4.

Furthermore, Figs. 14 and 15 show that the proposed methodology offers significant diagnosis performance improvement with 100% classification accuracy for several fault types. The other three methodologies, however, do not provide such significant diagnostic performance (see Table 3).

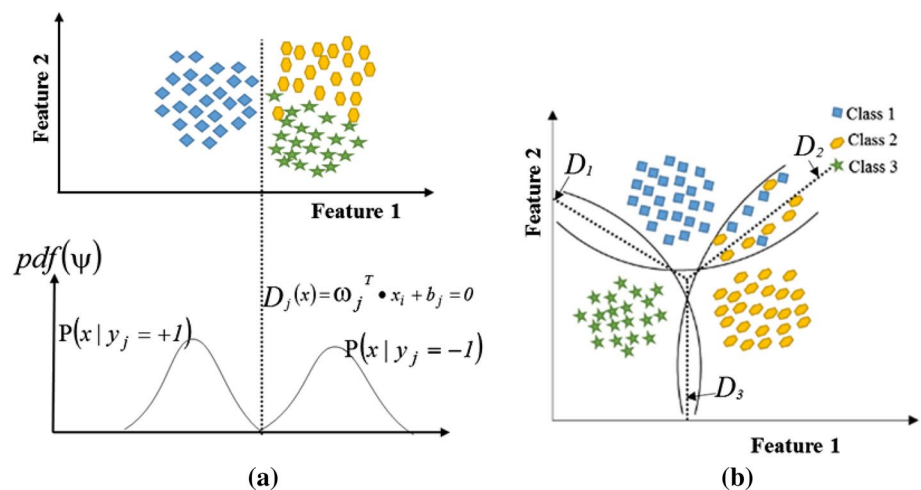
It is worthwhile to mention that our proposed methodology showed an improved performance because of its main conception regarding the utilization of all feature spaces as an appropriate prior and a formulation of MAP to achieve global optimization of class separation. On the other hand, the three traditional methods do not have any treatment regarding class distributions to maximize their classification performance, and these methods solely depend on initial feature distributions, even when they are combined in an arbitrary, one-against-all, fashion, in which the spatial variations among the classes are completely overlooked.

Figure 16 compares the overall performance based on all datasets. As expected, the proposed WPT-EA outperforms other conventional methods since it provides better feature distribution than other methodologies.

Table 3 Sensitivities and classification accuracies for identifying various single and multiple-combined defects and bearing with no defect (BND) using 20 times k-cv (in %)

