

Bearing fault diagnosis using multiclass support vector machines with binary particle swarm optimization and regularized Fisher's criterion

Ridha Ziani · Ahmed Felkaoui ·
Rabah Zegadi

Received: 28 March 2014 / Accepted: 21 October 2014 / Published online: 1 November 2014
© Springer Science+Business Media New York 2014

Abstract Condition monitoring of rotating machinery has attracted more and more attention in recent years in order to reduce the unnecessary breakdowns of components such as bearings and gears which suffer frequently from failures. Vibration based approaches are the most commonly used techniques to the condition monitoring tasks. In this paper, we propose a bearing fault detection scheme based on support vector machine as a classification method and binary particle swarm optimization algorithm (BPSO) based on maximal class separability as a feature selection method. In order to maximize the class separability, regularized Fisher's criterion is used as a fitness function in the proposed BPSO algorithm. This approach was evaluated using vibration data of bearing in healthy and faulty conditions. The experimental results demonstrate the effectiveness of the proposed method.

Keywords Support vector machines (SVMs) · Particle swarm optimization (PSO) · Regularized linear discriminant analysis (RLDA) · Features selection · Condition monitoring

Introduction

The importance of fault diagnosis of rotating machinery in manufacturing industry is increasing due to the demand for machines availability. However, the traditional engineering approaches require a significant degree of engineer-

ing expertise to apply them successfully. Therefore, simpler approaches are needed to allow relatively unskilled operators to make reliable decisions without the need of a specialist to examine the data and to diagnose the problems. Hence, there is a demand to incorporate techniques that can make decision on the health of the machine automatically and reliably (Yang et al. 2005). The vibration analysis, for machine condition monitoring, has proven to be an appreciated tool for a few decades for industries (Jack and Nandi 2002; Samanta et al. 2001; Wang and Too 2002; Rafiee et al. 2007; Kurek and Osowski 2010; Konar and Chattopadhyay 2011). Its use is articulated around three levels of analysis: the monitoring, the diagnosis, and the follow-up of the equipments health state. Fault diagnosis can be carried out by learning from known problems, such as unbalance, shaft misalignment, gears and bearing defects. Generally, it includes three crucial steps: feature extraction, sensitive feature selection, and fault patterns recognition.

The popularly diagnosis methods used in machine condition monitoring which are based on Artificial Intelligence (AI) belongs to two broad categories: supervised and unsupervised learning. If the classes of the observations in the data set used to train the model are known, then it is a supervised learning approach, otherwise it is an unsupervised learning approach (Mortada et al. 2013).

In (Gryllias and Antoniadis 2012) was reported that unsupervised learning procedures present some inherent disadvantages over supervised learning. The data clusters result in an unsupervised way cannot be easily attributed to specific faults, they require a-posteriori intervention of experienced personal. Moreover, most existing unsupervised learning methods still present stability, convergence and robustness problems.

In supervised learning methods, the most well known Artificial Neural Networks (ANN) have been extensively used in

R. Ziani (✉) · A. Felkaoui · R. Zegadi
Laboratory of Applied Precision Mechanics, Institute of Optics and Precision Mechanics, Ferhat Abbes University Setif 1, 19000 Sétif, Algeria
e-mail: ziani_lmpa@yahoo.fr

R. Ziani
National High School of Technology, ENST ex CT siege DG SNVI, RN No 5 Z.I Rouiba, Algiers, Algeria

fault diagnosis (Samanta et al. 2001; Jack and Nandi 2002; Samanta et al. 2003; and Rafiee et al. 2007). Expert system was applied in (Yang et al. 2012) for fault diagnosis of gear box in wind turbine. Another application of this method was presented in (Qian et al. 2008). In (Li et al. 2013b) a Fuzzy k-Nearest Neighbor (FKNN) classifier was proposed to the fault pattern identification of a gearbox, and two case studies were carried out to evaluate the effectiveness of the proposed diagnostic approach; one is for the gear fault diagnosis, and the other is to diagnose the rolling bearing faults of the gearbox.

Support vector machines (SVMs) introduced by Vapnik (1995) is a relatively new computational supervised learning method based on Statistical Learning Theory. Unlike to the above classification methods, SVM has a global optimum and exhibits better prediction accuracy due to its implementation of the Structural Risk Minimization (SRM) principle which considers both the training error and the capacity of the classifier model. Moreover, SVM does not require a large number of training samples (Borges 1998) and can solve the learning problem even when only a small amount of training samples is available (Gryllias and Antoniadis 2012). Due to the fact that it is hard to obtain sufficient fault samples in practice, SVMs have been already proposed, for numerous practical applications in rotating machine health condition monitoring (Samanta et al. 2001; Yang et al. 2007; Kurek and Osowski 2010; Konar and Chattopadhyay 2011). For all the above reasons SVM is considered in this paper.

One of the most important and indispensable tasks in any pattern recognition system is the use of feature selection methods to overcome the curse of dimensionality problem. Kudo and Sklansky (2000) indicate that the reasons behind using features selection methods are: (1) to reduce the cost of extracting features, (2) to improve the classification accuracy, and (3) to improve the reliability of the estimated performance. Usually two main approaches for features selection: wrapper methods, in which the features are selected using the classifier, and filter methods, in which the selection of features is independent of the classifier used. In the past few years, the choice of an algorithm for selecting features from an initial set was the focus of a great deal of research, and a large number of algorithms have been proposed such as Principal Components Analysis (PCA) (Sun et al. 2006), Kernel PCA (Zhang et al. 2013a), Independent Components Analysis (ICA) (He et al. 2013), Differential Evolution (DE) (Khushaba et al. 2011), and Simulated Annealing (SA) (Lin et al. 2008). In addition to the above different features selection methods, population based search procedures like: Ant Colony Optimization (ACO), Genetic Algorithms (GA), and Particle swarm optimization (PSO) were the focus of a great deal of research in the past few years (Chen et al. 2010; Jack and Nandi 2002; Yuan and Chu 2007). Some compar-

ative studies of features selection methods have been carried out in (Kudo and Sklansky 2000) and (Khushaba et al. 2011).

Among the population based approaches, the application of PSO to features selection has attracted a lot of attention. Samanta and Nataraj (2009) presented a study on the application of PSO combined with Artificial Neural Networks (ANNs) and SVMs for bearing fault detection. In this study PSO was even used to optimize classifier parameters such as number of nodes in the hidden layer for ANNs and kernel parameters for SVMs. Yuan and Chu (2007) proposed a new method that jointly optimizes the features selection and the SVM parameters with a modified discrete particle swarm optimization. Li et al. (2007) presented an improved PSO algorithm for training SVMs. The PSO algorithm presented in this paper is combined with Proximal Support Vector Machines (PSVM) for features selection. One of the advantages of PSO method is that the user does not have to specify the desired number of features, as it is embedded in the optimization process. Moreover, unlike GA and other evolutionary algorithms, PSO is easy to implement and does not have many parameters that need to be handled properly to achieve a reasonably good performance (Du et al. 2012; Gaitonde and Karnik 2012).

