



Linear feature selection and classification using PNN and SFAM neural networks for a nearly online diagnosis of bearing naturally progressing degradations



Jaouher Ben Ali^{a,b,*}, Lotfi Saidi^a, Aymen Mouelhi^a, Brigitte Chebel-Morello^b, Farhat Fnaiech^a

^a University of Tunis, National Higher School of Engineers of Tunis, Laboratory of Signal Image and Energy Mastery (SIME),

5 Av. Taha Hussein, 1008 Tunis, Tunisia

^b FEMTO-ST Institute, AS2M Department, UMR CNRS 6174 – UFC/ENSMM/UTBM, 25000 Besançon, France

ARTICLE INFO

Article history:

Received 1 April 2014

Received in revised form

12 February 2015

Accepted 25 March 2015

Available online 13 April 2015

Keywords:

Empirical Mode Decomposition (EMD)

Linear Discriminant Analysis (LDA)

Principal Component Analysis (PCA)

Probabilistic Neural Network (PNN)

Scatter matrix

Simplified Fuzzy Adaptive Resonance Theory Map (SFAM)

ABSTRACT

In this work, an effort is made to characterize seven bearing states depending on the energy entropy of Intrinsic Mode Functions (IMFs) resulted from the Empirical Modes Decomposition (EMD). Three run-to-failure bearing vibration signals representing different defects either degraded or different failing components (roller, inner race and outer race) with healthy state lead to seven bearing states under study. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are used for feature reduction. Then, six classification scenarios are processed via a Probabilistic Neural Network (PNN) and a Simplified Fuzzy Adaptive Resonance Theory Map (SFAM) neural network. In other words, the three extracted feature data bases (EMD, PCA and LDA features) are processed firstly with SFAM and secondly with a combination of PNN–SFAM. The computation of classification accuracy and scattering criterion for each scenario shows that the EMD–LDA–PNN–SFAM combination is the suitable strategy for online bearing fault diagnosis. The proposed methodology reveals better generalization capability compared to previous works and it is validated by an online bearing fault diagnosis. The proposed strategy can be applied for the decision making of several assets.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Anomalous operating condition and fault detection of industrial assets are the core of process engineers in industry. Effective detection ensures the product quality improvement and reduces the cost and the time of repair. Consequently, industrial assets safety and reliability are very important to ensure the continuity of production, to improve process operation, to increase plant throughput, to reduce process downtime and to comply with increasingly stringent environmental rules and safety regulations (Lau et al., 2013).

Rolling Element Bearing (REB) is the most used solution for industrial machinery guiding in rotation. This critical component requires running with high reliability to decrease fault occurrences and fatal breakdowns of machineries (Lei et al., 2007; Konar and Chattopadhyay, 2011). It is well known that bearing defects are one of the most common fault sources in induction machineries (about 40–50%) (Widodo and Yang, 2007; Frosini and Bassi, 2010). Failure surveys by the electric power research institute indicate that bearing-

related faults are about 40% among the most frequent faults in induction motors (Bellini et al., 2008). Therefore, accurate strategies for REB defect detection are a trivial task nowadays in the industry.

Over the past few years, various signal processing methods have been proposed to detect and diagnose bearing defects. These methods may be roughly classified into vibration and acoustic measurements, temperature measurements and wear debris analysis (Nandi et al., 2005). The mechanical bearing vibration signals remain the most immediate, simplest, and richest source of information for understanding phenomena related to bearing defects (Seungdeog et al., 2011; Žvokelj et al., 2011). Certainly, vibration analysis is a powerful source to extract interesting information and to apply appropriate techniques. However, REB accelerations are considered as non-stationary and nonlinear (Randall and Antoni, 2011). Besides, noises present a serious trouble in the study of this type of signals (Zhang and Randall, 2009). Moreover, the relatively weak bearing signals are always affected by quite stronger signals (gear meshing, unbalance, bars, etc.) (Randall and Antoni, 2011).

Several efforts based on vibration signals have been recently proposed in the open literature to deal with the diagnosis of bearing faults. When a fault occurs on one bearing surface, the acceleration is characterized by the presence of periodic repetitive

* Corresponding author. Tel.: +216 96 568 115, +216 52 276 629.

E-mail address: benalijaouher@yahoo.fr (J. Ben Ali).

sharp peaks and further modulated by a number of harmonic frequencies (Zhang and Randall, 2009). However, interesting information is often contained in the periodicity of impacts, rather than in the rest of the signal frequency content. Thereby, vibration signals are non-stationary and always are masked by machine noise (Randall and Antoni, 2011). For this reason, conventional vibration analysis techniques, which were successfully employed for stationary signals such as fast Fourier transform, have not provided good and significant results based on REB vibration signals (Braun and Feldman, 2011).

In the literature, non-stationary or transient signals are often decomposed into sub-bands where stationary and linear characteristics could be obtained and fault patterns are easily extracted (Priestley, 1988). Consequently, Wavelet Transforms (WT) (Wang et al., 2015), Empirical Mode Decomposition (EMD) (Liu et al., 2015), Local Characteristic-scale Decomposition (LCD) (Zheng et al., 2013a), Ensemble Empirical Mode Decomposition (Wang et al., 2014), and Generalized Empirical Mode Decomposition (Zheng et al., 2013b) were widely used for the processing of bearing vibration signals in the presence of nonlinear and non-stationary data. By analyzing some series of decomposed REB signals based on WT and EMD, it was demonstrated that the energy entropy changes in different frequency bands when a bearing fault occurs. Besides, EMD is more highlighted than WT in terms of mean fault characterization (Tavakkoli and Teshnehlab, 2007).

After decomposing bearing accelerations into some linear and stationary signals, it is recommended to extract the most useful features. Bearing fault diagnosis is not an easy task, it is essentially a problem of pattern recognition. The most effective features and accurate classifiers are needed to obtain higher diagnostic accuracy (Li and Zhang, 2011).

Several previous works based on feature extraction and classification techniques have been proposed for bearing fault diagnosis. Yang et al. have used the discrete WT to decompose vibration and acoustic signals into different frequency levels. Then, some statistical measures were extracted from the first-four levels. Also, an Artificial Neural Networks (ANNs) combined with a Support Vector Machine (SVM) were used to classify the faults of small reciprocating compressor used in refrigerators (Yang et al., 2007). In Lin (2010), the EEMD was applied to decompose vibration signals into different Intrinsic Mode Functions (IMFs) and then the Hilbert spectrum was computed to extract bearing fault features. Bin et al. (2012) have demonstrated that the WT generally has some shortcomings and difficulties in selecting suitable basic function because there are no standards or general selection rules for different tasks. However, the EMD is an automated method that extracts easily failure patterns thanks to the use of a wide scale of frequency. Saidi et al. (2014) have presented a combination of EMD and bi-spectrum. The computation of the bi-spectrum was based on the first IMF which contains more information. The proposed combination was promising but it needs always an expert intervention for the analyses of the bi-spectrum plot. Unfortunately, all these methods presented an offline bearing fault detection. Generally, some holes were created in the tested REB and subsequently the fault detection efficiency of the proposed method was given. In this case, the bearing vibration signals increase and the fault detection task become effortless. In an industrial environment, REBs are in continuous rotation where fault characteristics are always submerged and hidden by noises. So, it is recommended to investigate bearing defects online by using run-to-failure histories.

The online bearing monitoring is very important to avoid unexpected breakdowns where the breakdown of a single machine can halt the production of the entire process chain. However, investigations related to naturally induced and naturally progressed defects of REBs are relatively scarce. In fact, online detection of REB degradation for naturally progressing defect in damage stages is not available in the literature (Shakya et al.,

2014). Recently, some approaches have been proposed for the online bearing health monitoring. Pan et al. (2014) have proposed a new approach based on dynamic fuzzy neural networks where only consequent parameters were updated online. In Shakya et al. (2014), various damage stages for naturally progressing REB defect were given online. For this, time domain, frequency domain and time–frequency domain were together under study. The Mahalanobis distance was used as a trend parameter for bearing fault diagnosis. In Loutas et al. (2011), the combination of vibration, acoustic emission and oil debris measures is used in an online condition monitoring of bearings and gearboxes. Several techniques were combined such as fusion technique, Principal Component Analysis (PCA) and Independent Component Analysis (ICA) to identify different damage modes. The computational time was not discussed in this work and with respect the definition of failure thresholds was arbitrary. In Ben Ali et al. (2015b), the combination of traditional statistical features and EMD energy entropy were jointed with the back propagation neural network for an online bearing fault diagnosis. The classification results were relatively low (93%) that is why the authors have proposed a health index for a more reliable online monitoring.

In this paper, an attempt is made to diagnose the state of bearings using a new nearly online method. The proposed method is based on the EMD feature extraction, linear feature reduction and feature classification using neural networks. The validation of the proposed method is based on bearing run-to-failure histories. This method is able to detect online perfectly the state of the bearing. The remainder of this paper is organized as follows: Section 2 is devoted to describing the used techniques in this work. This section describes the different steps of the EMD method. Also, a brief description of Linear Discriminant Analysis and Principal Component Analysis algorithms is given. Besides, this section details the different steps to implement the used neural networks in this work. Section 3 is dedicated to the experimental results by analyzing the proposed feature extraction, feature reduction and feature classification techniques. Section 4 gives a numerical analysis of the most effective features based on a mathematical criterion. Also, a good discussion and analysis of the experimental results by comparing the performances of the proposed method with some previous works is provided in this section. Finally, conclusions and prospects of this work are given in Section 5.

2. Methods

2.1. Empirical Mode Decomposition (EMD)

EMD algorithm is developed from the simple assumption that any signal consists of different simple intrinsic modes of oscillations called Intrinsic Mode Functions (IMFs). This adaptive decomposition method is especially applicable to the analysis of nonlinear and non-stationary signals (Yu et al., 2006).

By the definition, any signal $x(t)$ can be decomposed as detailed in Table 1 (Huang et al., 2011).

The step (G) in Table 1 presents a stoppage criterion for the EMD which is one of the weaknesses of this algorithm (Lei et al., 2013). Huang et al. (1998) proposed a robust way to access the IMF component of a signal as “Sifting process”. The sifting process is stopped by limiting the size of the standard deviation (S_D), computed from two consecutive results as

$$0.2 < S_D = \sum_{t=0}^T \frac{[h_{k-1}(t) - h_k(t)]^2}{h_{k-1}^2(t)} < 0.3 \quad (1)$$

Numerous publications of the EMD for bearing fault diagnosis have shown several advantages and very encouraging results.

Table 1
EMD algorithm.

- (1) Initialize: $r_0 = x(t)$, and $i = 1$
- (2) Extract the i th IMF c_i
 - A) Initialize: $h_i(k-1) = r_i - 1$, $k = 1$
 - B) Extract the local maxima and minima of $h_i(k-1)$
 - C) Interpolate the local maxima and the minima by cubic spline lines to form upper and lower envelopes of $h_i(k-1)$
 - D) Compute the mean $m_i(k-1)$ of the upper and lower envelopes of $h_i(k-1)$
 - E) Let $h_{ik} = h_i(k-1) - m_i(k-1)$
 - F) If h_{ik} is an imf then set $c_i = h_{ik}$ else go to step (B) with $k = k+1$ define the residue $r_{i+1} = r_i - c_i$
 - G) If r_{i+1} still has least 2 extrema then go to step (2) with $i = i+1$ else the decomposition process is finished and r_{i+1} is the monotonic residue of the signal.