	BCO	BCI	BCR	BCIO	BCOR	BCIR	BCIOR	BND	ACA
Average sensitivity (standard deviation) of each fault class									
Dataset 1 (300 rpm with 3 mm crack)									
Methodology 1	75 (9)	78.66 (6.3)	74.33 (5.7)	76.68 (10)	78.5 (11)	79.67 (11)	75.01 (6.9)	83.33 (5.2)	77.65 (8.1)
Methodology 2	79.38 (8.6)	80.5 (9)	76.67 (9.5)	84.33 (9.8)	78.01 (7.6)	89.33 (6.7)	86.33 (10.9)	86.39 (6.6)	82.62 (8.6)
Methodology 3	80.33 (2.4)	88 (3.0)	81.77 (2.1)	82.33 (4.5)	84.6 (2.5)	86.33 (2.0)	87.33 (2.2)	80.33 (2.0)	83.88 (2.6)
Proposed	91.57 (1.2)	89.5 (3.0)	91.17 (1.5)	86.17 (2.2)	89.5 (1.7)	90.37 (1.6)	89.97 (1.8)	92.85 (1.5)	90.14 (1.8)
Dataset 2 (500 rpm with 3 mm crack)									
Methodology 1	78.33 (7.8)	73.66 (5.5)	73.33 (7.5)	78.33 (6.9)	78.33 (7.2)	75.68 (9.9)	78.33 (4.6)	82.78 (9)	77.35 (7.3)
Methodology 2	79.78 (7.9)	78.11 (8)	79.78 (6.3)	82.19 (7.7)	82.71 (7.6)	83.11 (8.4)	86.78 (10.9)	88.1 (6.3)	82.57 (7.9)
Methodology 3	81.42 (1.3)	79.77 (2.0)	81.42 (2.1)	83.84 (2.3)	82.94 (1.6)	84.75 (2.4)	82.42 (1.5)	90.75 (1.9)	83.41 (1.9)
Proposed	94.25 (1.0)	92.33 (1.9)	94.25 (1.6)	94.01 (1.8)	94 (1.1)	91.83 (1.3)	94.7 (1.5)	94.87 (0.7)	93.78 (1.4)
Dataset 3 (300 rpm with 12 mm crack)									
Methodology 1	82.43 (2.1)	81.67 (5.5)	81.67 (4.4)	84.63 (2.7)	85.99 (6.5)	85.05 (3.8)	83.88 (3.2)	90.99 (7.5)	84.54 (4.5)
Methodology 2	83 (3.9)	82.25 (3.0)	83.93 (2.2)	85.31 (2.9)	86.12 (5.0)	85.05 (5.4)	85 (5.8)	94.13 (6.0)	85.60 (4.3)
Methodology 3	85.35 (1.4)	83.6 (0.9)	86.29 (1.1)	87.67 (0.9)	95.48 (1.3)	87.4 (1.9)	87.35 (1.0)	95.49 (3.5)	88.58 (1.5)
Proposed	94.15 (0.4)	93.9 (0.3)	94.98 (0.5)	96.15 (0.9)	98.29 (0.9)	94.7 (1.7)	96.65 (0.2)	100 (0.0)	96.10 (0.6)
Dataset 4 (500 rpm with 12 mm crack)									
Methodology 1	86.68 (4.5)	85.92 (4)	85.92 (3.6)	88.88 (7.4)	90.24 (3.5)	89.92 (4.1)	92.13 (4.2)	95.24 (5)	89.37 (4.5)
Methodology 2	90.31 (5.5)	88.06 (7)	89.75 (9.5)	91.12 (6.3)	95.01 (6.2)	90.86 (3.2)	95.81 (4.9)	97.95 (4.9)	92.36 (5.9)
Methodology 3	90.34 (1.6)	88.59 (1.5)	91.27 (1.1)	92.65 (1.6)	95.5 (0.8)	92.39 (1.3)	95.34 (1.5)	98.47 (0.6)	93.07 (1.3)
Proposed	96.2 (0.6)	95.96 (0.6)	97.04 (0.5)	98.2 (1.4)	100 (0.0)	96.75 (1.5)	100 (0.0)	100 (0.0)	98.02 (0.57)

Fig. 9 **a** The basic concept of feature space utilization of a traditional-OAASVM with an improved probabilistic classifier, **b** well separated decision boundaries by the probabilistic classifier, as indicated by dotted straight lines



5 Conclusion

This paper presented a highly reliable multi-fault diagnosis methodology for identifying single and multiple-combined

faults of low-speed bearings with varied rotational speeds and crack sizes. This study mainly focused on two major contributions, namely WPT-EA with DDR for finding informative sub-bands for discriminative feature extraction