In the studies mentioned above, the fitness function used in PSO algorithms was chosen according to the classifier performance or a desired number of selected features. In this paper, we present a new Binary particle swarm optimization (BPSO) which selects a feature subset that maximizes class separability and consequently increases the classification performance. In order to maximize the class separability, Regularized Fisher's Criterion (RFC) (Friedman 1989) is chosen as a fitness function in the proposed BPSO algorithm. Another reason behind the choice of this features selection scheme is the formulation of SVM method which is based on maximizing the margin between two different classes. In addition, in real applications, the classification accuracy is widely penalized by the overlap between the classes, especially in the multiclass case where the classifier is trained with samples of different known levels of defects. On the other hand, the object of the proposed fault diagnosis scheme is not only limited to identify the presence of damage but also to quantify its extent.

The rest of the paper is organized as follows: in the next section, basic principle of SVMs, PSO, and Fisher's Linear Discriminant Analysis (LDA) is presented. Third section describes the proposed hybrid BPSO-RFC+SVM fault diagnosis scheme. The vibration data and features extraction procedure is given in "Experimental application" section. Fifth section presents the performance evaluation of the proposed fault diagnosis scheme. Finally, sixth section is dedicated to the conclusion.

Basic principles

Support vector machines (SVMs)

SVM is a computational learning method proposed by Vapnik (1995). The essential idea of SVM is to place a linear boundary between two classes of data, and adjust it in such a way that the margin is maximized, namely, the distance between the boundary and the nearest data point in each class is maximal. The nearest data points are used to define the margins and are known as support vectors (SVs) (Samanta et al. 2003; Konar and Chattopadhyay 2011). Once the support vectors are selected, all the necessary information to define the classifier are provided.

If the training data are non-separable (i.e. they are not linearly separable) in the input space, it is possible to create a hyperplane that allows linear separation in the higher dimension. This is achieved through the use of a transformation that converts the data from an N -dimensional input space to Q -dimensional feature space. A kernel can be used to perform this transformation and the dot product in a single step provided the transformation can be replaced by an equivalent kernel function. Among the kernel functions in common use are linear functions, polynomials functions, radial basis functions (RBF), and sigmoid functions. A deeper mathematical treatise of SVMs can be found in the book of Vapnik (1995) and the tutorials on SVMs (Burges 1998; Scholkopf 1998).

As mentioned before, SVM classification is essentially a two-class classification technique, which has to be modified to handle the multiclass tasks in real applications e.g. rotating machinery which usually suffer from more than two faults. Two of the common methods to enable this adaptation include the one-against-all (OAA) and One-against-one (OAO) strategies (Yang et al. 2005).

The one-against-all strategy represents the earliest approach used for SVMs. Each class is trained against the remaining $N - 1$ classes that have been collected together. The “winner-takes-all” rule is used for the final decision, where the winning class is the one corresponding to the SVM with the highest output (discriminant function value). For one classification, N two-class SVMs are needed.

The one-against-one strategy needs to train $N(N - 1)/2$ two-class SVMs, where each one is trained using the data collected from two classes. When testing, for each class, score will be computed by a score function. Then, the unlabeled sample x will be associated with the class with the largest score.

Particle swarm optimization (PSO)

In PSO technique (Kennedy and Eberhart 1995), individuals (particles) are composed of cells called position. The swarm composed from these particles is randomly initialized, and

every particle in the swarm represents a potential solution. The PSO successfully leads to a global optimum by an iterative procedure based on the processes of movement and intelligence in an evolutionary system.

Best values of each particle (personal best value $p_{best,i,j}$, global best value $g_{best,i,j}$) are accumulated to be used in the next step and also to obtain the optimal value. The velocity and the position of the particles at the next iteration ($t + 1$) are calculated in terms of the values at current iteration (t) as follows:

$$v_{k,l}(t + 1) = \omega v_{k,l}(t) + c_1 R_1 (p_{best,k,l} - X_{k,l}(t)) + c_2 R_2 (g_{best,k,l} - X_{k,l}(t)) \tag{1}$$

$$X_{k,l}(t + 1) = X_{k,l}(t) + v_{k,l}(t + 1) \tag{2}$$

where k is the index of particle, l is the index of position in particle, t shows the iteration number, ω is called the “inertia weight” that controls the impact of the previous velocity of the particle on its current one. $v_{k,l}(t)$ is the velocity of the k th particle in swarm on l th index of position in particle $v_{min} \leq v_{k,l}(t) \leq v_{max}$ and $X_{k,l}(t)$ is the position. R_1 and R_2 are the random numbers uniformly distributed in the interval $[0.0, 1.0]$. c_1 and c_2 are positive constants with default values 2, called “acceleration coefficients”.

In the BPSO technique (Kennedy and Eberhart 1997), each particle position is expressed as a binary bit vector composed of 0’s and 1’s. The velocity $v_{k,l}$ is used to compute the probability that the l th bit of the k th particle position $x_{k,l}$ takes a value of 1. This determination of the position is performed using the following formula:

$$X_{k,l}(t + 1) = \begin{cases} 0 & \text{if } rand() \leq s(v_{k,l}(t + 1)) \\ 1 & \text{otherwise} \end{cases} \tag{3}$$

where $rand()$ is the random numbers in the closed interval $[0.0, 1.0]$. $S(\cdot)$ is a sigmoid function used to transform the velocity vector into a probability vector as follows:

$$s(v_{k,l}(t + 1)) = \frac{1}{1 + e^{-v_{k,l}(t+1)}} \tag{4}$$

Fisher’s linear discriminant analysis (LDA)

In the present work, we need to evaluate how separable a set of classes are in a D -dimensional feature space by some criteria such as the one discussed here. Fisher’s Linear Discriminant Analysis (LDA) is a popular linear dimensionality reduction method. LDA is given by a linear transformation matrix W maximizing the so-called Fisher criterion (Duda et al. 2000):

$$J_F(W) = tr \left(\frac{W^T S_b W}{W^T S_w W} \right) \tag{5}$$

where S_b and S_w are the between-class scatter matrix and the within-class scatter matrix, respectively. They have the following expressions:

$$S_b = \sum_{i=1}^c n_i (m_i - m)(m_i - m)^T \quad (6)$$

$$S_w = \sum_{i=1}^c S_i \quad (7)$$

where $S_i = \sum_{x \in D_i} (x - m_i)(x - m_i)^T$ is the within-class scatter matrix of class i . $m = \frac{1}{n} \sum_{i=1}^c n_i m_i$ is the overall mean vector. c is the number of classes, m_i and n_i are the mean vector and number of samples of class i respectively. tr denotes the trace of a square matrix, i.e. the sum of the diagonal elements. W is a transformation matrix given by the eigenvectors of S_b/S_w . Fisher's criterion $J_F(W)$ is a measurement of the separability among all classes.

It is well-known that the applicability of LDA to high-dimensional pattern classification tasks often suffers from the so-called “*small sample size*” (SSS) problem arising from the small number of available training samples compared to the dimensionality of the sample space (Sharma and Paliwal 2012). Several methods have been proposed to overcome the SSS problem. These include LDA based on the generalized singular value decomposition (GSVD) (Howland and Park 2004), uncorrelated linear discriminant analysis (ULDA) (Ye et al. 2004), direct LDA method (DLDA) (Yu and Yang 2001), and Regularized LDA method (RLDA) (Friedman 1989). Some other related methods are reported in (Ye and Xiong 2006) and a comparative study is done by Park and Park (2007).

