

# Bearing Faults Detection by a Novel Condition Monitoring Scheme based on Statistical-Time Features and Neural Networks

Miguel Delgado, *Student Member, IEEE*, Giansalvo Cirrincione, *Member, IEEE*, Antonio Garcia Espinosa, *Member, IEEE*, Juan Antonio Ortega, *Member, IEEE*, Humberto Henao, *Senior Member, IEEE*

**Abstract** – Bearing degradation is the most common source of faults in electrical machines. In this context this work presents a novel monitoring scheme applied to diagnose bearing faults. Apart from detecting local defects, i.e. single point balls and raceways faults, it takes also into account the detection of distributed defects, such as roughness. The development of diagnosis methodologies considering both kind of bearing faults is, nowadays, subject of concern in fault diagnosis of electrical machines. First, the method analysis the most significant statistical-time features calculated from vibration signal. Then it uses a variant of the Curvilinear Component Analysis, a nonlinear manifold learning technique, for compression and visualization of the features behavior. It allows interpreting the underlying physical phenomenon. This technique has demonstrated to be a very powerful and promising tool in the diagnosis area. Finally, a hierarchical Neural Network structure is used to perform the classification stage. The effectiveness of this condition monitoring scheme has been verified by experimental results obtained from different operation conditions.

**Index Terms**— Ball bearings, Classification algorithms, Condition monitoring, Fault diagnosis, Feature extraction, Induction motors, Neural networks, Vibrations.

## NOMENCLATURE

$\Sigma$	Covariance matrix.
$D$	Data space dimension.
$d$	Latent space dimension.
$D_{ij}$	Euclidian distance between $i$ and $j$ points in data space.
$k$	Number of neighbors considered.
$L_{ij}$	Euclidian distance between $i$ and $j$ points in latent space.
$N$	Number of measurements in a data set.
$p$	Vector of operating conditions.
$X_i^l$	The $i$ -th data point from a $l$ set.
$Y_i^l$	The $i$ -th projection from a $l$ set.
$\alpha$	Learning rate.

$\lambda$	Neighborhood radius.
$\mathcal{R}^D$	$D$ dimensional feature space.
$x_i$	The $i$ -th vector of features in data space.
$y_i$	The $i$ -th vector of features in latent space.
$w_i$	Weighting factor.

## I. INTRODUCTION

THE reliability of electrical drives is an extensively investigated field for cost and maintenance savings, but also because a safe operation is desired in critical applications [1]-[4]. It is well known that the bearings represent one of the most common sources of faults in electromechanical systems [5]-[8]. Bearing defects have been categorized as local (due to cracks, pits and spalls, basically) and distributed (including surface roughness, waviness and misaligned races) [9], [10]. Localized bearing faults are classified by the fault specific location in: inner race, outer race and ball. Conversely, in generalized roughness fault, the bearing surface has been degraded considerably over a large area and becomes rough, irregular, or deformed [11].

Most of the bearings monitoring schemes are focused on bearing localized defects [12], [13]. They are based on the detection of some characteristic fault harmonic components of the vibration spectra [14]. However, apart from possible electrical and mechanical noise during the acquisition, it will be affected by the intrinsic vibration modes of the system, and the bearing deterioration stage. Besides, the absence of clear characteristic fault frequencies should not be interpreted as a completely healthy condition of the bearing. The analysis of the classical characteristic fault frequencies is not a simplistic matter since the basic diagnosis schemes may leads to a delayed diagnosis until the characteristic fault frequencies have enough presence in the spectra to be clearly localized. On the other hand, the generalized roughness faults produce unpredictable broadband effects which are not necessarily related with specific fault frequencies [15]. However, these faults are common in industry, while they are often neglected in the research literature.

Afterwards, advanced signal processing techniques, such as probabilistic models [16], high-resolution frequency analysis [6] or enhanced wavelets decompositions [17], [18], applied over the measured physical magnitude have been used to obtain reliable fault indicators. However, most of these approaches cannot deal with the identification of single and generalized bearing faults at the same time, which makes the development of a whole bearing fault diagnosis system still an open point. In this sense, over the last years, instead of looking for highly significative features, a trend towards the fusion of different features to

Manuscript received April 22, 2012; revised June 7, 2012; accepted August 26, 2012. This work was supported in part by the Spanish Ministry of Economy and Competitiveness under the TRA2010-21598-C02-01 Research Project. Copyright © 2012 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

M. Delgado and J. A. Ortega are with the department of Electronic Engineering, Technical University of Catalonia (UPC), MCIA research center, Rbla. San Nebridi s/n, 08222 Terrassa, Spain (phone: +34-93-739-8518; fax: +34-93-739-8972; e-mails: miguel.delgado@mcia.upc.edu, juan.antonio.ortega@mcia.upc.edu).