Nowadays, EMD is one of the most powerful advanced signal processing techniques and it has been extensively studied and widely applied in fault diagnosis of rotating machinery. To achieve better results, it is recommended to combine the EMD with other signal processing techniques or artificial intelligent techniques for an effective strategy for bearing diagnosis (Lei et al., 2013).

2.2. Brief description of Principal Component Analysis

In literature, Principal Component Analysis (PCA) is the most used technique for feature reduction which means transforming the original features into a lower dimensional space. PCA, as a linear technique, is a quantitatively rigorous method for achieving data dimensionality reduction of the extracted features. The method generates a new set of variables, called Principal Components (PCs), which maximize the variance of the projected vectors. Each PC is a linear combination of the original variables. All PCs are orthogonal to each other, so there is no redundant information. The PCs are a whole of an orthogonal basis for the data space. Thus, the first PC consists of the high variability, the second PC consists of the next highest variability and so on for other directions. The first few components are kept and others with less variability are discarded (Widodo et al., 2007). In this work, the PCA algorithm will be used to eliminate the non-significant features in the recorded vibration signals. The step by step procedure of PCA is provided below. Let us say we have N observations of M -dimensional data.

- **Step 1:** To perform PCA we should first compute the covariance matrix $\Sigma \in \mathbb{R}^{M \times M}$; $\Sigma = E\{(x - \bar{x})(x - \bar{x})^T\}$, where x is the given signal, \bar{x} is its mean vector, and $E\{\cdot\}$ is an average over an ensemble of N samples.
- **Step 2:** Find matrix Λ of eigenvectors and diagonal elements of matrix D as eigenvalues of covariance matrix Σ as given by, $\Lambda^{-1} \Sigma \Lambda = D$.
- **Step 3:** Sort the eigenvectors of PCs in the decreasing order of importance of eigenvalues in D .
- **Step 4:** Project the data into the directions of sorted eigenvectors by taking the dot product between the given data and eigenvectors.
- **Step 5:** Select the first few principal components depending on the containment of a given percentage of variability (like 95% or 98% depending on the problem).

2.3. Brief look into Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a mathematical transformation based on a linear discriminant combination. As the PCA method, LDA is used to reduce features by transforming the original features into a lower dimensional space. Typically, LDA is used to analyze the original dataset and to produce $K-1$

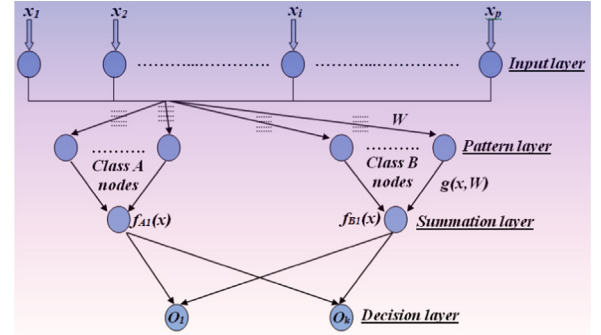


Fig. 1. Probabilistic Neural Network architecture.

independent components where K is the number of classes. Generally, LDA dataset preprocessing is more suitable for classification tasks than PCA (Mouelhi et al., 2013): PCA tends to minimize feature variances of the same class although LDA tends to minimize feature variances of the same class and to maximize feature variances of different classes. The aim of LDA is to create new variables by maximizing discrimination between all classes. New variables maximize the differences between the predefined classes and minimize differences in the same class (Subasi and Gursoy, 2010).

2.4. Probabilistic Neural Network

The first formulation of the Probabilistic Neural Network (PNN) was by Specht in 1990. The different steps of PNN algorithm are well described in Specht (1990). Brief, PNN is based on the computing of Bayes statistics for two categories (A and B). Consequently, for each p dimensional vector x , the state θ could either be θ_A or θ_B . The state decision is based on the Bayes rule as follows:

- If $h_A l_A f_A(x) > h_B l_B f_B(x)$ then $d(x) = \theta_A$: State A .
- Contrarily, if $h_A l_A f_A(x) < h_B l_B f_B(x)$ then $d(x) = \theta_B$: State B .

where $f_A(x)$ and $f_B(x)$ are respectively the probability density functions of the category A and B , l_A and l_B are the loss functions associated respectively with the decision $d(x) = \theta_B$ and $d(x) = \theta_A$, h_A is the prior probability of category A pattern occurrences and vice-versa for h_B . We should note that the boundary of density probabilities A and B is given by

$$f_A(x) = \frac{h_B l_B}{h_A l_A} \cdot f_B(x) \quad (2)$$

As shown in Fig. 1, PNN consists of four layers the input layer, the pattern layer, the summation layer and the decision layer. The p nodes in the input layer consist of the p input vector components. The pattern layer is designed to receive all components of

the input vector x . For each pattern node i , the product $Z_i = xW_i$ is computed. To deal with the nonlinear feature distributions for any classification task, a nonlinear function, where x and W_i should be normalized to the unit length, is introduced as follows:

$$\exp\left[\frac{Z_i - 1}{2}\right] = \exp\left[-\frac{(W_i - X)^T(W_i - X)}{2\sigma^2}\right] \quad (3)$$

After computing the nonlinear function for all nodes, this function will be transferred to the summation layer. Afterwards, summation units sum pattern unit inputs which correspond to the desired category. The decision layer consists of two neurons that produce binary outputs taking into account the variable weight C_k with the following mathematical expression:

$$C_k = -\frac{h_{B_k} l_{B_k} n_{A_k}}{h_{A_k} l_{A_k} n_{B_k}} \quad (4)$$

where n_A and n_B is the number of training patterns from the category A and B respectively. More details are given by [Specht \(1990\)](#).

2.5. Simplified Fuzzy Adaptive Resonance Theory Map neural network (SFAM)

The Simplified Fuzzy Adaptive Resonance Theory Map (SFAM) neural network is a fast, incremental, supervised learning system for analog inputs ([Ben Ali et al., 2014](#)). In 1993, Kasuba proposed the SFAM network which is a simplification of Fuzzy ARTMAP ([Venkatesan and Suresh., 2009](#)). This neural network is a step ahead of Fuzzy ARTMAP in reducing the computation and architectural redundancy.

As illustrated in [Fig. 2](#), the SFAM is formed of three layers; the input layer F_1 , the competitive layer F_2 and the output layer F^{ab} .

The activity patterns of layers F_1 and F_2 are fully interconnected. Each neuron is initially connected to every neuron on the other layer. All connections are weighted by synaptic weights lying between 0 and 1. Each neuron of the F_2 layer represents one category formed by the network and it is characterized by its weight vector W_j (j is the index of the neuron). The weight vector's size is equal to the dimension M of layer F_1 . In the training process, if the weights of a neuron are not modified, this neuron is called uncommitted. Inversely, once a neuron's weights have been modified, this neuron is named committed.

SFAM is based on a form of normalization called complement coding. The operation consists on taking the input vector and concatenating it with its complement. The resulting vector should be presented to the layer F_1 . Therefore, the dimension M of layer F_1 is equal to the double of the input vector's dimension ($M=2L'$).

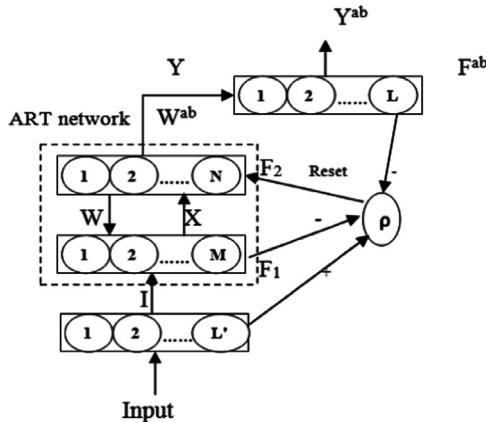


Fig. 2. SFAM neural network architecture.

The F^{ab} layer is connected with the competitive layer F_2 by the W^{ab} weights to locate the appropriate class t of the input vector in the test mode.

The SFAM has a self-determining architecture: the training vectors are presented to the network, the output category nodes are grown and assigned to a particular class. If there is a new category of input training example, a new output category node is grown and its weights are set equal to the elements of the complement-coded input vector I .

For further applied training vectors, the subset hood of the weight vectors to the complement coded input is computed using the activation function

$$T_j(I) = \frac{|I \wedge W_j|}{\alpha + |W_j|} \quad (5)$$

where \wedge is the fuzzy AND operator; α is a very small value (typically 10^{-5}); and j is the number of the current output category node.

The node with the highest $T_j(I)$, called the “winner”, is initially chosen (G) to represent the input vector. To know if associating W_G with I is a good match, the matching function $MF(I, W_G)$ between the two is calculated as

$$MF(I, W_G) = \frac{|I \wedge W_G|}{|I|} \quad (6)$$

The matching function should be compared to the vigilance parameter ρ .

- If $MF(I, W_G) \leq \rho$, the winner G is disabled and the next most activated node is selected.
- If $MF(I, W_G) > \rho$, a category agreement test between the winner and the input example is performed.

If the categories disagree, the vigilance ρ is increased by a small amount α and the winner is disabled. If they agree the SFAM is considered in a state of resonance and the weight vector W_G is updated using the following formula:

$$W_G^{new} = (1 - \beta)W_G^{old} + \beta(I \wedge W_G^{old}) \quad (7)$$

where β is the learning rate introduced to select the mode of learning as follows:

- If $\beta=1$, the fast learning mode is selected with just one iteration of the training data.
- Whereas if $0 < \beta < 1$, slow learning is selected and it requires a number of iterations for the training.

Finally, when classifying, each test vector is applied in turn and its predicted class is that of the node with the highest activation. More details are given in [Ben Ali et al. \(2014, 2015a\)](#).

3. Experimental results

3.1. Experimental setup and data recording

The vibration signals used in this paper were provided by the Center for Intelligent Maintenance Systems (IMS), University of Cincinnati, USA, in collaboration with the National Aeronautics and Space Administration (NASA). A schematic of experimental test rig is shown in [Fig. 3](#). Four Rexnord ZA-2115 double row bearings are installed on the shaft. Each bearing contains 16 rollers (for each row), a pitch diameter of 2.815 in., a roller diameter of 0.331 in., and a tapered contact angle of 15.17° ([Qiu et al., 2006](#)).