RLDA is a simple and competitive method. In this method, when S_w is singular or ill-conditioned, a diagonal matrix λI with $\lambda > 0$ is added to S_w . Since S_w is symmetric positive semi definite, $S_w + \lambda I$ is non singular with any $\lambda > 0$. The background theory of this method is well discussed in (Friedman 1989; Park and Park 2007). Following the same notation, and by replacing the regularized matrix S_w in (5), the RFC becomes:

$$J_F(W) = tr \left(\frac{W^T S_b W}{W^T (S_w + \lambda I) W} \right) \quad (8)$$

Therefore, the problem of singularity of the classical LDA is solved, and the RFC can be applied in our feature selection algorithm to measure the class separability.

The proposed BPSO-RFC+SVM based fault diagnosis method

As shown in Fig. 1, the vibration signals are processed for the extraction of different features. Then, the obtained dataset matrix of size $(M \times L)$ is normalized within ± 1 , where M is the number of individuals (signals) and L is the number of features. The main advantage of the normalization is to

avoid higher valued features to suppress the influence of the smaller ones. Another advantage is to make machine learning perform well during the calculation. Kernel values usually depend on the inner products of feature vectors, and as a result large features values might cause numerical problems.

BPSO is used to select the most suitable features that maximize the class separability, and consequently improve the classification performance. The BPSO algorithm starts with a population of particles (swarm) wherein each particle represents a possible solution of the problem of class separability which requires to be maximized. The position X and the velocity v of all particles of the population are initialized randomly and have the same dimensions as the number of features (L) in the dataset considered. The particle position is initialized randomly with 0's and 1's. For example $X = [0 \ 1 \ 1 \ 0 \ 1 \ 0 \ 0 \ 1 \ \dots \ 1]$ is a position vector of a particle where the bit 1 when assigned causes the selection of the corresponding feature in the dataset and bit 0 causes the feature to be discarded. This generates a new feature subset corresponding to the particle under consideration. Hence, for a population of N_p particles, N_p corresponding subsets are generated. The objective of the BPSO algorithm is to find the optimal solution (particle) where its corresponding subset maximizes the class separability.

The fitness value of each particle is evaluated via the RFC according to Eq. (8). The RFC measures the distribution of between-class scatter over the within-class scatter. The particle having a high fitness value indicates that the difference between classes is large since the magnitude of RFC value determines the degree of separation of classes. During the evolutionary process looking for larger value of fitness, the between-class scatter is maximized and at the same time the within-class scatter is minimized. For the fitness computation, the following procedure is executed:

1. Suppose there is total K number of 1's in the position X of the particle under consideration.
2. Generate a new subset from the initial dataset with only K features to which bit 1 has been assigned. The new subset generated is of size $(M \times K)$. Where K represents the number of selected features. $1 \leq K \leq L$.
3. Compute the scatter matrixes S_b and S_w of the subset generated by this particle using Eqs. (6) and (7) respectively.
4. Estimate the transformation matrix W by the eigenvectors of $S_b/(S_w + \lambda I)$, where λ is the regularization parameter ($\lambda > 0$) and I is an identity matrix.
5. When S_b , S_w and W are obtained, then the RFC value (considered as the fitness value of the particle) is calculated according to the Eq. (8)

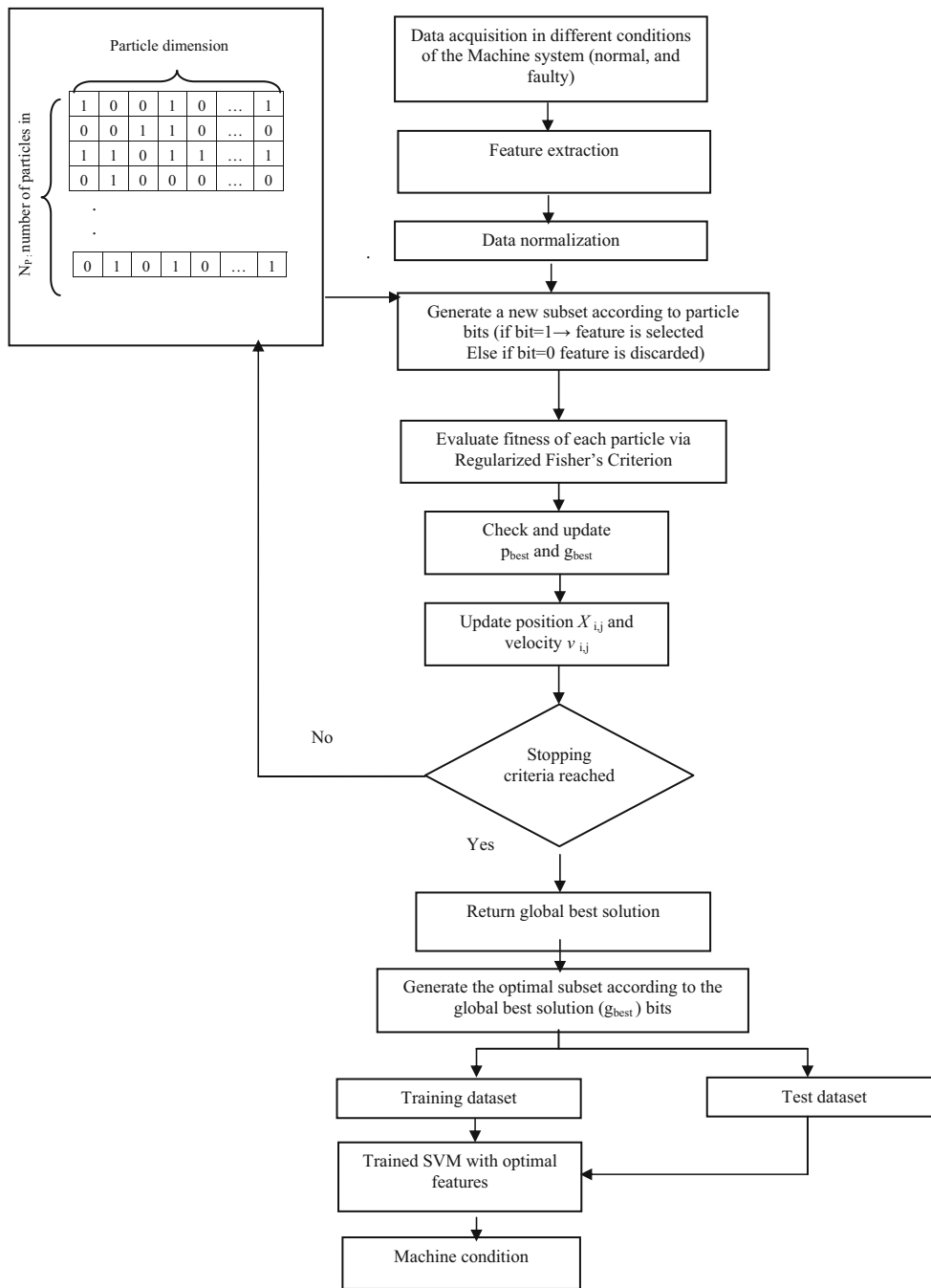


Fig. 1 Flow chart of the proposed BPSO-RFC+SVM based fault diagnosis

At each iteration of the BPSO algorithm, the fitness value of each particle is compared with the fitness value of its previous best personal position (P_{best}). If the current position has the better fitness value it is designated as the new P_{best} of the particle. Then, the current positions of all particles are compared with the previous best global position (g_{best}) of the population in terms of fitness value. If current position of any one of particles is better than the previous g_{best} , then the current position is designated as the new g_{best} .