A. Garcia is with the department of Electrical Engineering, Technical University of Catalonia (UPC), MCIA research center, Rbla. San Nebridi s/n, 08222 Terrassa, Spain (e-mail: antoni.garcia@mcia.upc.edu).

G. Cirrincione and H. Henao are with the department of Electrical Engineering, University of Picardie (UPJV), LTI 7 Rue du Moulin Neuf, 80000 Amiens, France (e-mails: giansalvo.cirrincione@u-picardie.fr, humberto.henao@u-picardie.fr).

enhance the performance of the diagnosis system have been carried out. In order to analyze and manage the significance and relations between the features, advanced techniques based on artificial intelligent are used [19]-[24].

In this sense a general diagnosis methodology based on pattern recognition scheme is composed by three blocks as follows. Firstly, the calculation of numerical features from the acquired physical measurement, secondly a feature reduction procedure to highlight the hidden patterns and to compress the information and, finally, a classification stage in which the different classes (considered degradation types) are identified. The feature reduction represents the most critical stage in this diagnosis process. Classification systems fed by raw vectors of features will decrease its performance, while an inaccurate reduction may remove useful information. In this way, the feature reduction process has been typically implemented with linear techniques such as Principal Component Analysis (PCA). However, PCA techniques has been discussed by many authors emphasizing its limitation dealing with large data sets, because it seeks for a global structure of the data [20], [25].

A feature vector is formed by the  $D$  calculated features and these vectors are represented in a  $D$ -dimensional space. The information contained in such a  $D$ -space mostly has a nonlinear structure. Concerning with this problem manifold learning methods has been applied in the last years [26], [27] to preserve this information in a lower  $d$ -dimensional space, where  $D > d$ . Among them, Self-Organizing Maps (SOM) is the most used, which is based on developing a neural network grid to preserve most of the original distances between feature vectors representations in the  $D$ -space.

This procedure, although exhibits generally good performances, requires to fix an initial shape of the neural network. That is, some idea of the data behavior in the  $d$ -dimensional space should be known *a priori*. This fact may present convergence problems, difficulties to unfold high nonlinearities and large computational time, which make them very difficult to use with high-dimensional data sets.

Other approaches, as Curvilinear Component Analysis (CCA), automatically “finds” the correct shape of the sub-manifold, that is, the neurons “search” a proper position in the output space preserving as much as possible the original feature space distances. For that reason, CCA has been applied in different fields: initially to pattern recognition [28], after in electrical power systems [29] and, finally this work introduce the technique in the electromechanical systems diagnosis field.

The originality of this work includes a complete fault analysis and diagnosis methodology applied to detect different bearing faults including localized and generalized defects. The method begins with the selection of the most significant features from an initial set formed by statistical-time features calculated from vibration signals. Next, it has implemented a new set of improvements in the CCA with the objective that it can be applied in fault classification structures, in such a way that the feature reduction stage is performed. In this context, this implemented CCA allows the data visualization and interpretation of the underlying physical phenomenon. Finally, the classification stage is proposed to be solved by a hierarchical neural network. This work represents an important step to the introduction of manifold learning techniques, and advanced

classification structures, to the development of electromechanical system diagnosis procedures, being the first time that it has been applied.