Three tests were made. Each test is an experience of 4 bearings. In this way 12 bearings are used but only 4 bearings have reached

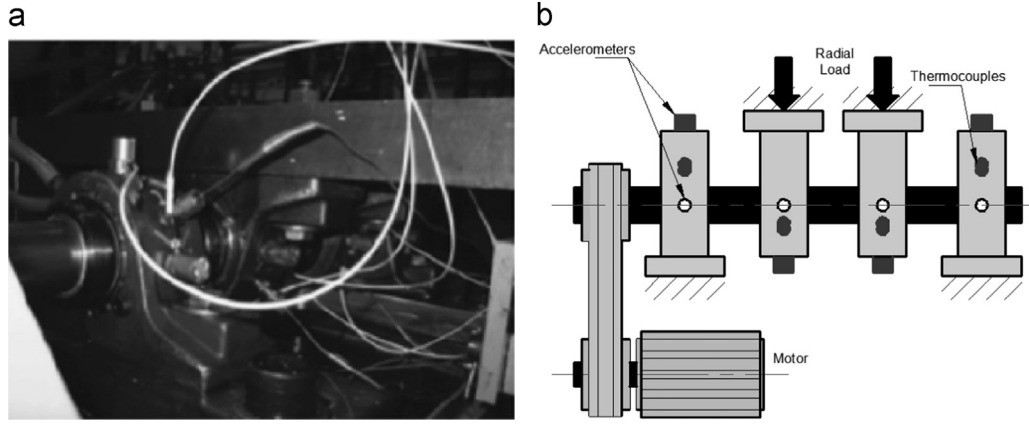


Fig. 3. Experimental test rig, (a) photo from the IMS center (Qiu et al., 2006) and (b) simplified schematic of the test rig.

Table 2
IMS bearing dataset description.

	Test 1	Test 2	Test 3
Recording duration	October 22, 2003 12:06:24 to November 25, 2003 23:39:56.	February 12, 2004 10:32:39 to February 19, 2004 06:22:39.	March 4, 2004 09:27:46 to April 4, 2004 19:01:57.
#. of files	2156	984	4448
#. of channels	8	4	4
File record-interval	Every 10 min (except the first 43 files were taken every 5 min).	Every 10 min	Every 10 min.
Description	At the end of the test-to-failure experiment, inner race defect occurred in bearing 3 and roller element defect in bearing 4.	At the end of the test-to-failure experiment, outer race failure occurred in bearing 1.	At the end of the test-to-failure experiment, outer race failure occurred in bearing 3.

failure with known defects. Each data set describes a run-to-failure experiment. It consists of individual files that are 1-second vibration signal snapshots recorded at specific intervals (every 10 min). Each file consists of 20,480 points with the sampling rate set at 20 kHz. Records (row) in the data ASCII files are data points. Data collection is provided by NI DAQ Card 6062E (Ben Ali et al., 2015b).

The three tests have been carried out for 35 days until a significant amount of metal debris is found on the magnetic plug of the test bearing. Larger intervals of time stamps (showed in file names) indicate resumption of the experiment in the next working day. Table 2 summarizes the necessary information of the IMS bearing dataset and full details are provided at (<http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/>).

In the first test, two perpendicular sensors (12 o'clock and 3 o'clock positions) were used to record bearing vibration signals. However, only one sensor (12 o'clock) was used. In this paper only the 12 o'clock is used. We should also note that the last file in the first test is uninformative because the record was made in the stoppage phase of the machine.

Table 2 focuses on some important points. The REB life is a highly nonlinear phenomenon. The 12 used bearings in the experimental setup were tested in the same conditions of speed and load. Despite that, only four bearings were broken and one bearing was slightly degraded (bearing 3 of testing 3). Moreover, the failure time and the failure nature of the three broken bearings were not the same. All these confirm that online bearing fault assessment is a very hard task.

3.2. Feature extraction

In this section, the three broken bearings in the IMS database are explored. These bearings are bearing 4 and 3 of testing 1 ending respectively with a failure roller and a failure inner race

and bearing 1 of testing 2 ending with a failure outer race. The run-to-failure vibration signals of these bearings are shown in Fig. 4. Each run-to-failure history is divided into three zones the healthy zone (number 1), the degradation zone (number 2) and the faulting zone (number 3). This partition was done based on the magnitude of vibration signals by proposing two thresholds for each history. Consequently, seven REB states are considered: healthy (H), degraded roller (DR), failure roller (FR), degraded inner race (DIR), failure inner race (FIR), degraded outer race (DOR), and failure outer race (FOR).

Every 10 min, an individual file is recorded during 1 s vibration signal with the sampling rate 20 kHz. The original vibration signals recorded during 1 s are decomposed into some IMFs. For each decomposition, the number of the found IMFs is higher than 12. Consequently, the first 12 IMFs are chosen to extract features. However, the computational time was very important and sometimes the EMD algorithm diverges and no IMF is given even after 10 min. To solve these troubles, we propose the following steps for EMD feature extraction:

- Every 10 min, extract vibration signals with recording duration of one second from bearing run-to-failure accelerations.
- Decompose just the first part of the extracted signal (with 0.1 s as length) into some IMFs. If the EMD algorithm is still running after 10 s, stop it and consider the second part of the signal (from 0.1 s to 0.2 s) and so on.
- Compute the total energy E_i of the first 12 IMFs as (where i is the IMF number):

$$E_i = \sum_{j=1}^{2048} |c_{ij}|^2 \quad (8)$$

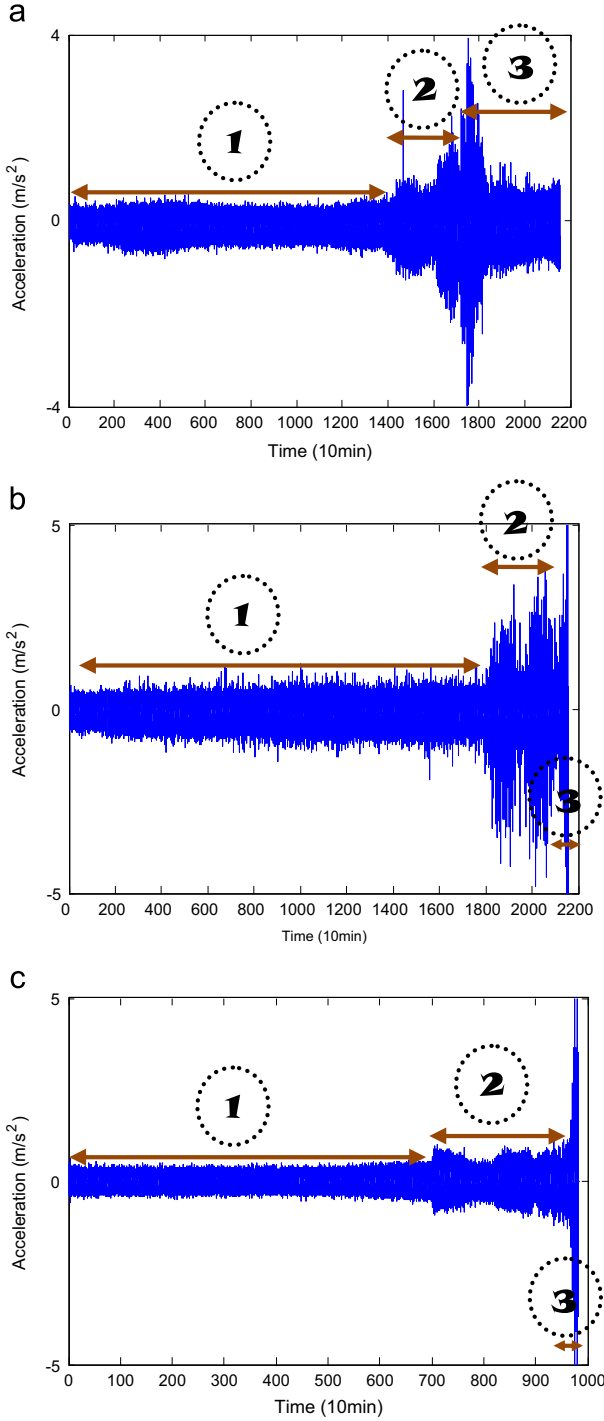


Fig. 4. (a) Acceleration of bearing 4 of testing 1 ending with failure roller. (b) Acceleration of bearing 3 of testing 1 ending with failure inner race. (c) Acceleration of bearing 1 of testing 2 ending with failure outer race.

iv. Compute the total energy of all IMFs:

$$E = \sum_{i=1}^{12} E_i \quad (9)$$

v. Construct the feature vector as

$$\left[H_{en}, \frac{E_1}{E}, \frac{E_2}{E}, \dots, \frac{E_{12}}{E} \right] = [H_{en}, H_{en}IMF_1, H_{en}IMF_2, \dots, H_{en}IMF_{12}] \quad (10)$$

where H_{en} is EMD energy entropy of the whole original signal and $H_{en}IMF_G$ is EMD energy entropy of the IMF number G . The EMD energy entropy is designated as (Hong et al., 2014)

$$H_{en} = - \sum_{i=1}^n p_i \log(p_i) \quad (11)$$

where $p_i = E_i/E$ is the percent of the energy of $c_i(t)$ in the whole signal energy.

By considering just the first part of the vibration signal, the number of the treated points by the EMD algorithm is reduced from 20,480 to 2048 and the computational time is reduced impressively. Besides, if the computational time of the EMD algorithm exceeds 10 s, we consider the next part of the vibration signals. This is a very smart step to ensure a quick convergence of the proposed feature extraction strategy. We note that for all the considered vibration signals, just the first three parts of the signal were considered and thereby the maximum computational was time less than 30 s. Note that the reported time is on a typical PC configured with CPU i3 2.3 GHz, 4 GB RAM, and running Windows 7.

The energy entropy of the original signal and the first twelve IMFs of the bearing 1 of testing with an outer race defect (shown in Fig. 4c) are shown in Fig. 5. In this plot, feature number zero is the energy entropy of the original signal and the other features (from 1 to 12) are the energy entropies of IMFs (from 1 to 12 respectively). Through Fig. 5 it is obvious that the energy entropy is near zero for the last IMFs.

Yu et al. (2006) demonstrated that the EMD energy entropy changes with the energy variation of bearing fault vibration signals when the REB operates with different faults. Thereby, we propose to use the energy entropy of the first 12 IMFs as features to characterize REB defects. We select just the first 12 IMFs because it is the minimum IMF number found using the EMD. Thus, we are sure that any REB vibration signals produce more than 12 IMFs.

3.3. Feature reduction and classification

To show the repartition of the extracted features, 50 vectors were selected arbitrarily from three classes (H, DR, FR). Fig. 6 shows the energy repartition of the first eight IMF energy entropies and the original signal energy entropy. This plot clearly demonstrates that the first IMFs consist of the highest discrimination. For more clarification, Figs. 7 and 8 show two bi-dimensional plots of the highest features' discrimination (shown in Fig. 6a) and lowest features' discrimination (shown Fig. 6g) respectively. Figs. 7 and 8 reveal that the overlapping of features increases for IMFs higher order than 4.