To generate the next population (Swarm), velocities and positions of each particle are updated according to Eqs. (1) and (3) respectively.

Stopping the algorithm is fixed according to the number of iterations which is initially given. The number of iterations should be sufficient to allow the algorithm to converge to the best solution.

The final best solution (g_{best}) found by the BPSO algorithm is used to generate the optimal subset from the initially

dataset (i.e the subset which allows the best class separability). Then, the M individuals of this subset are divided into two equally parts; the first one is used to train SVM, while the remaining part is used to test the performance of SVM in machine condition prediction.

Experimental application

Vibration data

Vibration data used in this paper have been obtained from the bearing test data set of the Western Reserve University Bearing Data Center website (Loparo 2012). These bearing fault signals have been widely used to validate the effectiveness of new algorithms for bearing fault diagnosis (Gryllias and Antoniadis 2012; Zhang et al. 2013a; Shen et al. 2013). As shown in Fig. 2, the test bed consists of a motor (left), a coupling (center), a dynamometer (right) and control circuits (not shown).

Motor bearings were seeded with faults using electro-discharge machining (EDM). Faults ranging from 0.007 in. in diameter to 0.040 in. in diameter were introduced separately at the inner raceway, rolling element (i.e. ball) and outer raceway. Faulted bearings were reinstalled into the test motor and vibration data was recorded for motor loads (0, 1, 2 and 3 HP), and respective rpm of each load is 1,797, 1,772, 1,750 and 1,730. The bearing monitored is a deep groove ball bearing manufactured by SKF. The drive end bearing is a 6205-2RS JEM with a BPFI, BPFO, and a BSF equalling 5.4152, 3.5848, and 4.7135 times the shaft frequency respectively. The theoretical estimations of the expected BPFO BPFI and BSF frequencies are presented at Table 1. The Vibration data were collected at a sampling rate equal to 12,000 Hz using accelerometers, which were attached to the housing with magnetic bases.

Signal processing and features extraction

The features extraction is very important in vibration based fault diagnosis. Different features and different feature

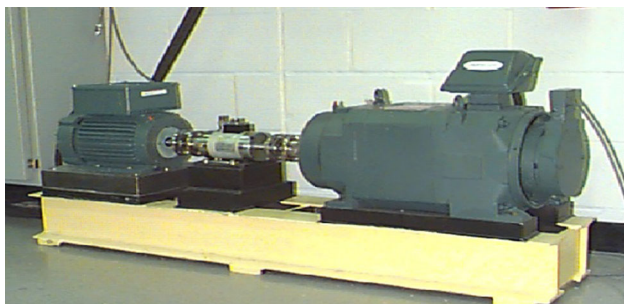


Fig. 2 The test bed

Table 1 Faults characteristic frequencies

Motor load (HP)	Motor speed (rpm)	Fault characteristic frequency (f)		
		BPFI (Hz)	BPFO (Hz)	BSF (Hz)
0	1,797	162.18	107.36	141.16
1	1,772	159.92	105.87	139.20
2	1,750	157.94	104.56	137.47
3	1,730	156.13	103.36	135.90

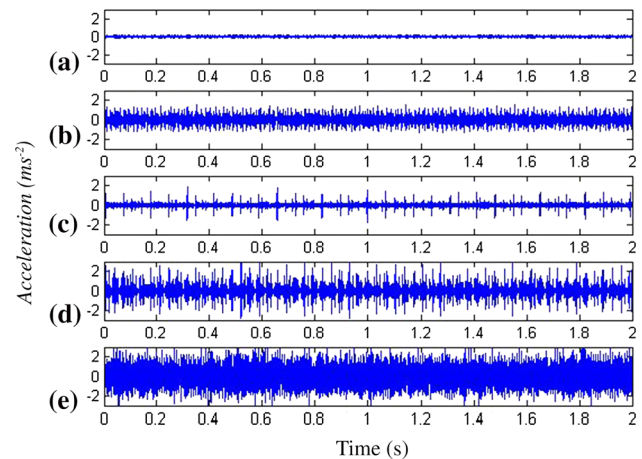


Fig. 3 Time domain signals acquired under 2hp motor load for normal and faulty bearing with inner race fault. **a** Normal, **b** Fault diameter of 0.007 in., **c** Fault diameter of 0.014 in., **d** Fault diameter of 0.021 in., **e** Fault diameter of 0.028 in.

extraction methods have been proposed, including signal statistical analysis in the time domain, low and high-pass filtering, time synchronous averaging (TSA), Empirical Mode Decomposition (EMD), envelope analysis, Fourier transform, cepstral analysis, and wavelet transform. See (Teti et al. 2010) for more details. This section presents a brief discussion of features extraction from time-domain, frequency-domain, and time–frequency domain of vibration signals which will be used in this paper.

In time-domain (Fig. 3), signals are processed to extract the nine following statistical features: mean, crest factor, skewness, kurtosis, and normalised five to nine central statistical moments. Mathematical formula of these features can be found in (Soong 2004).

In spectral domain, the fact that the spectrum of the raw signal often contains little diagnostic information about bearing faults because fault impulses are amplified by structural resonances (Randall 2011), It has been established over the years that the benchmark method for bearing diagnostics is the envelope analysis (Sheen and Liu 2012; Stepanic et al. 2009; Yang et al. 2007; Randall et al. 2001; Li et al. 2012). This is the reason why the envelope analysis method is used in this paper. Usually, an envelope analysis consists of four

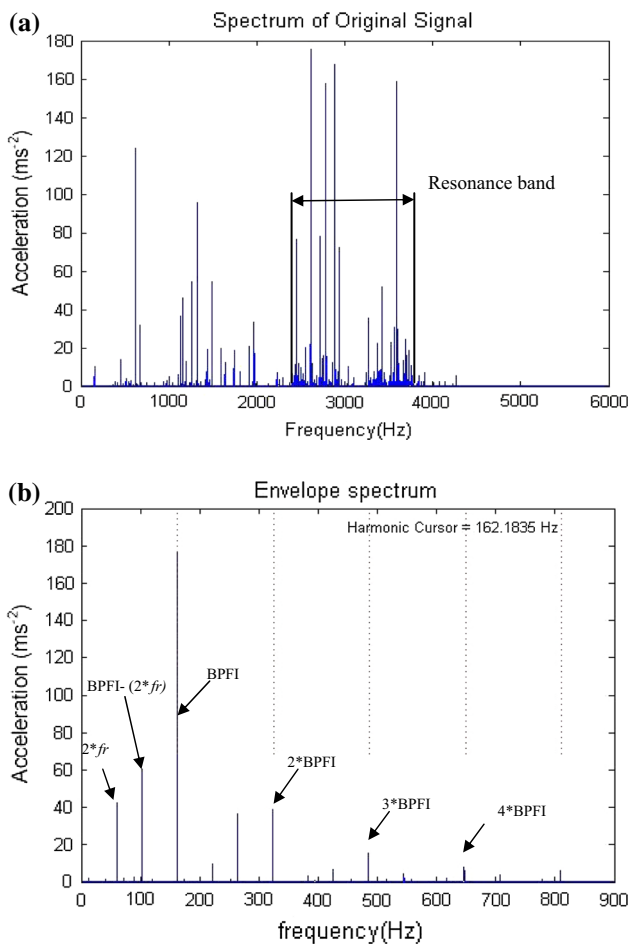


Fig. 4 Spectrum signal of the inner race fault acquired under OHP motor load. **a** spectrum of original signal, **b** envelope spectrum

operations: (a) the resonant frequency band of the structure is determined in the original signal spectrum (Fig. 4a); (b) a band-pass filtering is performed on the original signal in the resonant frequency band, by which most disturbances beyond this band are removed or greatly suppressed, and the weak impulsive components become prominent in the rest components; (c) the envelope signal of the filtered signal is obtained using Hilbert transform (HT); (d) fast Fourier transform (FFT) of the envelope signal is calculated to obtain the envelope spectrum. As shown in Fig. 4b, the fault characteristics frequencies are clearly identified in the envelope spectrum than in the original signal spectrum. In our case, the resonance frequency band was found between 2,400 and 3,800 Hz.