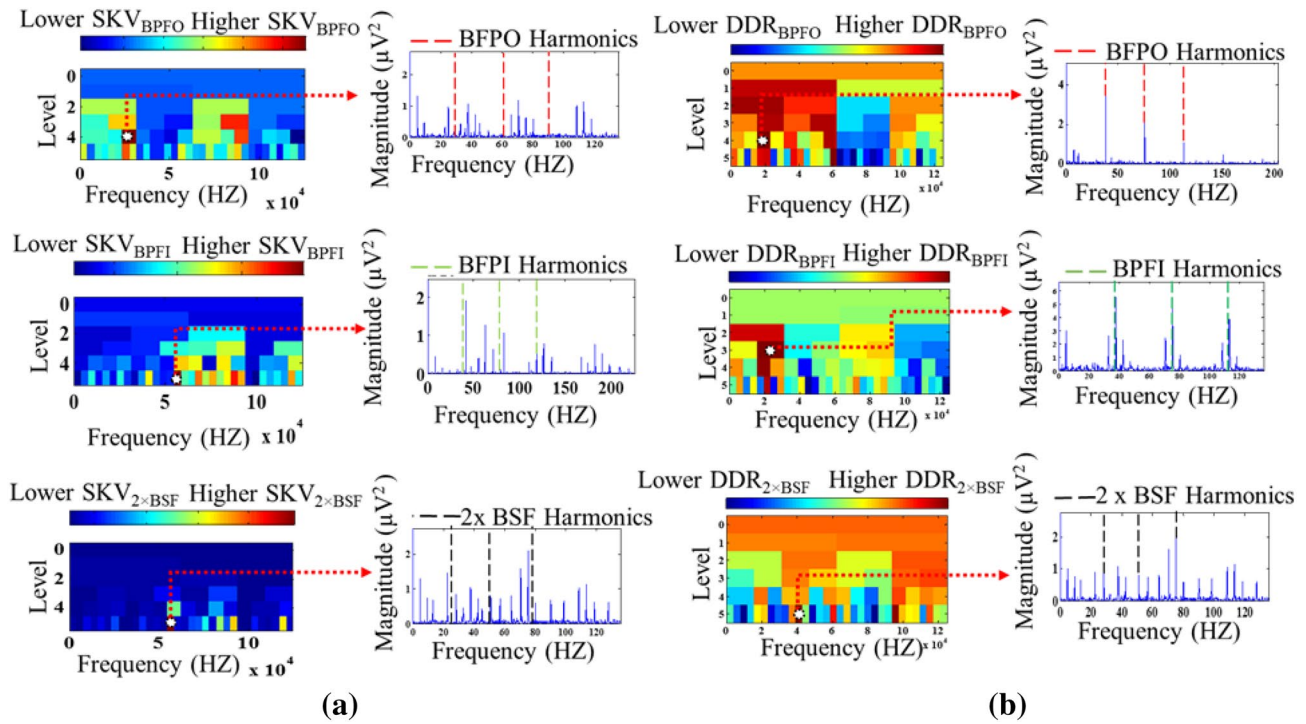
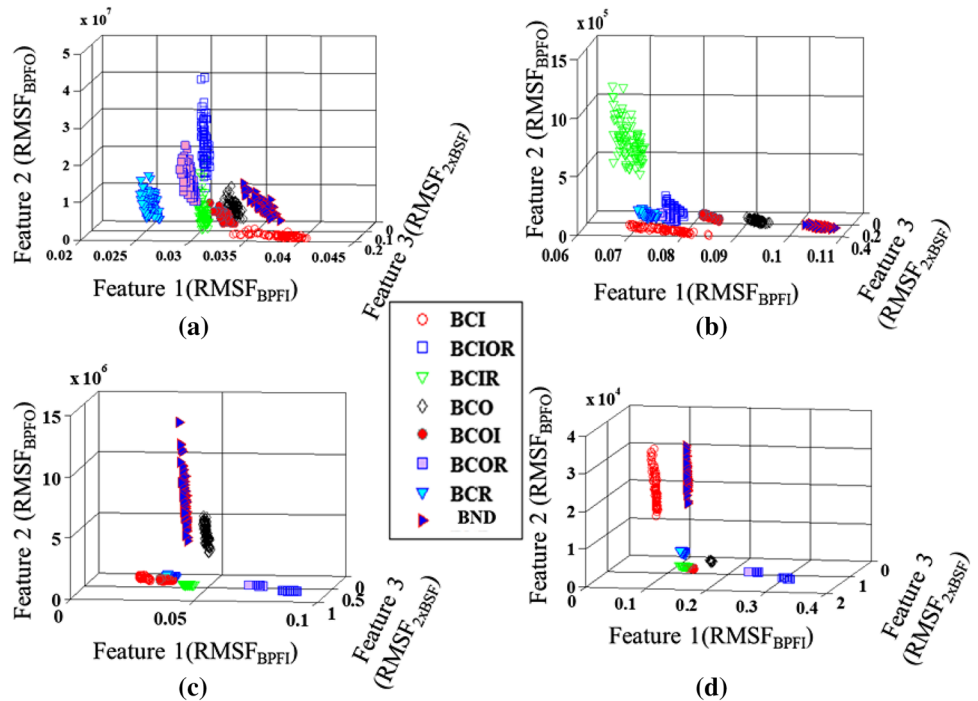