## II. CCA IMPLEMENTATION FOR FAULT CLASSIFICATION STRUCTURES

As it has been mentioned, one of the most novel strategies of nonlinear feature extraction is based on distance preservation concept. For every pair of different features vectors in the original feature space (data space), a between-points distance  $D_{ij}$ , is calculated,  $D_{ij} = \|x_i - x_j\|$ . The objective is to preserve these distances between the same points in the reduced feature space (latent space),  $L_{ij} = \|y_i - y_j\|$ , formed by a reduced set of features. The perfect projection of the feature vectors in the latent space is only possible if the original feature set contains redundant features which can be removed without negative effect from the loss of information point of view. Otherwise, some information will be lost. In order to face this problem the CCA technique defines a distance function threshold in order to determine short and long distances between feature vectors,  $D_{ij}$ . By this way, the CCA prioritizes the short distances, which means local distance preservation. The CCA used in this study considers the distance function threshold as a decreasing exponential function [30]; whose corresponding error function is the right-Bregman divergence [31]. Indeed, this function penalizes long distances and its asymmetry allows a better unfolding of data. The basic procedure of the CCA is shown schematically in Fig. 1.

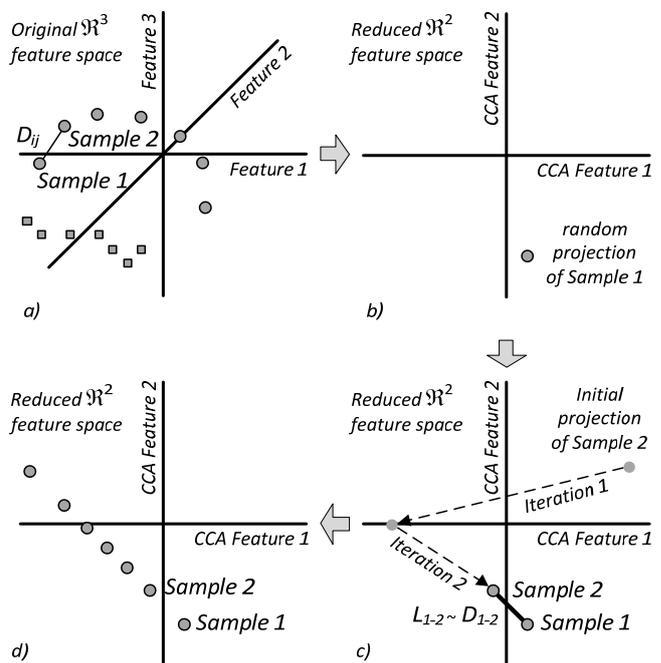


Fig. 1. CCA operation scheme sequence. (a) Seven feature vectors for each of the two operating conditions (circles and squares) represented in a three-dimensional data space. (b) CCA projection of the first feature vector of one operating condition (circles) in the latent space. (c) CCA projection of the second feature vector of the same operating condition (circles) in the latent space. Two iterations are represented until reach  $L_{1,2} \sim D_{1,2}$ . (d) Resultant CCA projection of the feature vectors corresponding to one operating condition (circles).

Although the projected topology in the latent space will exhibit the same performance, the global position of the projected map in the latent space changes at each new CCA execution, as it is represented in Fig. 2. That is, the CCA

projection is not invariant. Indeed, it changes because it is only constrained by the distance preservation. There are two sources of randomness: the projection of the first sample in the latent space, from which the rest of data will be projected, and the index sequence of the fixed samples.

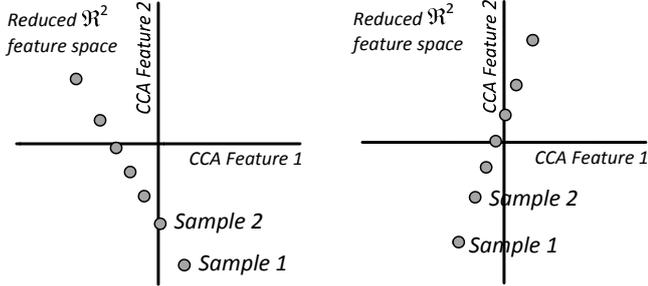


Fig. 2. Representation of two CCA executions over a same original set of feature vectors. The resulting two-dimensional projection are identical, but the global latent space position changes.

As a part of the proposed methodology, in order to maximize the performance of the CCA projection, one different CCA will be executed over each set of feature vectors corresponding to a different operating condition, i.e speed and torque. The quality of the projection is generally reduced while increasing the complexity of the data set distribution. Then, dividing the whole data set in different subsets implies a CCA projection improvement. Therefore, the CCA variability is a problem in case of classification, where the different regions defined in the latent space should be used to classify after new projections.