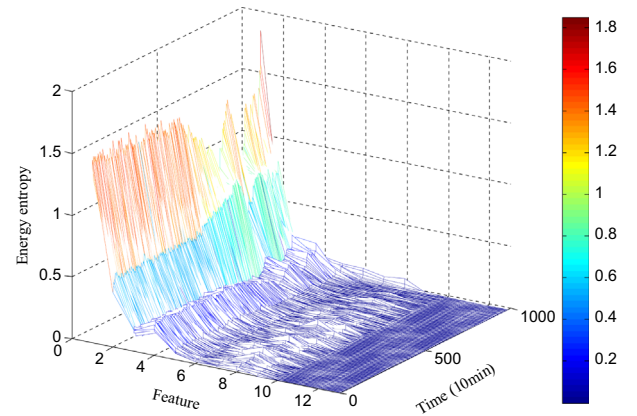


Fig. 5. Original extracted features from the bearing 1 of testing 2.

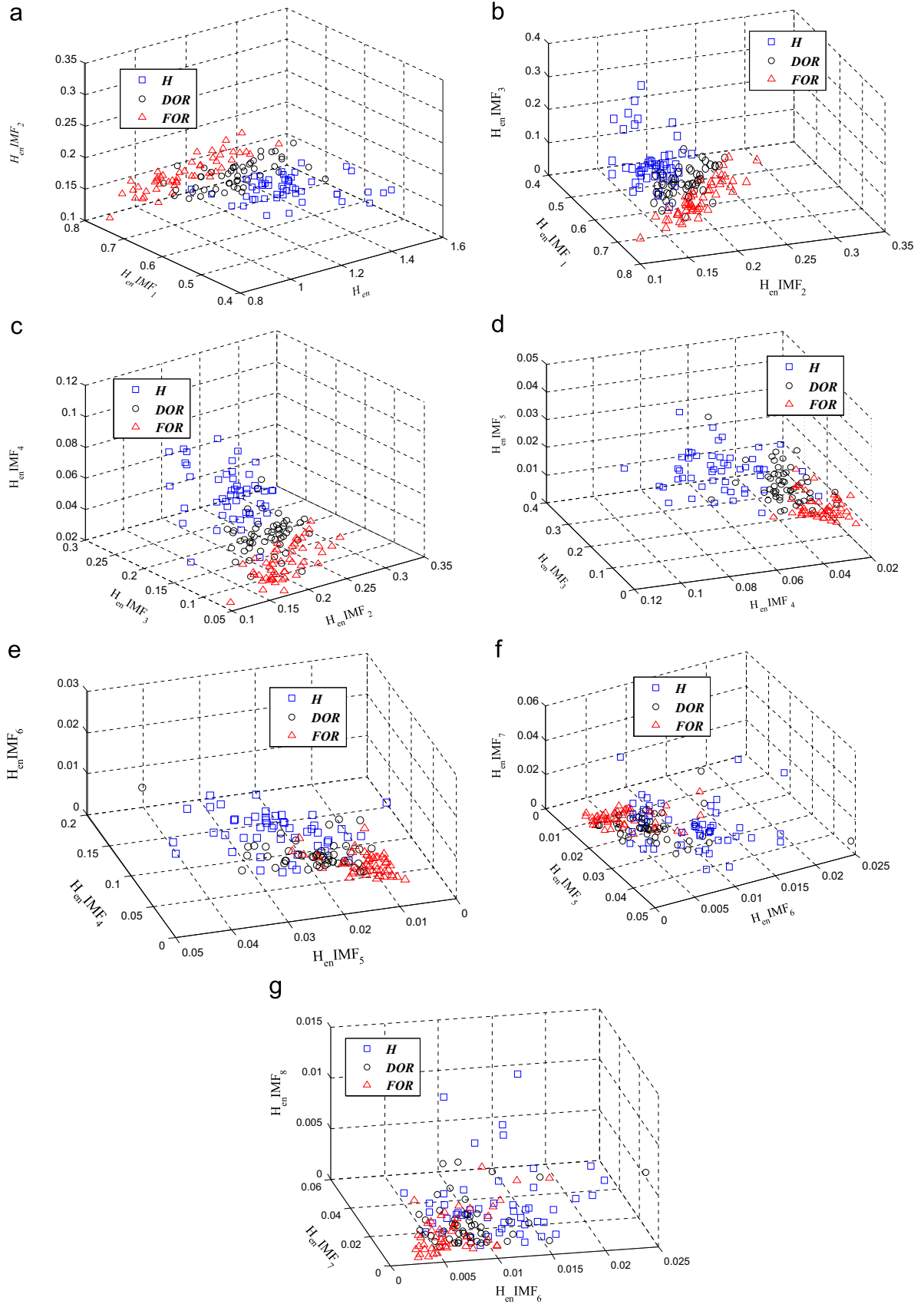


Fig. 6. Original feature distributions.

As shown previously, plotting EMD energy entropy features indicates the necessity of the preprocessing step them to generate separable features more suitable for the classification task.

Disorder distribution of original features tends to decrease the performance of the used classifier. So, the combination of these features directly with a classifier is not recommended.

To avoid this drawback, Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are proposed to extract and to reduce the feature dimensionality based on the eigenvalues of the covariance matrix. Firstly, because the overlap between normal-degradation zones and degradation-faulting zones, the number of records of the broken three bearings is reduced from 5394 to 3911. This reduction is made with a very great caution to ensure the reliability of measures. Then, the new data set is divided into two sub-sets: one for training (containing 60% of the samples), and the other for test (containing 40% of the samples).

By using the PCA for the training dataset, the first eight principal components have been selected to replace the original feature vector (which contains 13 components). As shown in Fig. 9, the number of features is reduced from 13 to 8 where principal components 9, 10, 11, 12 and 13 are discarded.

By using LDA technique, original features of the training dataset are reduced to automatically to $K-1$ where K is the number of classes. In this paper, we have used seven REB states. Thereby, the original features are reduced from 13 to 6 by applying the LDA.

Fig. 10b and c shows the distribution of the first three features extracted using PCA and LDA respectively. To evaluate the

importance of PCA and LDA techniques, in Fig. 10 we have used the same features shown in Fig. 6. To have a good interpretation, Fig. 10a shows the optimal and the desired feature distribution where classes are perfectly separated.

Compared to previous works, this paper is based on a large data set which contains 3911 records extracted from run-to-failure REB vibration signals. For each record, the EMD energy entropy is used to extract 13 features. Then, the PCA and LDA are employed to reduce features to 8 and to 6 respectively.

To obtain a good analysis of the proposed method, six scenarios are proposed

- i. **Scenario 1:** Combination of original features with SFAM.
- ii. **Scenario 2:** Combination of PCA and SFAM.
- iii. **Scenario 3:** Combination of LDA and SFAM.
- iv. **Scenario 4:** Combination of original features, PNN and SFAM.
- v. **Scenario 5:** Combination of PCA, PNN and SFAM.
- vi. **Scenario 6:** Combination of LDA, PNN and SFAM.

In the first three scenarios, the SFAM neural network is used to classify seven REB states (H, DIR, FIR, DOR, FOR, DR, FR). However,

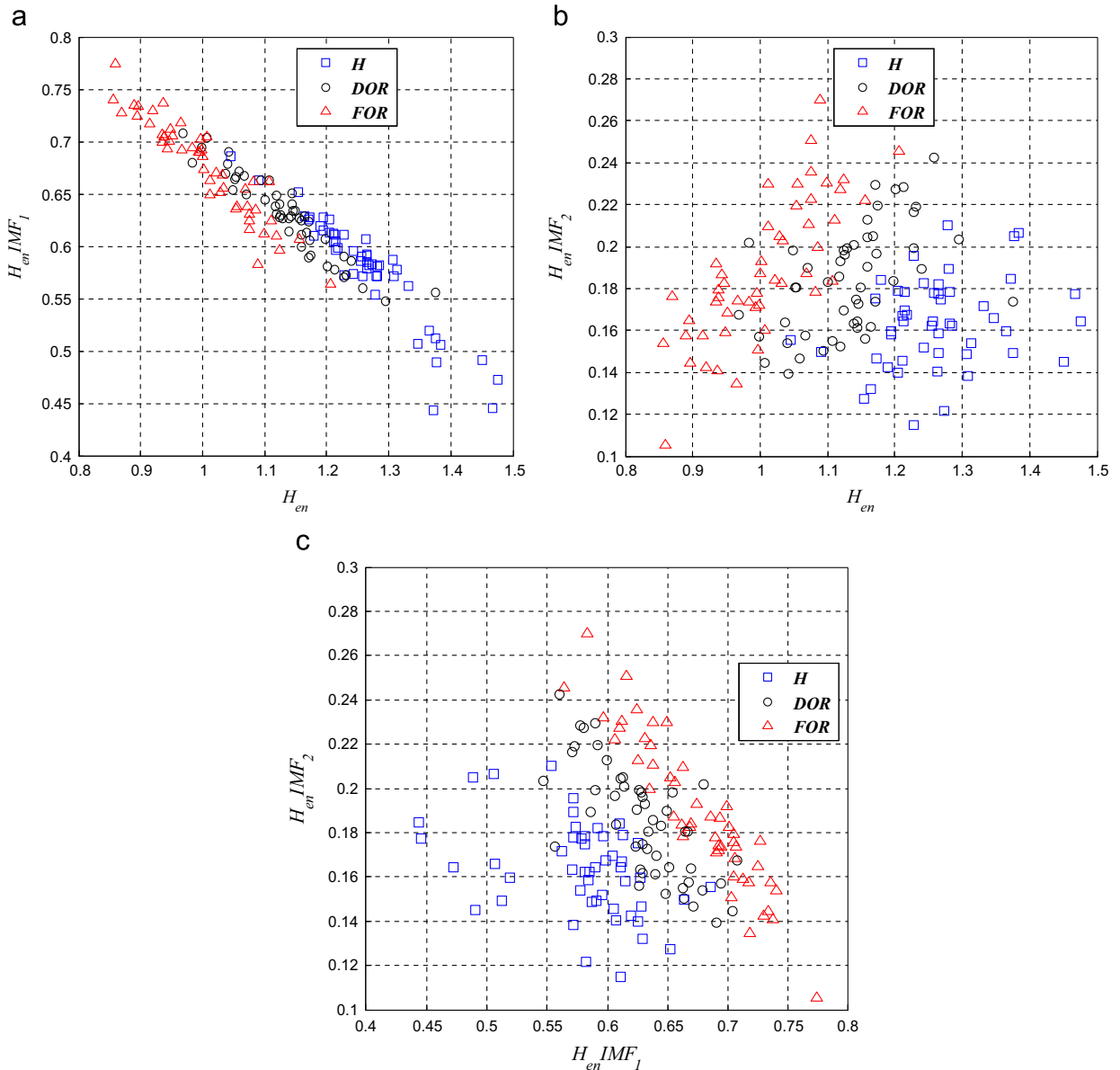


Fig. 7. Bi-dimensional distributions of highest features' discrimination (shown in Fig. 6a).