Fig. 10 2-D visualization tool for finding informative sub-band signals based on **a** SKV and **b** the proposed DDR values

Fig. 11 3-D visualization of three DDR values of each defect frequency from feature pool using **a** Dataset 1, **b** Dataset 2, **c** Dataset 3, and **d** Dataset 4, for all single and multiple-combined faults (i.e. 8 types in this study)



and a probabilistic classifier (probabilistic-OAASVM) for improved diagnostic results. This probabilistic classifier improves the traditional-OAASVM by introducing a new

feature space utilization scheme as a Gaussian process prior and maximized the classification performance using Bayesian inference. Overall, the probabilistic-OAASVM

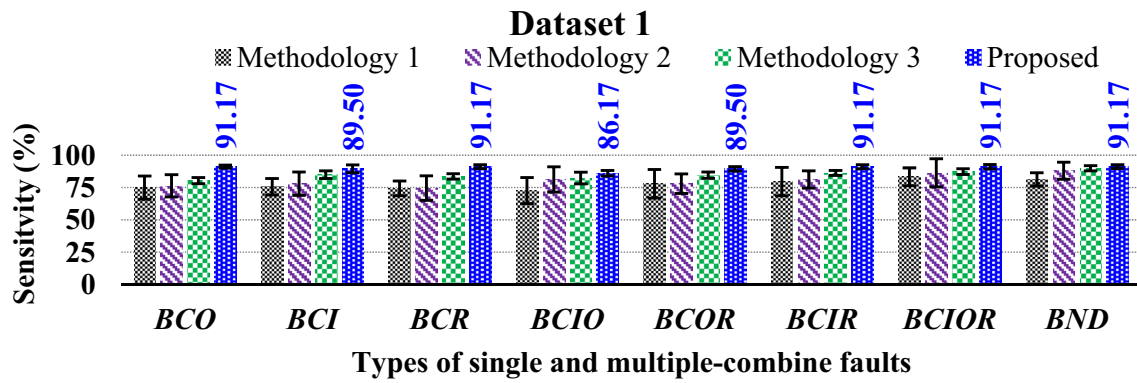


Fig. 12 The average sensitivity of each fault class with standard deviation for dataset 1

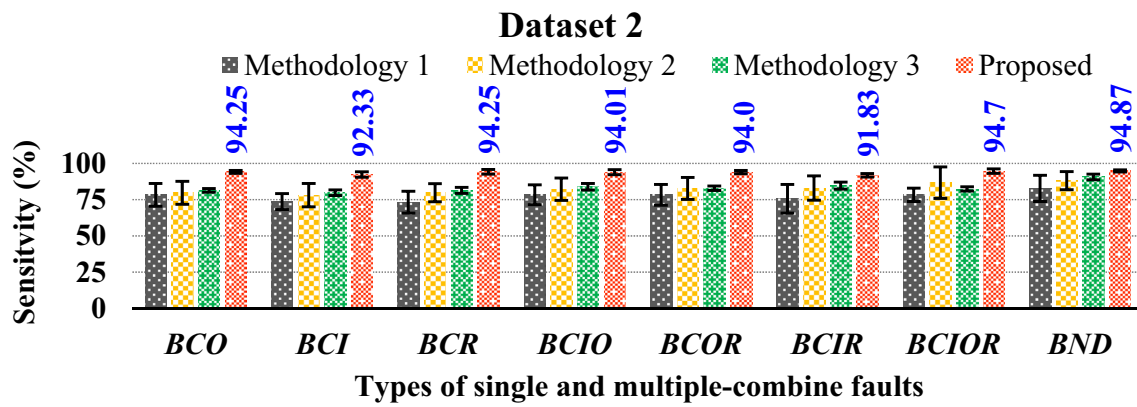


Fig. 13 The average sensitivity of each fault class with standard deviation for dataset 2