In order to achieve the CCA projection invariance, the following new developed procedure, consisting on three steps is presented.

- 1) The first CCA projection of samples corresponding to the first operating condition is executed.
- 2) The resulting projection of the previous CCA is used as initial projection of the next CCA in the latent space. The learning rate  $\alpha$  is lowered in order to have fewer variations during the iterations. This is applied consecutively until no more operating conditions are available. With this step, all the CCA projections share approximately the global latent space position.
- 3) Then, in order to match perfectly all the resulting CCA projection maps, a refinement (affine) transformation is applied.

Indeed, one of the resulting CCA projection maps is chosen as a rated CCA. For the estimation of the affine transformation, consider that for each data point  $X_i^l$ , two different  $d$ -dimensional projected points  $Y_i^{rated}$  and  $Y_i^l$  are obtained from the rated CCA and the  $p$ -th CCA respectively. Then, the entire data set  $l$  is transformed to the rated projection, by means of  $Y_i^{rated} = AY_i^l + b$  for each  $i$ , where is defined  $\theta = [(vecA)^T \ b^T]^T$  as the affine parameter vector containing the necessary  $A$  rotations and  $b$  translations as is schematically represented in Fig 3.

The affine transformation can be solved with success by ordinary least squares (OLS) because the differences between the different CCA projection maps references are small. Finally, the obtained rated CCA projection map allows developing a fault classification strategy common to all the range of operating conditions.

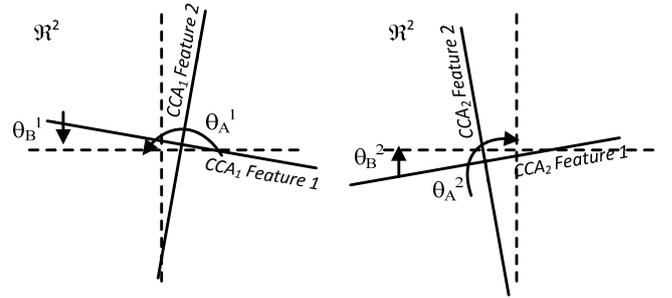


Fig. 3. Representation of the refinement transformation applied to two resulting CCAs projection maps (solid axes) to the rated CCA (dotted axes). Slight rotations and translations are applied by the affine parameters  $\theta_A$  and  $\theta_B$  respectively.

If it is required to project a new input, a recall phase is needed, which depends on the operating conditions. That is the new input will be projected using the corresponding CCA projection map, and then the corresponding affine transformation  $\theta$  will be applied to obtain its representation in the rated CCA map. However, this approach is acceptable only if new inputs match to one of the considered operating conditions. It is necessary, for a practical industrial application, to propose a solution to take into account whatever operating condition comprised between the considered operating conditions range. This is carried out by means of a nonlinear interpolation as follows. The nearest CCAs projection maps (regarding the operating conditions) are used. This fact implies that the distances of the new feature vector regarding the nearest sets of features vectors will be preserved. However, the weighting factor will prioritize to maintain the distances with the feature vectors corresponding to the nearest operating conditions. The proposed factor uses Gaussian probability density functions, as is represented as:

$$w_l(p_{new}) = \frac{e^{-(p_{new}-p_l)^T \sum (p_{new}-p_l)}}{\sum_{j=1}^k e^{-(p_{new}-p_l)^T \sum (p_{new}-p_l)}} \quad (1)$$

The  $k$  parameter yields the problem-dependent number of neighbor CCA's whose working conditions are close to  $p_{new}$  between all the CCAs (in Fig. 4, four neighbors). That is, choosing  $k$  means deciding how many CCAs are used in the recall phase. Therefore, when the operating condition point is too far, the associated projection is given less importance by decreasing the weight factor.

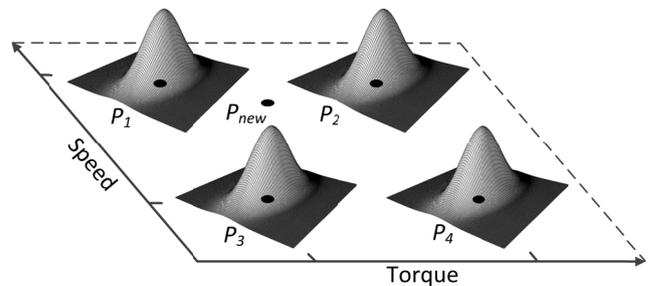


Fig. 4. Representation of the weighting factor between four operating conditions ( $p_1$ ,  $p_2$ ,  $p_3$  and  $p_4$ ) and the new one  $p_{new}$ . With  $k=4$ , the four nearest CCA projections will be used to project the new sample/s under  $p_{new}$ . Each of the CCA projections will be weighted by a weighting factor obtained from Gaussian probability density functions as is represented.