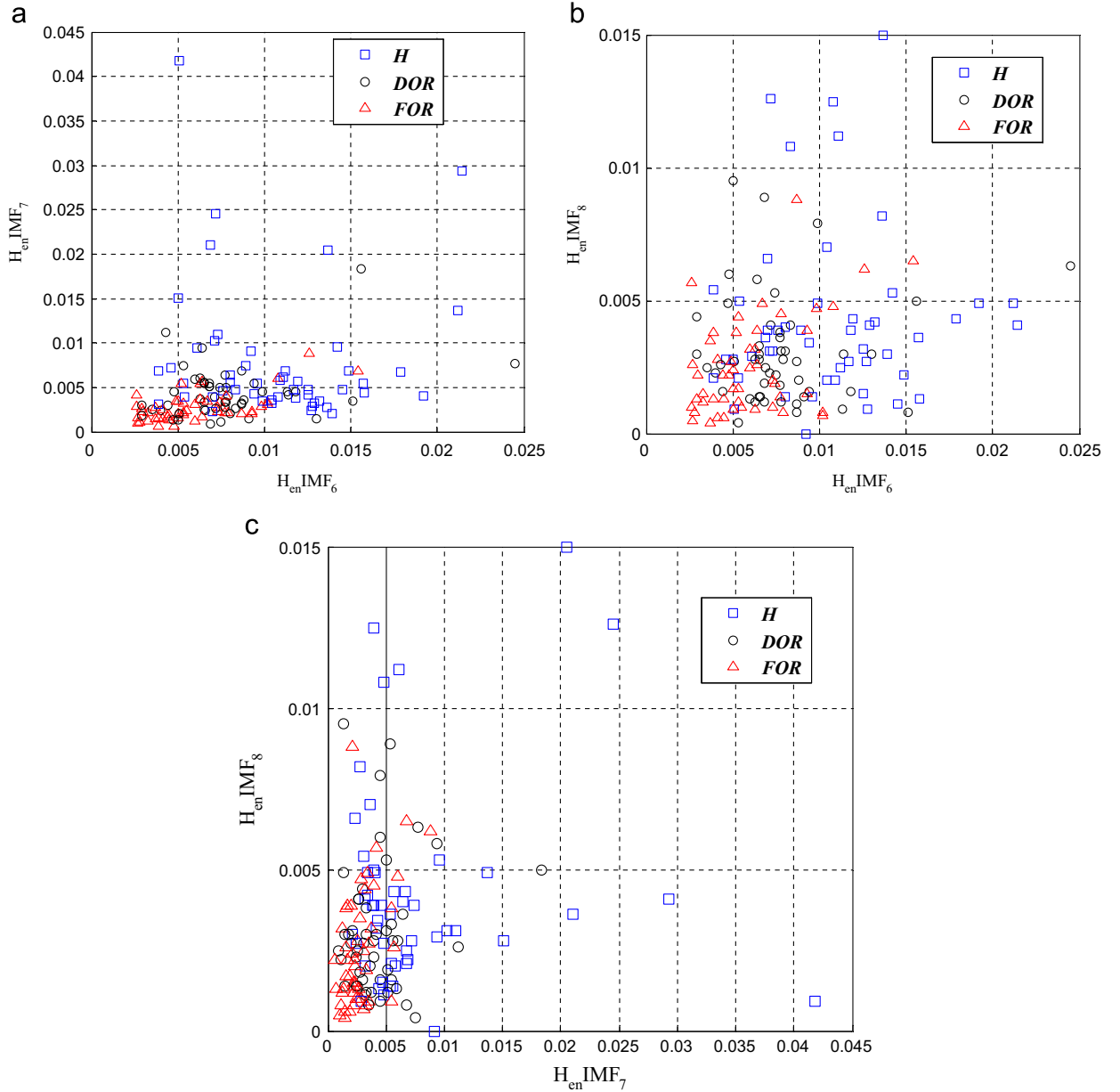


Fig. 8. Bi-dimensional distributions of the lowest features' discriminations (shown in Fig. 6g).

in the last three scenarios, the PNN is introduced to classify REB states in two classes healthy and not healthy. If the PNN decision shows that the processed state is unhealthy, then the SFAM is used to detect the type of failure (DIR, FIR, DOR, FOR, DR, FR).

For each scenario, the training-bearing dataset is divided into ten equal sized sub-datasets (ten folds). Then, ten iterations of training and validation are performed. One iteration presents nine folds for the training and one fold for the validation. Also, ten folds cross-validation experiments are done to validate the classification accuracy of each scenario.

After the validation phase, we have projected the testing dataset using PCA and LDA techniques. Then, the classification accuracy (CA) is computed for each scenario. The CA is the ratio between the total number of correctly classified test samples to the total number of test samples and it can be expressed as follows:

$$CA[\%] = \frac{(\text{number of correctly classified samples})}{(\text{total number of samples in testing dataset})} \times 100 \quad (12)$$

Table 3 shows the CA for each scenario where the introduction of PNN is highlighted compared to the direct use of the SFAM neural network. The PNN neural network is able to classify the whole dataset in two classes (healthy and not healthy) with a CA about 100%. Compared to the use of SFAM only, the proposed combination of PNN–SFAM is very promising for classification tasks. Besides, LDA is better than PCA in terms of feature discrimination. This result is validated when comparing the CA of scenarios (2,3) and (5,6). Better experimental results are insured by the combination of LDA, PNN and SFAM as illustrated by the 6th scenario summarized in Fig. 11.

3.4. Nearly online bearing fault diagnosis

As mentioned above, the online diagnosis is a very hard task for scientists. With the recent advances in modern technology, industry and scientist researchers are progressing toward providing enhanced strategies to detect online REB anomalies.

In this section, we will present the application of the proposed EMD–LDA–PNN–SFAM combination for an online diagnosis task.

We recall that the three used bearings (see Fig. 4) contain 5394 records were just 3911 are used in the training and testing modes. Consequently, the other 1483 records are unknown for the validated SFAM and PNN neural networks. Experimental results, given in Fig. 12, show that the application of the proposed strategy is very promising for a nearly online bearing diagnosis.

The online evaluation of the proposed strategy for the diagnosis of naturally bearing progressing defects is very encouraging. The proposed strategy is able to indicate qualitatively a clear

distinction between the healthy and damaged states. In addition, this approach proposes automatic fault detection without any human intervention. The application of the proposed methodology perfectly separates the healthy stage, the degraded stage and the faulting stage. Besides, the type and the severity of defects are given for all tests which provide an idea about the remaining useful life of the tested bearing. However, the three tested bearings are already modeled by the PNN and SFAM neural networks and another test is needed. Another bearing is needed where records were not yet treated by the neural networks. Hence, we thought to use the bearing 3 of testing 3 where any record was used in the training phase. Experimental results shown in Fig. 13 confirm that the proposed methodology is really interesting. The probability density of failure increases at the end of the experiment although that the magnitude of vibration signals remains substantially constant. For more clarity and better resolution, experimental results are enlarged in Fig. 14.

To improve the quality of the proposed diagnosis approach, we propose a smart step to minimize the false alarm rate: It is a sliding window that filters external tests which were not used in the training mode and it can be summarized as follows:

- Select a sliding window (sliding step ($ss=1$), window length ($wl=16$)) to browse the nearly online classification results.
- Set $k_{DIR}=k_{FIR}=k_{DOR}=k_{FOR}=k_{DR}=k_{FR}=0$, where k_i is the occurrence of the defect i along the sliding window.
- For each point of the window, if the classification result is i then $k_i=k_i+1$.
- After 16 sliding steps, determine the highest occurrence as $occ=\max(k_{DIR}, k_{FIR}, k_{DOR}, k_{FOR}, k_{DR}, k_{FR})$.
- Determine the fault with the highest occurrence as $occtype=\arg\max(k_{DIR}, k_{FIR}, k_{DOR}, k_{FOR}, k_{DR}, k_{FR})$.
- If $occ \geq 2$ then the fault type is $occtype$ else the state of bearing is healthy.
- Run the next 16 classification results.

Table 3

Testing accuracies of the proposed methodologies.

Scenario	1	2	3	4	5	6
CA [%]	92.32	93.41	94.05	92.77	95.07	97.82

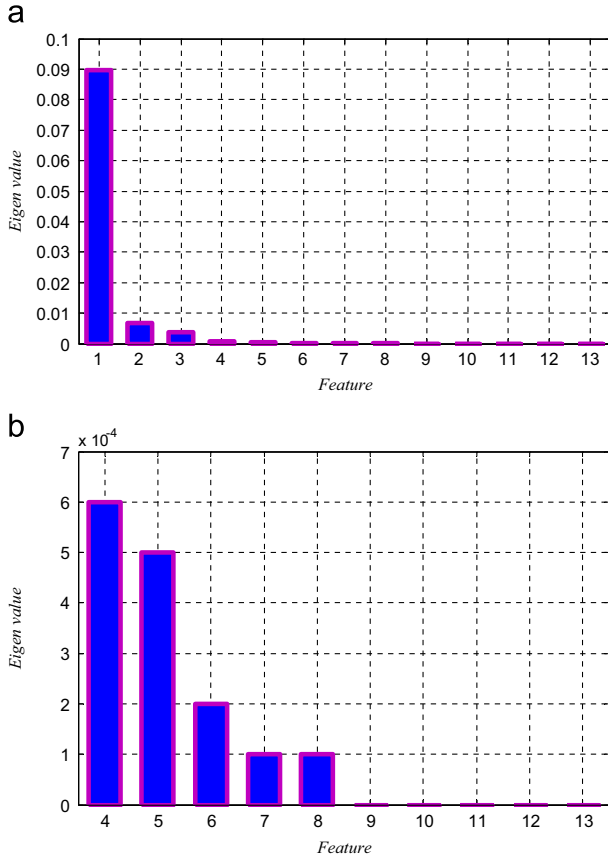


Fig. 9. PCA eigenvalue: (a) entire distribution of covariance matrix and (b) zoom in the area of the last features.

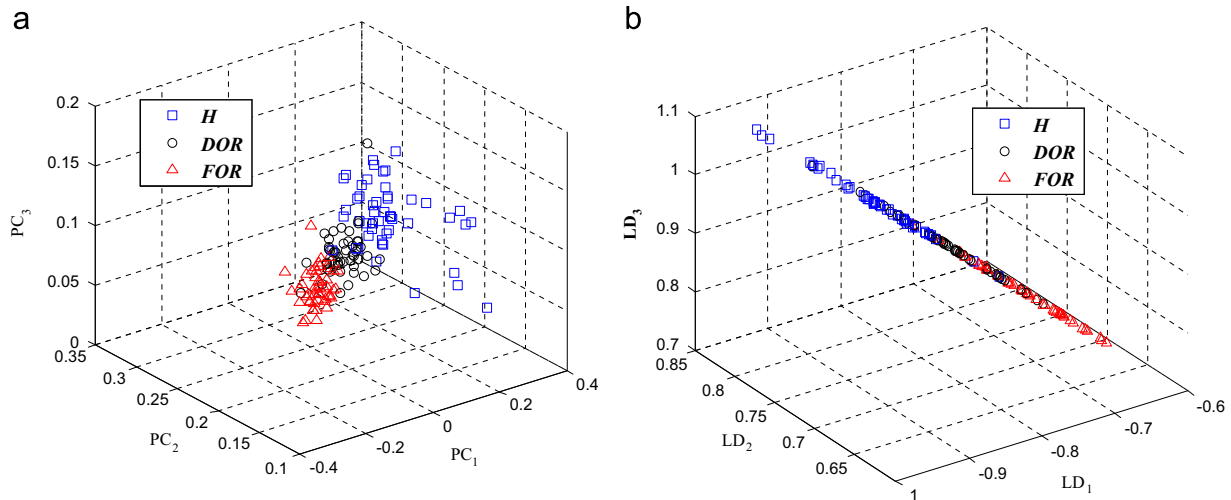


Fig. 10. Distribution of the best original features: (a) obtained from PCA and (b) obtained from LDA.