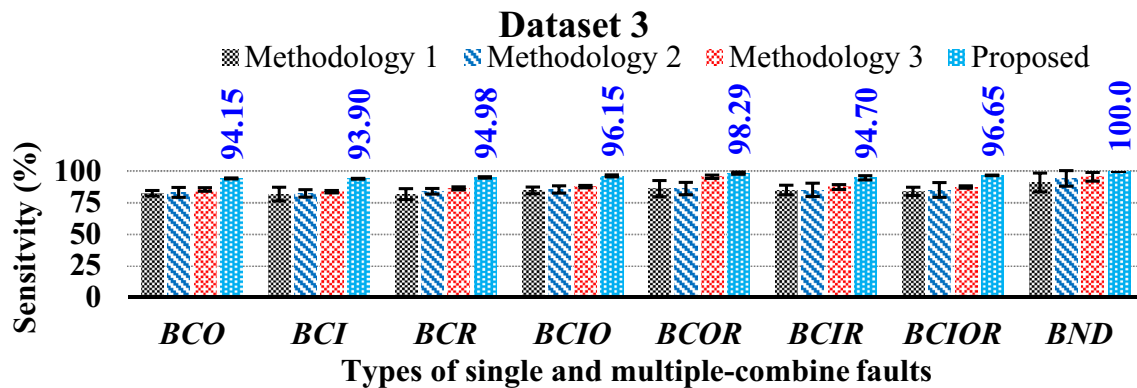


Fig. 14 The average sensitivity of each fault class with standard deviation for dataset 3

method provided superior diagnostic performance in all aspects of the experiments. It especially showed an increasing trend in the diagnostic performance when the rotational speeds and crack sizes are increased. In addition

to validating the effectiveness, the proposed classifier outperformed three state-of-art algorithms, yielding a 4.95–20.67% diagnostic performance improvement in the average classification accuracy.

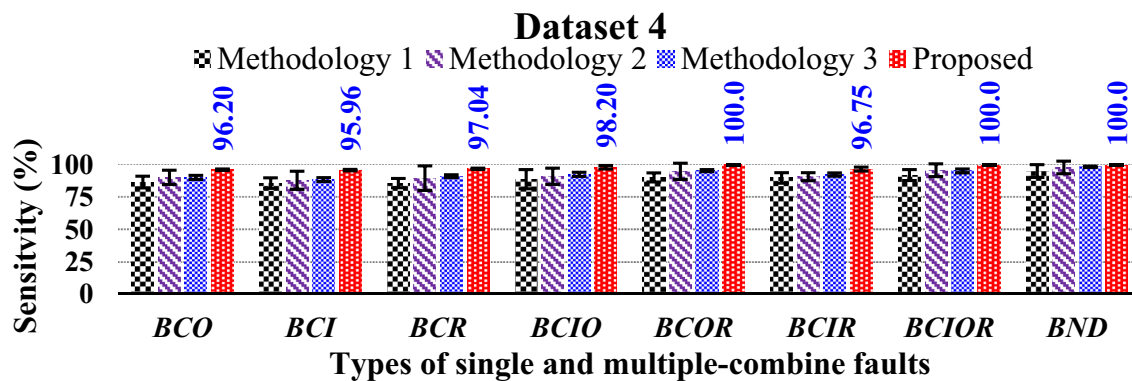


Fig. 15 The average sensitivity of each fault class with standard deviation for dataset 4

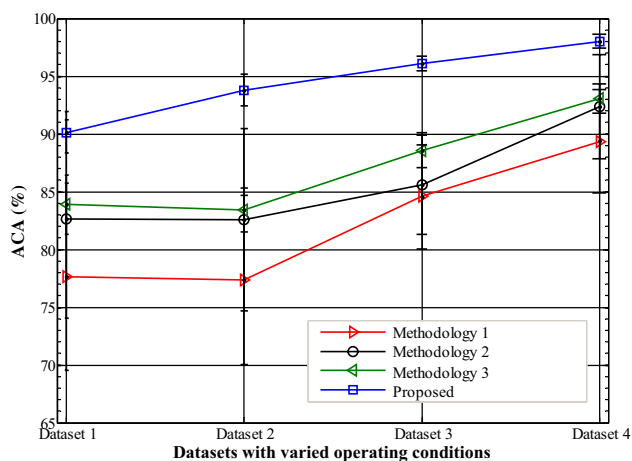


Fig. 16 Effects of crack sizes and rotational speeds (varied from lower to higher) on diagnosis performance

Acknowledgements This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. 20162220100050, No. 20161120100350, and No. 20172510102130). It was also funded in part by The Leading Human Resource Training Program of Regional Neo industry through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and future Planning (NRF-2016H1D5A1910564), and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2016R1D1A3B03931927).