By using this method, the low frequency noise is eliminated so that the characteristic bearing frequencies can be extracted successfully. Afterwards, features extracted from enveloped signal are composed of the sum of Power Spectral Density (PSD) values, calculated at $f \pm \sigma_f$, $2 * f \pm \sigma_f$, $3 * f \pm \sigma_f$, $4 * f \pm \sigma_f$, where f is the average fault characteristic fre-

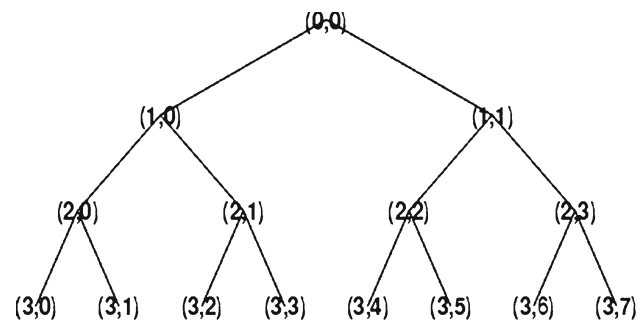


Fig. 5 Wavelet packet Decomposition tree at level 3

quency (BPF0, BPF1, or BSF), and σ_f is the standard deviation of fault frequencies estimated with four motor speeds of Table 1. Hence, a feature set containing 5 features for each sample is obtained, where the fifth one is the sum of PSD values calculated in the total band $[f - \sigma_f, 4 * f + \sigma_f]$.

Taking into account the non-stationary property of the bearing vibration signals, which contains numerous non-stationary or transitory characteristics, Wavelet Packet Decomposition WPD is a suitable tool which has been intensively investigated and applied on nonstationary vibrations signal processing, especially on vibration signal features extraction (Li et al. 2013; Zhang et al. 2013b). Wavelet packet decomposition is developed from wavelet, which shows good performance on both high and low frequency analysis (Mallat 2003). The selection of the mother wavelet can influence the WPT efficiency. Rafiee et al. (2010) have shown that the Daubechies 44 wavelet is the most effective for both faulty gears and bearings. Hence, db44 is adopted in this paper. The signal is firstly decomposed into p wavelet coefficients ($p = 2^q$, and q denotes the wavelet level). In general, the maximum wavelet packet decomposition depth of 3 is effective for features extraction purpose (Shen et al. 2013). By applying three depths WPT decomposition to the original signal with Db44 mother wavelet, the WPT decomposition coefficients are obtained (Fig. 5). In order to obtain further input features for SVM, the kurtosis and energy of the 14 coefficients obtained from all depths are calculated. As result another feature set containing 28 features is obtained.

The procedure of features extraction in time domain, spectral domain, and time-spectral domain is repeated with all vibration signals, and as result a total of 42 features are obtained.

Performance evaluation of the proposed fault diagnosis scheme

In the present section, the ability of the proposed method to detect faults is evaluated with two different cases. First, SVM performance is evaluated using the entire feature set extracted

Table 2 Description of data set considered in fault identification case

Case study	Number of classes	Training samples	Testing samples	Operating condition	Fault size (in.)
Fault identification	4	32	32	Normal	–
				Inner race	0.007
				Outer race	0.007
				Rolling element	0.007

Table 3 Description of data sets considered in 3 cases of fault level identification

Case study	Number of classes	Training samples	Testing samples	Operating condition	Fault size (in.)
Inner race fault level	5	40	40	Normal	–
				Inner race	0.007
				Inner race	0.014
				Inner race	0.021
				Inner race	0.028
Outer race fault level	4	32	32	Normal	–
				Outer race	0.007
				Outer race	0.014
				Outer race	0.021
Rolling element fault level	5	40	40	Normal	–
				Rolling element	0.007
				Rolling element	0.014
				Rolling element	0.021
				Rolling element	0.028

in the above sub section (42 features). In the second, SVM performance is evaluated with only the optimal feature set.

SVM performance with the entire feature set

In real cases of studies when damage appears, the estimation of the bearing's remaining useful life and the machine performance would require not only the process of identifying the presence of damage but also to quantify the extent of damage based on the information extracted from the measured system response. For this reason, the performance of SVM is firstly evaluated in fault identification case (inner race, outer race, or rolling element). Table 2 describe the vibration data set used in this case which is composed of 20 vibration signals and cover a normal condition and the three above faulty conditions of bearing with the smallest fault size (0.007 in.) in each one, which means early detection of the defect. Secondly, after detection and identification of fault, SVM performance is evaluated in fault level identification. In this case, three vibration data sets were used where each one cover a normal condition and all levels of the faulty condition. Table 3 describe the vibration data sets used in these cases of fault level identification.

In order to obtain sufficient samples for all classification cases, each signal was divided into 4 equal samples. Afterwards, the 42 features described in “Signal processing and features extraction” section are extracted from each sample. The procedure of features extraction was repeated with all samples in different cases studies. Hence, we obtain a data base of 64×42 in fault identification case. While in fault level identification we obtain three data bases; 80×42 in inner race case, 64×42 in outer race case, and 80×42 in rolling element case. Then, each data base is partitioned into two equally sized subsets; the first one is used to train SVMs, while the second is used for the test. Data sets were normalised by dividing each column by its absolute maximum value keeping the inputs features within ± 1 for better speed and success of the SVM training.

A large corpus of experiments has been carried out. Table 4 and Table 5 illustrate the classification performance using the entire feature set with two different multiclass SVM strategies; OAO and OAA. Each value indicates the classification accuracy obtained with three different kernels; linear, RBF, and sigmoid. A specific point worth noting is that the penalty parameter “c” and kernel parameters “ σ ” are selected among those which lead to the best classification performance using

Table 4 SVMs performance in fault identification using the entire feature set

Case study	Kernel	SVM test accuracy (%)	
		OAO	OAA
Fault identification	Linear	96.87	93.75
	RBF	100	96.87
	Sigmoid	93.75	90.62

Table 5 SVMs Performance in fault level identification using the entire feature set

Case study	Kernel	SVM test accuracy (%)	
		OAO	OAA
Inner race fault level	Linear	95	92.5
	RBF	97.5	97.5
	Sigmoid	95	95
Outer race fault level	Linear	84.37	81.25
	RBF	96.87	90.62
	Sigmoid	93.75	87.5
Rolling element fault level	Linear	92.5	72.5
	RBF	90	90
	Sigmoid	82.5	77.50

cross validation method where “c” varies in the range [1, 10³] and “σ” varies in the range [10⁻¹, 10]. Results show that the use of different kernels affects significantly the classification performance. Clearly, the best performance for both multi-class SVM strategies is obtained using RBF kernel. Further analysis of these results shows that OAO strategy has higher classification accuracies than OAA in all considered cases. Using RBF kernel and OAO strategy, SVM achieved 100% in fault identification case, while in fault level identification cases it achieved respectively 97.5% in inner race case, 96.87% in outer race case, and 90% in rolling element case.