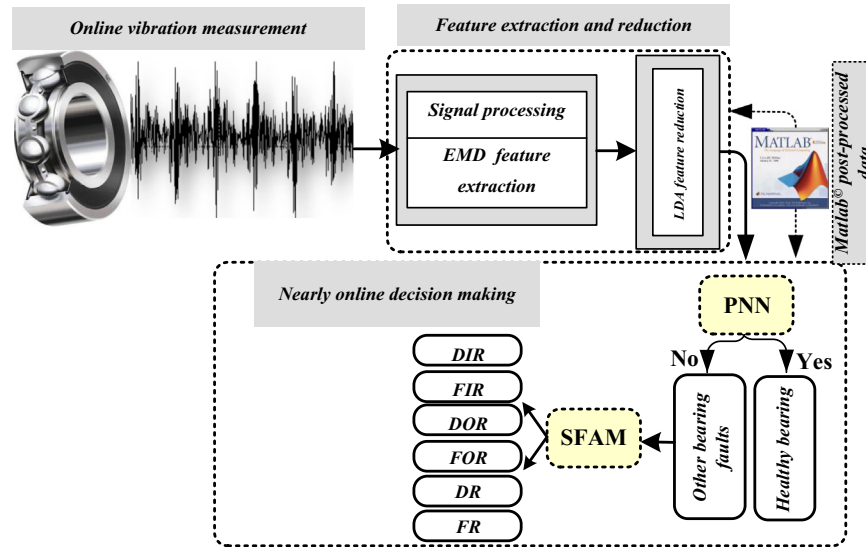


Fig. 11. General synoptic of the selected EMD-LDA-PNN-SFAM methodology.

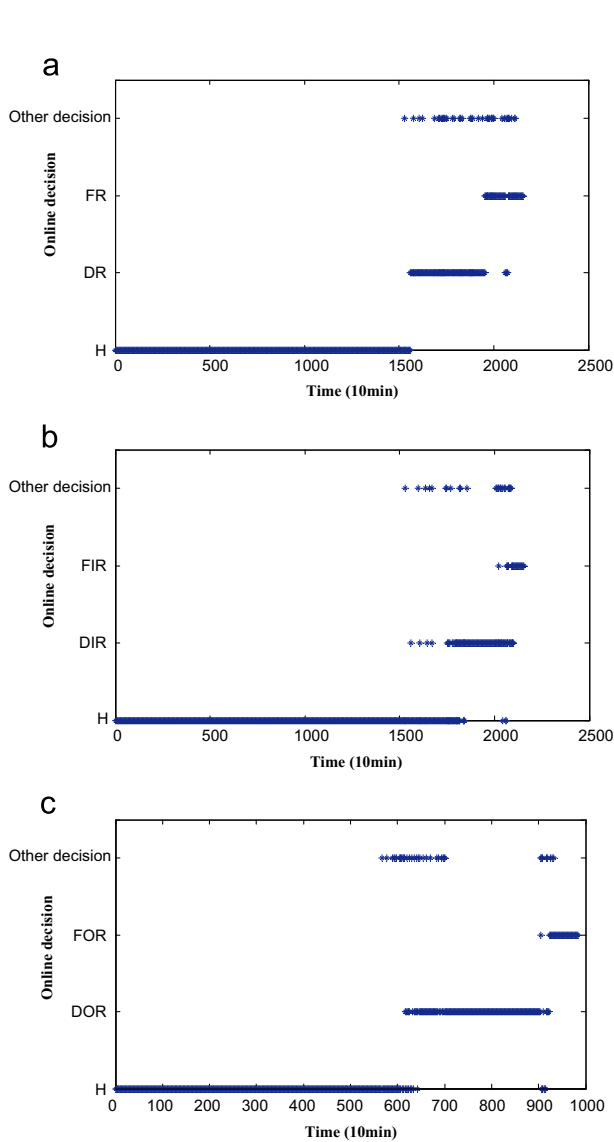


Fig. 12. Application of the proposed method for an online automatic decision making, (a) bearing 4 of testing 1, (b) bearing 3 of testing 1 and (c) bearing 1 of testing 2.

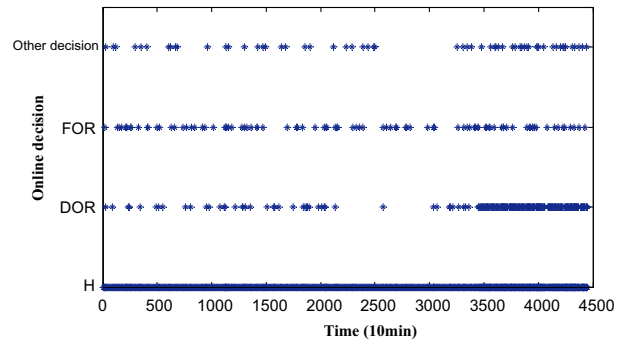


Fig. 13. Application of the proposed method for an online diagnosis of bearing 3 of testing 3.

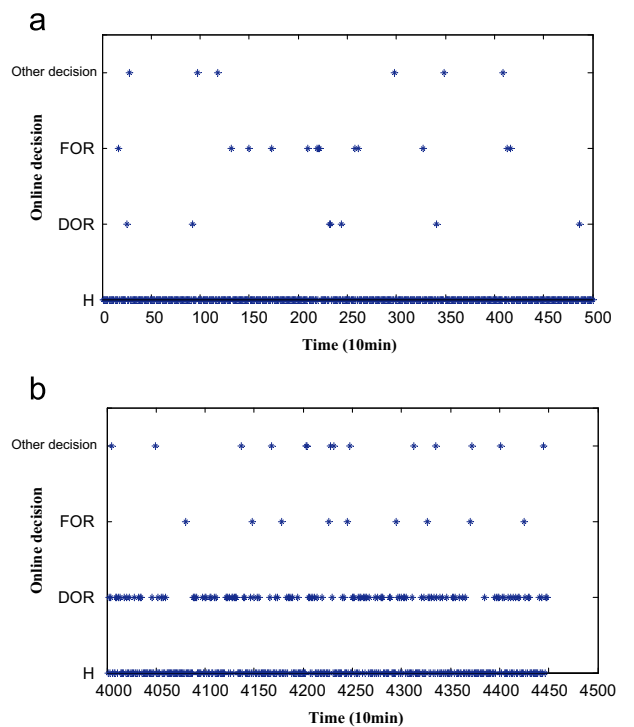


Fig. 14. Zoom of Fig. 13 in the area between, (a) 0 and 500 (unit of time) and (b) 4000 and 4500 (unit of time).

Fig. 15 shows the adverse effect of the proposed smart filtering phase on the bearing 3 of testing 3. Compared to Fig. 14, Fig. 15 exhibits fewer false alarms, which can improve the classification task.

The first fifteen points in Fig. 15 are equal to zero (omitted) due to the sliding window length which is set to sixteen. Despite the fact that increasing window length decreases the false alarm rate, it increases the number of zero points and the computational time. Consequently, several tests have been performed in this work to select the most effective window length and to select the threshold of occurrences. By evaluating the percentage of false alarms, the best window length is chosen to ensure more sensitivity to the proposed method.

4. Discussion

4.1. Feature performance measures

The ultimate goal in pattern classification is to find the best feature dataset. It is a common practice to get a feature dataset with high performance measures. To achieve this goal, a mathematical treatment is presented in this section. From various measures known in literature, we discuss the interclass/intraclass distance well known as scatter matrix. This measure is applied to the multi-class case and it is well known as the J criterion. In this paper, the J criterion is used to evaluate original extracted features, PCA features and LDA features. As clarified previously, seven REB states are defined in this work. For each class, 50 extracts are used to be evaluated. Also, these features are projected with PCA and LDA to be evaluated. In this way seven classes have been used ($K=7$) and 50 extracts are evaluated ($N=50$). The procedure of the J criterion is summarized as follows (Ben Ali et al., 2013):

- i. Compute the average feature vectors of the k th class:

$$\underline{\mu}_k = \frac{1}{N} \sum_{n=1}^N x_{k,n} \quad (13)$$

where $x_{k,n}$ the n th extract of the k th class.

- ii. Compute the total average feature vectors of all classes

$$\underline{\mu}_c = \frac{1}{K} \sum_{k=1}^K \underline{\mu}_k \quad (14)$$

- iii. Compute the intra-class variance matrix of average dispersion coefficients

$$S_{\text{intra}} = \frac{1}{KN} \sum_{k=1}^K \sum_{n=1}^N (x_{k,n} - \underline{\mu}_k)(x_{k,n} - \underline{\mu}_k)^t \quad (15)$$

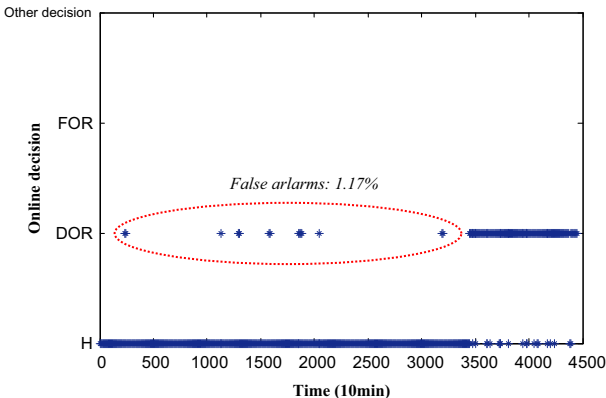


Fig. 15. Enhanced nearly online diagnosis of bearing 3 of testing 3 using a smart average technique.

- iv. Compute the inter-class variance of average dispersion between different classes

$$S_{\text{inter}} = \frac{1}{K} \sum_{k=1}^K (\underline{\mu}_k - \underline{\mu}_c)(\underline{\mu}_k - \underline{\mu}_c)^t \lim_{x \rightarrow \infty} \quad (16)$$

- v. Finally, the J degree of the average scattering within classes is given by

$$J = \text{trace}(S_{\text{intra}}^{-1} S_{\text{inter}}) \quad (17)$$

The computation of the J criterion based on original features, PCA features and LDA features is given respectively in Table 4, Tables 5 and 6. These three tables improve the importance of the LDA feature reduction technique. First, the use of LDA reduces the 13 original features to 6 features which decrease the training computational time of PNN and SFAM. In addition, the first component LD_1 shows the greatest scattering criterion compared to all features. Besides, the scattering criterion of all LDA features is more discriminant compared to original and PCA features. In conclusion, the comparison of the six proposed scenarios confirms the important of EMD–LDA–PNN–SFAM combination.