References

- Abe S (2015) Fuzzy support vector machines for multilabel classification. *Pattern Recognit* 48:2110–2117. doi:[10.1016/j.patcog.2015.01.009](https://doi.org/10.1016/j.patcog.2015.01.009)
- Aydin I, Karakose M, Akin E (2011) A multi-objective artificial immune algorithm for parameter optimization in support vector machine. *Appl Soft Comput* 11:120–129. doi:[10.1016/j.asoc.2009.11.003](https://doi.org/10.1016/j.asoc.2009.11.003)
- Chakrabarty S, Cauwenberghs G (2007) Support vector machine: quadratic entropy based robust multi-class probability regression. *J Mach Learn Res* 8:813–839
- Chen X, Zhou J, Xiao J, Zhang X, Xiao H, Zhu W, Fu W (2014) Fault diagnosis based on dependent feature vector and probability neural network for rolling element bearings. *Appl Math Comput* 247:835–847. doi:[10.1016/j.amc.2014.09.062](https://doi.org/10.1016/j.amc.2014.09.062)
- Chih-Wei H, Chih-Jen L (2002) A comparison of methods for multiclass support vector machines. *IEEE Trans Neural Networks* 13:415–425. doi:[10.1109/72.991427](https://doi.org/10.1109/72.991427)
- de Bessa IV, Palhares RM, D'Angelo MFSV, Chaves Filho JE (2016) Data-driven fault detection and isolation scheme for a wind turbine benchmark. *Renew Energy* 87(Part 1):634–645. doi:[10.1016/j.renene.2015.10.061](https://doi.org/10.1016/j.renene.2015.10.061)
- Dong G, Chen J, Zhao F (2015) A frequency-shifted bispectrum for rolling element bearing diagnosis. *J Sound Vib*, 339:396–418. doi:[10.1016/j.jsv.2014.11.015](https://doi.org/10.1016/j.jsv.2014.11.015)
- Finogeev AG, Parygin DS, Finogeev AA (2017) The convergence computing model for big sensor data mining and knowledge discovery. *Hum Cent Comput Inf Sci* 7:11. doi:[10.1186/s13673-017-0092-7](https://doi.org/10.1186/s13673-017-0092-7)
- Hamadache M, Lee D, Veluvolu KC (2015) Rotor speed-based bearing fault diagnosis (RSB-BFD) under variable speed and constant load. *IEEE Trans Ind Electron* 62:6486–6495. doi:[10.1109/TIE.2015.2416673](https://doi.org/10.1109/TIE.2015.2416673)
- He Y, Zhang X (2012) Approximate entropy analysis of the acoustic emission from defects in rolling element bearings. *J Vib Acoust* 134:061012–061012. doi:[10.1115/1.4007240](https://doi.org/10.1115/1.4007240)
- Hyun-Chul K, Ghahramani Z (2006) Bayesian Gaussian process classification with the EM-EP algorithm. *IEEE Trans Pattern Anal Mach Intell* 28:1948–1959. doi:[10.1109/TPAMI.2006.238](https://doi.org/10.1109/TPAMI.2006.238)
- Islam MMM, Khan SA, Kim J-M (2015) Multi-fault diagnosis of roller bearings using support vector machines with an improved decision strategy. In: Huang D-S, Han K (eds) *Advanced intelligent computing theories and applications: 11th international conference, ICIC 2015, Fuzhou, China, August 20–23, 2015. Proceedings, Part III*. Springer International Publishing, Cham, pp 538–550. doi:[10.1007/978-3-319-22053-6_57](https://doi.org/10.1007/978-3-319-22053-6_57)
- Islam MMM, Khan SA, Kim J-M (2017) Reliable bearing fault diagnosis using Bayesian inference-based multi-class support vector machines. *J Acoust Soc Am* 141:EL89–EL95. doi:[10.1121/1.4976038](https://doi.org/10.1121/1.4976038)
- Jeong I-K, Kang M, Kim J, Kim J-M, Ha J-M, Choi B-K (2015) Enhanced DET-based fault signature analysis for reliable diagnosis of single and multiple-combined bearing defects. *Shock Vib* 2015:10. doi:[10.1155/2015/814650](https://doi.org/10.1155/2015/814650)
- Kandukuri ST, Klausen A, Karimi HR, Robbersmyr KG (2016) A review of diagnostics and prognostics of low-speed machinery