BPSO-RFC+SVM performances

In order to investigate SVMs classification performance with a sensitive selected feature subset, the proposed BPSO-RFC+SVM is applied on the all cases mentioned in Table 1 and Table 2. BPSO-RFC algorithm was implemented in Matlab and has been initialized with the following parameters values:

- Swarm size = 30 particles.(values recommended by Samanta and Nataraj (2009) between 20 and 50)
- Particle size = 42 (Equal to the number of the extracted features, see “Signal processing and features extraction” section)

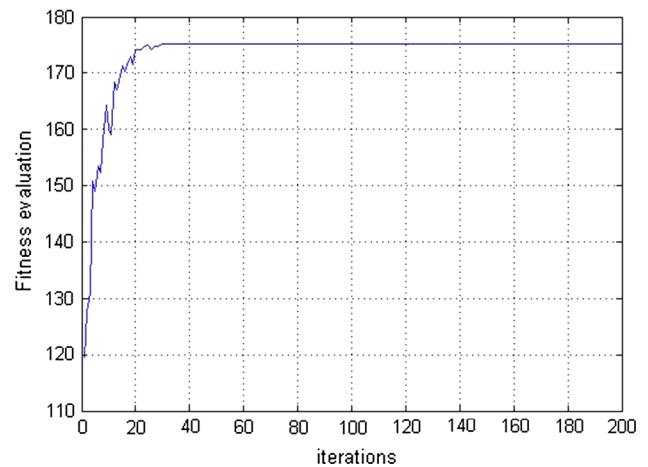


Fig. 6 Convergence of BPSO-RFC algorithm to the best fitness respect to the iterations number

- $\omega_{min} = 0.1, \omega_{max} = 0.6, v_{min} = -2, v_{max} = 2, c_1 = 2, c_2 = 2, R_1$ and R_2 : randomly generated between 0 and 1 (see “Particle Swarm Optimization (PSO)” section).
- Number of iteration $N_i = 200$.

In order to analyze the results, one can start by looking at the convergence of the proposed BPSO-RFC based features selection algorithm. Figure 6 shows that BPSO-RFC algorithm reaches the global best solution after around 30 generations. This can prove that the number of iterations initially given is sufficient. On the other hand, Figs. 7, 8, 9, and 10 present 3D scatter plot of data using PCA which illustrate graphically the influence of the selected feature subset on class separability. It is very clear that in all cases of study, data is well separated with the selected feature subset than using the entire feature set initially extracted.

In order to evaluate how the proposed BPSO-RFC+SVM approach improves the classification performance, SVM is trained with the optimal feature subset, and then the test data set is used to evaluate SVM performance. Table 6 shows the classification performances in fault identification case, while Table 7 shows the performance in fault level identification cases. By comparison of results in Tables 6 and 7 with those of Tables 4 and 5, respectively, it can be seen that BPSO-RFC+SVM has high classification accuracy than SVM with the entire feature set. Sure enough, BPSO-RFC+SVM with RBF kernel achieve 100% in fault identification case with only 21 features, and 100% in all fault level identification cases with 28 features in inner race case, 19 features in outer race case, and only 13 features in rolling element case. This can confirm the efficiency of the proposed BPSO-RFC algorithm in selecting the optimal feature set which maximize class separability and consequently increase the classification accuracy of SVM.

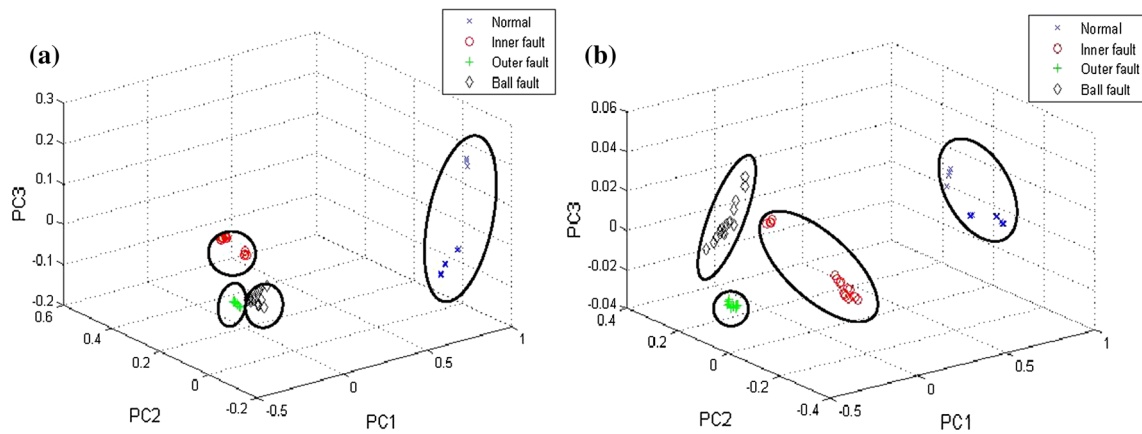


Fig. 7 Scatter plots of data used in fault identification case; **a** using the entire feature set, **b** using the selected feature subset (21 features)

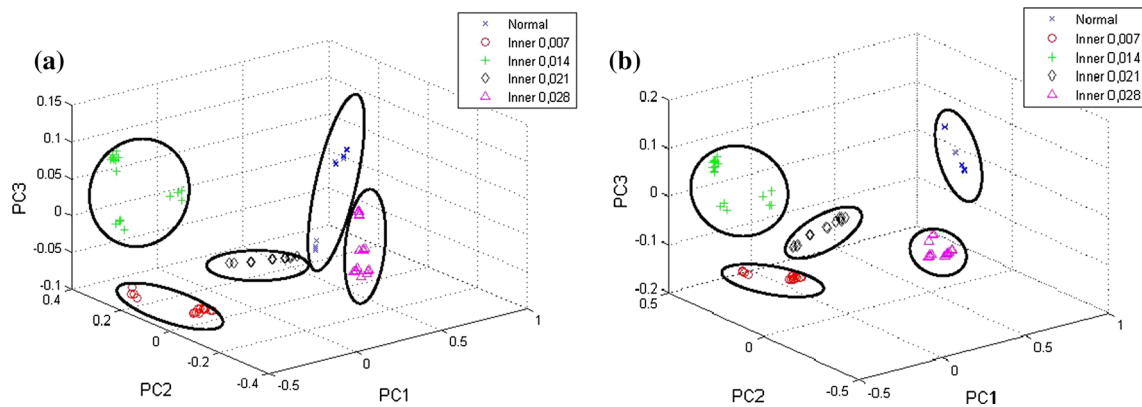


Fig. 8 Scatter plots of data used in inner race fault level identification; **a** using the entire feature set, **b** using the selected feature subset (28 features)