4.2. Comparison with some previous works

Classification accuracy is much more improved for all bearing diagnosis conditions by the use of methodologies based on feature extraction, feature selection (or reduction) and classification tools. The classification accuracy results are all greater than 92%. To get a good classification, the majority of scientific researches in the previous works have used four bearing states (H, FIR, FOR and FR). The superiority of this work is highlighted in the introduction of the default severity which was not held before. Severity consists of three types: normal, degraded and faulty. Thereby, seven bearing states have been used in our work. These added modifications allow the real time monitoring degradation. Table 7 summarizes previous works on automated identification of bearing faults.

The importance of our work over others is the use of real data which results from real time vibrations describing the dynamic response of defective REBs. However, previous works are based on simulation defects by creating holes with different diameters and depths. The proposed approach produces good classification accuracy despite the important number of classes and records. Moreover, our proposed methodology is tested online however the majority of the previous works do not satisfy this aim and scope.

Compared to the work proposed in Ben Ali et al. (2015b), this paper presents a strategic improvement. As shown in Table 7, the number of features used in Ben Ali et al. (2015b) was 18 and the used classifier was the back propagation neural network (which is

Table 4
 J criterion of original features.

Features	J
H_{en}	0.6446
$H_{en}IMF_1$	0.4052
$H_{en}IMF_2$	0.3163
$H_{en}IMF_3$	0.3560
$H_{en}IMF_4$	1.3972
$H_{en}IMF_5$	0.7359
$H_{en}IMF_6$	0.1853
$H_{en}IMF_7$	0.1357
$H_{en}IMF_8$	0.0774
$H_{en}IMF_G (G > 8)$	≈ 0
All features	3.3292

perhaps, with respect, even a bit outdated). The structure of the used neural network was important (four layers and 60 nodes) which increase remarkably the computational time. Besides, the use of statistical features certainly degrades the quality classification: it is well known that they are redundant and consequently we do not need all of them as they provide the same information (Tandon and Choudhury, 1999). Also, statistical features present strange behavior: their values come down to the level of an undamaged bearing when the damage is well advanced (Tandon and Choudhury, 1999). In this paper, we have applied the LDA technique to select best features derived from the EMD method and guarantee an optimal separation between features of different classes. Moreover, we have proposed a new combination of PNN and SFAM neural networks. This combination is very promising and it is through her that we obtained 97.82% as classification accuracy. The combination of PNN and SFAM do not require structure initialization, however, the back propagation neural network needs to be compiled many times to confirm a structure which can be not optimal. In the literature there are some

empirical recommendations to define the structure of the back propagation neural network, however, no mathematical justification was given (Capocchi et al., 2011).

Although that classification accuracy, of the previous works given in Table 7, was very satisfactory, these works were based on synthetic bearing defects which were always introduced by the user (holes with different diameters, depths and numbers). In reality, this is not the case of a bearing defect in industrial environment. Many factors can cause deterioration of bearings such as

- Contamination and corrosion.
- Lack of lubrication causing heating and abrasion.
- Defect of bearing's mounting, by improperly forcing the bearing onto the shaft.
- Misalignment defects.

REB degradation is a highly nonlinear phenomenon; using the same type of bearing in the same experimental conditions never produces the same kind of failure in terms of time, type, and severity. This property is confirmed by Intelligent Maintenance Systems (IMS) bearing run-to-failure data set.

Relatively, only few papers discussed the bearing diagnosis based on run-to-failure vibration signals. This is due to the difficulties in processing these signals. The major problem is the noise; it is very difficult to detect degradation when it is smaller than the noise measurements. This form of degradation is often known as “naturally progressing degradations”. Even though many researchers have performed fault bearing detection, they do not perform the diagnosis and identification of naturally progressing degradations.

The twelve used bearings in this work confirm that bearing degradation is a highly nonlinear phenomenon. The failure time and the failure nature of the three broken bearings were not the same. All these confirm that bearing fault assessment is a very hard task. In conclusion, the proposed method is very adequate in terms of fault accuracy detection.

4.3. Statistical validation

The confusion matrix is a useful tool for analyzing how well the classifier can recognize tuples of different groups, which contains information about real and predicted classifications done by a classification system (Fawcett, 2001).

From the confusion matrix of the proposed method given in Table 8 and presented in Fig. 16 by a compact box plot, it can be seen that the EMD–LDA–PNN–SFAM combination recognize the type and the severity of defects effectively.

Table 5

J criterion of PCA features.

PCs	<i>J</i>
PC ₁	1.7677
PC ₂	0.8489
PC ₃	0.8242
PC ₄	0.7149
PC ₅	0.4005
PC ₆	0.3445
PC ₇	0.34
PC ₈	0.2925
All features	5.6205

Table 6

J criterion of LDA features.

LDs	<i>J</i>
LD ₁	3.8574
LD ₂	0.3629
LD ₃	0.228
LD ₄	0.3539
LD ₅	0.5555
LD ₆	0.0513
All features	6.8044

Table 7

Summary of some studies reporting bearing fault diagnosis.

Literature	Features/techniques used	Classifier	REB classes	Accuracy [%]	Online diagnosis
Gryllias and Antoniadis (2012)	Frequency domain based vibration energy features.	Support Vector Machine (SVM)	3: H, FIR, FOR.	100	No
Yu et al. (2006)	EMD energy entropy of the first eight IMFs	Artificial neural networks (ANN)	3: H, FIR, FOR	93	No
Liu et al. (2013)	EMD features extraction in time and frequency domains	Particle swarm optimization (PSO) and SVM	4: H, FIR, FOR, FR.	97.5	No
Xu and Chen (2013)	EMD energy entropy of the first six IMFs	Improved PSO and least squares SVM	3: H, FIR, FOR.	97	No
Zarei et al. (2014)	Removing non-bearing fault component (RNFC) filter	ANN	4: H, FIR, FOR (one hole), FOR (double holes).	97.66	No
Pandya et al. (2013)	Hilbert Huang transform (HHT) based on EMD	Asymmetric proximity function and K-nearest neighbor	5: H, FIR, FOR, FR, combined defect.	96.67	No
Ben Ali et al. (2015b)	EMD energy entropy of the first seven IMFs and classical statistical features	ANN	7: H, DIR, DOR, DR, FIR, FOR, FR.	93	Yes
This work	EMD energy entropy and LDA	PNN and SFAM	7: H, DIR, DOR, DR, FIR, FOR, FR.	97.82	Yes

For more statistical significance and better generalization, some series of experiments should be conducted to support the author's claims. In fault detection problems, the performance of a detection algorithm usually depends on the trade-off between robustness and sensitivity. The sensitivity and robustness of the proposed method need to be explored by running a series of experiments. A receiver operating characteristic (ROC) curve will make the results more convincing. The ROC curves are often the only valid method of evaluation (Fawcett, 2006). An ROC curve is a detection performance evaluation methodology that shows how successfully a certain detector can separate two groups in a quantitative manner (Fawcett, 2001). An ROC curve gives the trade-off between the probability of detection or true positives rate, also called sensitivity and recall versus the probability of false alarm or false positives rate (Fawcett, 2006).

In order to compute the performance of the developed diagnosis methodology, the ten folds cross-validation experiments mentioned in Section 3.3 are re-used in this validation phase. Based on the correct classification of patterns in the test set (in each fold), the performance measures are computed. The average of these performance measures is computed to report the overall performance of the proposed method. To handle the problem of multiple classes, we used the one-against-all strategy: Each class was treated as appositive class in turn, while all remaining six classes were combined to form a negative class.

In each fold, based on correct classification of test patterns, the following parameters are defined:

- True Positive (TP): It is the number of actually abnormal test patterns which were classified as abnormal by our automated classifiers.
- True Negative (TN): It is the number of normal patterns in the test set which were classified into normal class by the automated classifiers.
- False Positive (FP): It is the number of normal test patterns classified as abnormal by the automated classifiers.
- False Negative (FN): It is the number of abnormal test patterns which are classified as normal by the automated classifiers.

Based on TP, TN, FP and FN, the sensitivity (Sn), and specificity (Sp), were defined as follows:

$$Sn = \frac{TP}{TP + FN} \quad (18)$$

$$Sp = \frac{TN}{TN + FP} \quad (19)$$

Fig. 17 shows one of the found ROC performances of our proposed strategy. It can be seen that PNN–SFAM classifier provides good accuracies of sensitivity (91.7%) and specificity (93.4%) and an area under curves equal to 0.847. We should note

that this curve has the least important features among all ROC curves found which validates the strength of our strategy.

5. Conclusion and prospects

In this work, a new online automated bearing fault diagnosis methodology is proposed. This effective method is introduced by feature extraction, characterization, reduction and classification based on non-stationary vibration signals. Compared to previous works, experimental results demonstrate that the proposed method is extremely promising and it can be summarized as follows:

- EMD is a self-adaptive signal processing method applied to decompose nonlinear and non-stationary vibration signals into some stationary IMFs.
- The diagnosis method is based on eigenvalues of LDA method to investigate a higher classification accuracy compared to the use of EMD energy entropy and PCA features.
- The combination of PNN and SFAM neural networks presents a better categorization than the use of SFAM only.

Finally, the proposed method accurately identifies bearing faults online and a better generalization capability is shown compared to previous works. As future works, we will focus on the variable operating conditions of speed and charge which can generalize the proposed methodology.

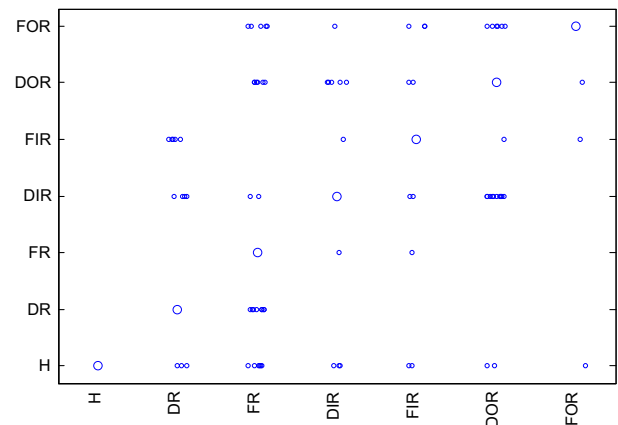


Fig. 16. Compact box plot of the confusion matrix.