- towards wind turbine farm-level health management. *Renew Sustain Energy Rev* 53:697–708. doi:[10.1016/j.rser.2015.08.061](https://doi.org/10.1016/j.rser.2015.08.061)
- Kang M, Kim J, Choi B-K, Kim J-M (2015a) Envelope analysis with a genetic algorithm-based adaptive filter bank for bearing fault detection. *J Acoust Soc Am* 138:EL65–EL70. doi:[10.1121/1.4922767](https://doi.org/10.1121/1.4922767)
- Kang M, Kim J, Wills LM, Kim JM (2015b) Time-varying and multiresolution envelope analysis and discriminative feature analysis for bearing fault diagnosis. *IEEE Trans Ind Electron* 62:7749–7761. doi:[10.1109/TIE.2015.2460242](https://doi.org/10.1109/TIE.2015.2460242)
- Kang M, Islam MR, Kim J, Kim JM, Pecht M (2016) A hybrid feature selection scheme for reducing diagnostic performance deterioration caused by outliers in data-driven diagnostics. *IEEE Trans Ind Electron* 63:3299–3310. doi:[10.1109/TIE.2016.2527623](https://doi.org/10.1109/TIE.2016.2527623)
- Khoobjou E, Mazinan AH (2017) On hybrid intelligence-based control approach with its application to flexible robot system. *Hum Cent Comput Inf Sci* 7:5. doi:[10.1186/s13673-017-0086-5](https://doi.org/10.1186/s13673-017-0086-5)
- Lalani S, Doye D (2017) Discrete wavelet transform and a singular value decomposition technique for watermarking based on an adaptive fuzzy inference system. *J Inf Process Syst* 13:340–347
- Li R, He D (2012) Rotational machine health monitoring and fault detection using EMD-based acoustic emission feature quantification. *IEEE Trans Instrum Meas* 61:990–1001. doi:[10.1109/TIM.2011.2179819](https://doi.org/10.1109/TIM.2011.2179819)
- Murphy KP (2012) *Machine learning: a probabilistic perspective*. MIT, Cambridge, pp 290–295
- Nasiri JA, Moghadam Charkari N, Jalili S (2015) Least squares twin multi-class classification support vector machine. *Pattern Recognit*, 48:984–992. doi:[10.1016/j.patcog.2014.09.020](https://doi.org/10.1016/j.patcog.2014.09.020)
- Rasmussen CE, Williams C (2006) *Relationships between GPs and other models*. MIT, Cambridge. <http://www.gaussianprocess.org/gpml/chapters/RW6.pdf>
- Skolidis G, Sanguinetti G (2011) Bayesian multitask classification with Gaussian process priors. *IEEE Trans Neural Netw* 22:2011–2021. doi:[10.1109/TNN.2011.2168568](https://doi.org/10.1109/TNN.2011.2168568)
- Smola AJ, Schölkopf B, Müller K-R (1998) The connection between regularization operators and support vector kernels. *Neural Netw* 11:637–649. doi:[10.1016/S0893-6080\(98\)00032-X](https://doi.org/10.1016/S0893-6080(98)00032-X)
- Sollich P (2000) Probabilistic methods for support vector machines. In: *Advances in neural information processing systems*, MIT Press, pp 349–355
- Wang D, Tse PW, Tsui KL (2013) An enhanced Kurtogram method for fault diagnosis of rolling element bearings. *Mech Syst Sig Process* 35:176–199. doi:[10.1016/j.ymssp.2012.10.003](https://doi.org/10.1016/j.ymssp.2012.10.003)
- Yala N, Fergani B, Fleury A (2017) Towards improving feature extraction and classification for activity recognition on streaming data. *J Ambient Intell Hum Comput* 8:177–189. doi:[10.1007/s12652-016-0412-1](https://doi.org/10.1007/s12652-016-0412-1)