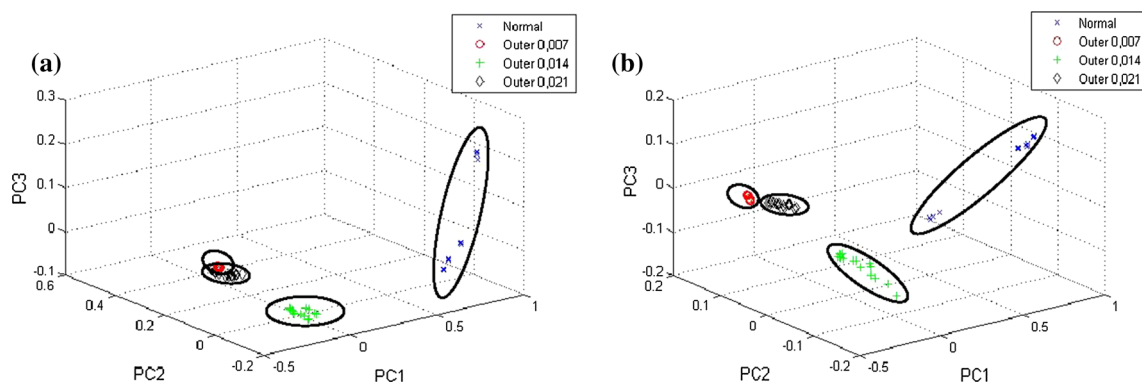


Fig. 9 Scatter plots of data used in outer race fault level identification case; **a** using the entire feature set, **b** using the selected feature subset (19 features)

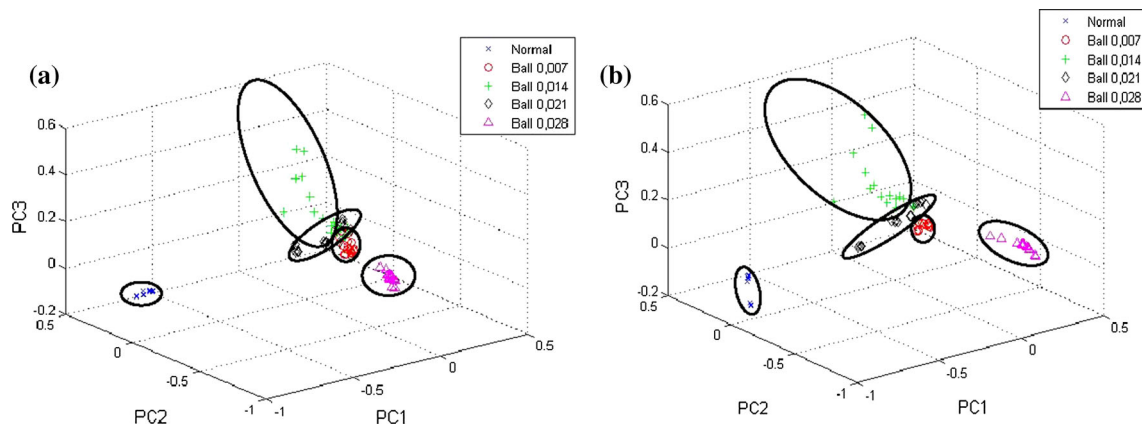


Fig. 10 Scatter plots of data used in rolling element fault level identification case; **a** using the entire feature set, **b** using the selected feature subset (13 features)

Table 6 BPSO-RFC+SVM Performance in fault identification case

Case study	Input features	Kernel	BPSO-RFC+SVM test accuracy (%)	
			OA0	OAA
Fault identification	21	Linear	100	100
		RBF	100	100
		Sigmoid	100	96.87

Table 7 BPSO-RFC+SVM Performance in fault level identification cases

Case study	Input features	Kernel	BPSO-RFC+SVM test accuracy (%)	
			OA0	OAA
Inner race fault level	28	Linear	100	100
		RBF	100	100
		Sigmoid	100	97.5
Outer race fault level	19	Linear	100	96.87
		RBF	100	100
		Sigmoid	100	100
Rolling element fault level	13	Linear	92.5	90
		RBF	100	97.5
		Sigmoid	95	95

Conclusion

In this paper, a BPSO-RFC+SVM algorithm is described. In this approach, the selection of sensitive features is done according to RFC which measures the class separability. This later is used as a fitness function in the proposed BPSO algorithm. Experimental data sets are used to evaluate the performance of the proposed method in fault detection in addition to fault level identification of bearing. Experimental results demonstrate the effectiveness of our method. More-

over, BPSO-RFC has the ability to quickly converge to the best solution. On the other hand, the performance of SVMs has been found to be substantially better with the OA0 strategy and the best accuracy of SVMs was obtained with RBF kernel.

Acknowledgements This work was completed in the laboratory of applied precision mechanics LAPM (University of Setif1, Algeria). The authors would like to thank the Algerian Ministry of Higher Education and Scientific Research (MESRS) and the Delegated Ministry for Scientific Research (MDRS) for granting financial support for CNEPRU

Research Project No. J0301220120001. The authors would like to thank Professor K. A. Loparo of Case Western Reserve University for his kind permission to use their bearing data.