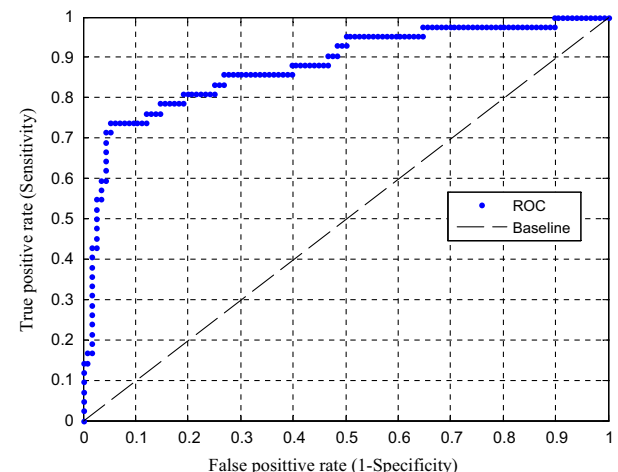


Fig. 17. Validation of the proposed method using ROC curves.

Table 8
Confusion matrix of the proposed method.

Real	Estimated						
	H	DR	FR	DIR	FIR	DOR	FOR
H	2900	0	0	0	0	0	0
DR	3	139	0	4	5	0	0
FR	6	8	423	1	1	7	5
DIR	3	0	1	140	2	4	1
FIR	2	0	1	2	46	2	3
DOR	2	0	0	12	1	130	6
FOR	1	0	0	0	1	1	48

Acknowledgment

Authors wish to thank the Center for Intelligent Maintenance Systems (IMS), University of Cincinnati, USA, for his permission to use his bearing data. Also, authors wish to thank Dr. Barrie William Jervis, former professor in Sheffield Hallam University (UK) for his relevant advices during the implementation of PNN and SFAM neural network.

References

- Ben Ali, J., Sayadi, M., Fnaiech, F., Morello, B., Zerhouni, N., 2013. Importance of the fourth and fifth intrinsic mode functions for bearing fault diagnosis. In: Proceedings of the 14th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), pp. 259–264.
- Ben Ali, J., Abid, S., Jervis, B.W., Fnaiech, F., Bigan, C., Besleaga, M., 2014. Identification of early stage Alzheimer's disease using SFAM neural network. *Neurocomputing* 143, 170–181.
- Ben Ali, J., Chebel-Morello, B., Saidi, L., Malinowski, S., Fnaiech, F., 2015a. Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network. *Mech. Syst. Signal Process.* 56–57, 150–172.
- Ben Ali, J., Fnaiech, N., Saidi, L., Chebel-Morello, B., Fnaiech, F., 2015b. Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Appl. Acoust.* 89, 16–27.
- Bellini, A., Filippetti, F., Tassoni, C., Capolino, G.-A., 2008. Advances in diagnostic techniques for induction machines. *IEEE Trans. Ind. Electron.* 55, 4109–4126.
- Bin, G.F., Gao, J.J., Li, X.J., Dhillon, B.S., 2012. Early fault diagnosis of rotating machinery based on wavelet packets-empirical mode decomposition feature extraction and neural network. *Mech. Syst. Signal Process.* 27, 696–711.
- Braun, S., Feldman, M., 2011. Decomposition of non-stationary signals into varying times scales: some aspects of the EMD and HVD methods. *Mech. Syst. Signal Process.* 25, 2608–2630.
- Capocchi, L., Toma, S., Capolino, G.A., Fnaiech, F., Yazidi, A., 2011. Wound-rotor induction generator short-circuit fault classification using a new neural network based on digital data. In: Proceedings of the IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics & Drives (SDMPED), Bologna, pp. 638–644.
- Fawcett, T., 2001. Using rule sets to maximize ROC performance. In: Proceedings of International Conference of Data Mining, pp. 131–138.
- Fawcett, T., 2006. An Introduction to ROC analysis. *Pattern Recognit. Lett.*, 861–874.
- Frosini, L., Bassi, E., 2010. Stator current and motor efficiency as indicators for different types of bearing faults in induction motors. *IEEE Trans. Ind. Electron.* 57, 244–251.
- 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. *Eng. Appl. Artif. Intell.* 25, 326–344.
- Hong, S., Zhou, Z., Zio, E., Wang, W., 2014. An adaptive method for health trend prediction of rotating bearings. *Digit. Signal Process.* 35, 117–123.
- Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., Liu, H.H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. A* 454, 903–995.
- Huang, J., Hu, X., Geng, X., 2011. An intelligent fault diagnosis method of high voltage circuit breaker based on improved EMD energy entropy and multi-class support vector machine. *Electr. Power Syst. Res.* 81, 400–407.
- Konar, P., Chattopadhyay, P., 2011. Bearing fault detection of induction motor using wavelet and Support Vector Machines (SVMs). *Appl. Soft Comput.* 11, 4203–4211.
- Lau, C.K., Ghosh, K., Hussain, M.A., Hassan, C.R.C., 2013. Fault diagnosis of Tennessee Eastman process with multi-scale PCA and ANFIS. *Chemom. Intell. Lab. Syst.* 120, 1–14.
- Lei, Y., He, Z., Zi, Y., Hu, Q., 2007. Fault diagnosis of rotating machinery based on multiple ANFIS combination with gas. *Mech. Syst. Signal Process.* 21, 2280–2294.
- Lei, Y., Lin, J., He, Z., Zuo, M.J., 2013. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mech. Syst. Signal Process.* 35, 108–126.
- Li, B.W., Zhang, Y., 2011. Supervised locally linear embedding projection for machinery fault diagnosis. *Mech. Syst. Signal Process.* 25, 3125–3134.
- Lin, J., 2010. Fault diagnosis of natural gas compressor based on EEMD and Hilbert marginal spectrum. In: Proceedings of the International Conference on Information Science and Engineering, pp. 3701–3704.
- Liu, X., Bo, L., Luo, H., 2015. Bearing faults diagnostics based on hybrid LS-SVM and EMD method. *Measurement* 59, 145–166.
- Liu, Z., Cao, H., Chen, X., He, Z., Shen, Z., 2013. Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings. *Neurocomputing* 99, 399–410.
- Loutas, T.H., Roulias, D., Pauly, E., Kostopoulos, V., 2011. The combined use of vibration, acoustic emission and oil debris on-line monitoring towards a more effective condition monitoring of rotating machine. *Mech. Syst. Signal Process.* 25, 1339–1352.
- Mouelhi, A., Sayadi, M., Fnaiech, F., Mrad, K., Romdhane, K.B., 2013. Automatic image segmentation of nuclear stained breast tissue sections using color active contour model and an improved watershed method. *Biomed. Signal Process. Control* 8, 421–436.
- Nandi, S., Toliyat, H.A., Xiaodong, L., 2005. Condition monitoring and fault diagnosis of electrical motors—a review. *IEEE Trans. Energy Convers.* 20, 719–729.
- Pan, Y., Er, M.J., Li, X., Yu, H., Gouriveau, R., 2014. Machine health condition prediction via online dynamic fuzzy neural networks. *Eng. Appl. Artif. Intell.* 35, 105–113.
- Pandya, D.H., Upadhyay, S.H., Harsha, S.P., 2013. Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN. *Expert Syst. Appl.* 40, 4137–4145.
- Priestley, M.B., 1988. *Nonlinear and Nonstationary Time Series Analysis*. Academic Press, New York, pp. 11–31.
- Randall, R.B., Antoni, J., 2011. Rolling element bearing diagnostics—a review. *Mech. Syst. Signal Process.* 25, 485–520.
- Qiu, H., Lee, J., Lin, J., Yu, G., 2006. Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *J. Sound Vib.* 289, 1066–1090.
- Saidi, L., Ben Ali, J., Fnaiech, F., 2014. Bi-spectrum based-EMD applied to the non-stationary vibration signals for bearing faults diagnosis. *ISA Trans.* 53, 1650–1660.
- Shakya, P., Kulkarni, M.S., Darpe, A.K., 2014. A novel methodology for online detection of bearing health status for naturally progressing defect. *J. Sound Vib.* 333, 5614–5629.
- Seungdeog, C., Akin, B., Rahimian, M.M., Toliyat, H.A., 2011. Implementation of a fault-diagnosis algorithm for induction machines based on advanced digital-signal-processing techniques. *IEEE Trans. Ind. Electron.* 58, 937–948.
- Specht, D.F., 1990. Probabilistic neural networks. *Neural Netw.* 3, 109–118.
- Subasi, A., Gursay, M.I., 2010. EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Syst. Appl.* 37, 8659–8666.
- Tandon, N., Choudhury, A., 1999. A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. *Tribol. Int.* 32, 469–480.
- Tavakkoli, F., Teshnehlab, M., 2007. A ball bearing fault diagnosis method based on wavelet and EMD energy entropy mean. In: Proceedings of the International Conference on Intelligent and Advanced Systems, pp. 1210–1212.
- Venkatesan, P., Suresh, M., 2009. Classification of renal failure using simplified fuzzy adaptive resonance theory map. *Int. J. Comput. Sci. Netw. Secur.* 9, 129–134.
- Wang, H., Chen, J., Dong, G., 2014. Feature extraction of rolling bearing's early weak fault based on EEMD and tunable Q-factor wavelet transform. *Mech. Syst. Signal Process.* 48, 103–119.
- Wang, Y., Xu, G., Liang, L., Jiang, K., 2015. Detection of weak transient signals based on wavelet packet transform and manifold learning for rolling element bearing fault diagnosis. *Mech. Syst. Signal Process.* 54–55, 259–276.
- Widodo, A., Yang, B.-S., 2007. Support vector machine in machine condition monitoring and fault diagnosis. *Mech. Syst. Signal Process.* 21, 2560–2574.
- Widodo, A., Yang, B.-S., Han, T., 2007. Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors. *Expert Syst. Appl.* 32, 299–312.
- Xu, H., Chen, G., 2013. An intelligent fault identification method of rolling bearings based on LSSVM optimized by improved PSO. *Mech. Syst. Signal Process.* 35, 167–175.
- Yang, J., Zhang, Y., Zhu, Y., 2007. Intelligent fault diagnosis of rolling element bearing based on SVMs and fractal dimension. *Mech. Syst. Signal Process.* 21, 2012–2024.
- Yu, Y., Yu, D., Junsheng, C., 2006. A roller bearing fault diagnosis method based on EMD energy entropy and ANN. *J. Sound Vib.* 294, 269–277.
- Zarei, J., Tajeddini, M.A., Karimi, H.R., 2014. Vibration analysis for bearing fault detection and classification using an intelligent filter. *Mechatronics* 24, 151–157.
- Zhang, Y., Randall, R.B., 2009. Rolling element bearing fault diagnosis based on the combination of genetic algorithms and fast kurtogram. *Mech. Syst. Signal Process.* 23, 1509–1517.
- Zheng, J., Cheng, J., Yang, Y., 2013a. A rolling bearing fault diagnosis approach based on LCD and fuzzy entropy. *Mech. Mach. Theory* 70, 441–453.
- Zheng, J., Cheng, J., Yang, Y., 2013b. Generalized empirical mode decomposition and its applications to rolling element bearing fault diagnosis. *Mech. Syst. Signal Process.* 40, 136–153.
- Žvokelj, M., Zupan, S., Prebil, I., 2011. Non-linear multivariate and multiscale monitoring and signal denoising strategy using Kernel Principal Component Analysis combined with Ensemble Empirical Mode Decomposition method. *Mech. Syst. Signal Process.* 25, 2631–2653.