References

- Burges, C. A. (1998). Tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2, 955–974.
- Chen, Y., Miao, D., & Wang, R. (2010). A rough set approach to feature selection based on ant colony optimization. *Pattern Recognition Letters*, 31, 226–233.
- Du, S., Lv, J., & Xi, L. (2012). A robust approach for root causes identification in machining processes using hybrid learning algorithm and engineering knowledge. *Journal of Intelligent Manufacturing*, 23, 1833–1847.
- Duda, R., Hart, P., & Stork, D. (2000). *Pattern classification* (2nd ed.). New York: Wiley.
- Friedman, J. H. (1989). Regularized discriminant analysis. *Journal of the American Statistical Association*, 84, 165–175.
- Gaitonde, V. N., & Karnik, S. R. (2012). Minimizing burr size in drilling using artificial neural network (ANN)-particle swarm optimization (PSO) approach. *Journal of Intelligent Manufacturing*, 23, 1783–1793.
- Gryllias, K. C., & Antoniadis, I. A. (2012). A support vector machine approach based on physical model training for rolling element bearing fault detection in industrial environments. *Engineering Applications of Artificial Intelligence*, 25, 326–344.
- He, Y., Pan, M., Luo, F., Chen, D., & Hu, X. (2013). Support vector machine and optimised feature extraction in integrated eddy current instrument. *Measurement*, 46, 764–774.
- Howland, P., & Park, H. (2004). Generalizing discriminant analysis using the generalized singular value decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(8), 995–1006.
- Jack, L. B., & Nandi, A. K. (2002). Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms. *Mechanical Systems and Signal Processing*, 16, 373–390.
- Kennedy, J., & Eberhart, R. C. (1995). Particle swarm optimization. In *Proceedings of IEEE international conference on neural networks*, Vol. 4, pp. 1942–1948.
- Kennedy, J., & Eberhart, R. C. (1997). A discrete binary version of the particle swarm optimisation algorithm. In *Proceedings of the IEEE International Conference on Neural Networks* (pp. 4104–4108). Australia: Perth.
- Konar, P., & Chattopadhyay, P. (2011). Bearing fault detection of induction motor using wavelet and Support Vector Machines (SVMs). *Applied Soft Computing*, 11, 4203–4211.
- Kudo, M., & Sklansky, J. (2000). Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition*, 33(1), 25–41.
- Kurek, J., & Osowski, S. (2010). Support vector machine for fault diagnosis of the broken rotor bars of squirrel-cage induction motor. *Neural Computing & Application*, 19, 557–564.
- Khushaba, R. N., Al-Ani, A., & Al-Jumaily, A. (2011). Feature subset selection using differential evolution and a statistical repair mechanism. *Expert Systems with Applications*, 38, 11515–11526.
- Li, R., Sopon, P., & He, D. (2012). Fault features extraction for bearing prognostics. *Journal of Intelligent Manufacturing*, 23, 313–321.
- Li, Y., Tong, Y., Bai, B., & Zhang, Y. (2007). An improved particle swarm optimization for SVM training. *Proceedings of the third international conference on natural computation* (pp. 611–615). Los Alamitos: IEEE Computer Society.
- Li, H., Lian, X., Guo, C., & Zhao, P. (2013a). Investigation on early fault classification for rolling element bearing based on the optimal frequency band determination. *Journal of Intelligent Manufacturing*. doi:10.1007/s10845-013-0772-8.
- Li, Z., Yan, X., Tian, Z., Yuan, C., Peng, Z., & Li, L. (2013b). Blind vibration component separation and nonlinear feature extraction applied to the nonstationary vibration signals for the gearbox multi-fault diagnosis. *Measurement*, 46, 259–271.
- Lin, S. W., Lee, Z. J., Chen, S. C., & Tseng, T. Y. (2008). Parameter determination of support vector machine and feature selection using simulated annealing approach. *Applied Soft Computing*, 8, 1505–1512.
- Loparo, K. A. (2012). Bearings Vibration Data Sets, Case Western Reserve University: <http://csegroups.case.edu/bearingdatacenter/home>.
- Mallat, S. G. (2003). *A wavelet tour of signal processing. The sparse way* (3rd ed.). New York: Academic Press.
- Mortada, M. A., Yacout, S., & Lakis, A. (2013). Fault diagnosis in power transformers using multi-class logical analysis of data. *Journal of Intelligent Manufacturing*. doi:10.1007/s10845-013-0750-1.
- Park, C. H., & Park, H. (2007). A comparison of generalized linear discriminant analysis algorithms. *Pattern Recognition*. doi:10.1016/j.patcog.2007.07.022.
- Qian, Y., Xu, L., Li, X., Lin, X., Kraslawski, L., & Lubres, A. (2008). An expert system development and implementation for real-time fault diagnosis of a lubricating oil refining process. *Expert Systems with Applications*, 35(3), 1251–1266.
- Rafiee, J., Arvani, F., Harifi, A., & Sadeghi, M. H. (2007). Intelligent condition monitoring of a gearbox using artificial neural network. *Mechanical Systems and Signal Processing*, 21, 1746–1754.
- Rafiee, J., Rafiee, M. A., & Tse, P. W. (2010). Application of mother wavelet functions for automatic gear and bearing fault diagnosis. *Expert Systems with Applications*, 37, 4568–4579.
- Randall, R. B., Antoni, J., & Chobsaard, S. (2001). The relationship between spectral correlation and envelope analysis in the diagnosis of bearing faults and other cyclostationary machine signals. *Mechanical Systems and Signal Processing*, 15(945–962), 2001.
- Randall, R. B. (2011). *Vibration-based condition monitoring: Industrial, aerospace and automotive applications*. New York: Wiley.
- Samanta, B., Al-Balushi, K. R., & Al-Araimi, S. A. (2001). Use of genetic algorithm and artificial neural network for gear condition diagnostics. *Proceedings of COMADEM*, (pp. 449–456). University of Manchester, UK.
- Samanta, B., Al-Balushi, K. R., & Al-Araimi, S. A. (2003). Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. *Engineering Applications of Artificial Intelligence*, 16, 657–665.
- Samanta, B., & Nataraj, C. (2009). Use of particle swarm optimization for machinery fault detection. *Engineering Applications of Artificial Intelligence*, 22, 308–316.
- Sharma, A., & Paliwal, K. K. (2012). A new perspective to null linear discriminant analysis method and its fast implementation using random matrix multiplication with scatter matrices. *Pattern Recognition*, 45, 2205–2213.
- Scholkopf, B. (1998). SVMs—a practical consequence of learning theory. *IEEE Intelligent Systems*, 13, 18–19.
- Sheen, Y. T., & Liu, Y. H. (2012). A quantified index for bearing vibration analysis based on the resonance modes of mechanical system. *Journal of Intelligent Manufacturing*, 23, 189–203.
- Shen, C., Wang, D., Kong, F., & Tse, P. W. (2013). Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier. *Measurement*, 46, 1551–1564.
- Soong, T. T. (2004). *Fundamentals of probability and statistics for engineers*. New York: Wiley.
- Stepanic, P., Latinovic, I. V., & Djurovic, Z. (2009). A new approach to detection of defects in rolling element bearings based on statistical

- pattern recognition. *International Journal of Advanced Manufacturing Technology*, 45, 91–100.
- Sun, W., Chen, J., & Li, J. (2006). Decision tree and PCA based fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing*, 21, 1300–1317.
- Teti, R., Jemielniak, K., O'Donnell, G., & Dornfeld, D. (2010). Advanced monitoring of machining operations. *CIRP Annals—Manufacturing Technology*, 59, 717–739.
- Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York: Springer.
- Wang, C. C., & Too, G. P. J. (2002). Rotating machine fault detection based on HOS and artificial neural networks. *Journal of Intelligent Manufacturing*, 13, 283–293.
- Yuan, S. F., & Chu, F. L. (2007). Fault diagnosis based on particle optimization and support vector machines. *Mechanical Systems and Signal Processing*, 21(4), 1787–1798.
- Yang, B. S., Han, T., & Hwang, W. W. (2005). Fault diagnosis of rotating machinery based on multi-class support vector machines. *Journal of Mechanical Science and Technology*, 19(3), 846–859.
- Yang, Y., Yu, D., & Cheng, J. (2007). A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM. *Measurement*, 40, 943–950.
- Yang, Z. L., Wang, B., Dong, X. H., & Liu, H. (2012). Expert system of fault diagnosis for Gear Box in wind turbine. *Systems Engineering Procedia*, 4, 189–195.
- Ye, J., Janardan, R., Li, Q., & Park, H. (2004). Feature extraction via generalized uncorrelated linear discriminant analysis. In *The Proceedings of the international conference on machine learning*, pp. 895–902.
- Ye, J., & Xiong, T. (2006). Computational and theoretical analysis of null space and orthogonal linear discriminant analysis. *Journal of Machine Learning Research*, 7, 1183–1204.
- Yu, H., & Yang, J. (2001). A direct LDA algorithm for high-dimensional data-with application to face recognition. *Pattern Recognition*, 34, 2067–2070.
- Zhang, Y., Zuo, H., & Bai, F. (2013a). Classification of fault location and performance degradation of a roller bearing. *Measurement*, 46, 1178–1189.
- Zhang, Z., Wang, Y., & Wang, K. (2013b). Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network. *Journal of Intelligent Manufacturing*, 24, 1213–1227.