Therefore, the projection process is carried out by the successive application of the next stochastic gradient algorithm:

$$Y_j^{pnew} \leftarrow Y_j^{pnew} \alpha w_l(p_{new}) \frac{L_{ij} - D_{ij}}{L_{ij}} e^{-\frac{L_{ij}}{\lambda}} (Y_j^{pnew} - Y_j^{l(k)}) \quad (2)$$

where sample  $i \neq j$ . This formula assumes that only the projected data from the  $k$  nearest neighbor CCA's are to be considered in training and recall. It must be noted that the proposed weight acts by multiplying (modulating) the learning rate  $\alpha$  of the CCA gradient algorithm. In the same way, in the recall phase, the parameters of the affine transformation  $\theta$  are estimated by interpolation, using a radial basis function (RBF) neural network. This approach is possible because all CCA frameworks are coherent, and then an interpolation of the affine transformations is possible. The final projection of the new sample respects the global latent space positions of the rated CCA projection, and therefore can be applied in classification structures.

### III. DIAGNOSIS METHODOLOGY

It can be said that the diagnosis methodology is composed by four steps, first the calculation of the different features of the vibration signal, the feature calculation. Second feature selection, where only the most significant features are selected during the training process. Then the feature reduction, to compress and maintain the useful information for diagnosis purposes, and finally the classification stage based on neural network, wherein the different considered faults are diagnosed. The complete proposed diagnosis methodology is represented in Fig. 5.

In this work six bearing fault scenarios have been considered, namely: healthy ( $h$ ), inner race fault ( $i$ ), outer race fault ( $o$ ), ball fault ( $b$ ), inner-outer-ball faults at the same time ( $iob$ ), and generalized degradation in inner and outer races ( $gdi$ ). Moreover, twenty five different steady state operating conditions have been considered: from 80% to 100% of rated speed, and from 80% to 100% of rated torque every variation of 5% of each parameter. For each combination of bearing fault scenario and operating condition, twenty  $x$  and  $y$  vibration axis measurements (radial and axial axis) have been acquired respectively.

#### A. Features calculation

From each acquired vibration signal axis ( $x$  and  $y$  axis), a set of statistical-time features is computed. A total of 15 features from time-domain are proposed for each acquired vibration axis: mean, maximum value, root mean square, square root mean, standard deviation, variance, root mean square shape factor, square root mean shape factor, crest factor, latitude factor, impulse factor, skewness, kurtosis, normalized 5-th and 6-th moments.

#### B. Features Selection

The proposed features will contain a large portion of the information contained in the vibration signal; however only some of them will be really significant. These, in turns depend on the considered bearing defects, the apparition of additional sources of vibration, and the bearing location. In the latter case, be them either in the motor frame or externally mounted such as the case here presented, the selected features do not have to be the same. In this way the proposed methodology can be applied regardless where the bearings are installed.

Different techniques can be applied to analyze the features significance regarding the considered diagnosis scenario. The Discriminant Analysis (DA) is one of the classical techniques for a feature selection procedure.

The DA evaluates quantitatively the discriminant capabilities of the proposed features regarding the considered classes. A large value of DA implies that the analyzed feature/s contributes to a proper representation of the measurements in the data space. That is the classes are well delimited and well separated.

Every two and three combinations of the calculated features, as well as their individual capabilities have been evaluated, obtaining finally an ordered list with the most significance set of features. It has been observed, that although the  $x$  axis exhibits a bigger discriminant capability than the  $y$  axis, the most significant features in each axis correspond to the RMS, Variance and Shape factor. Then, a set of six features is proposed to define the considered classes in the data space.

#### C. Features Extraction

The projections of the six-dimensional vectors are computed by using CCA as it has been explained in Section II to avoid the CCA projection variability.

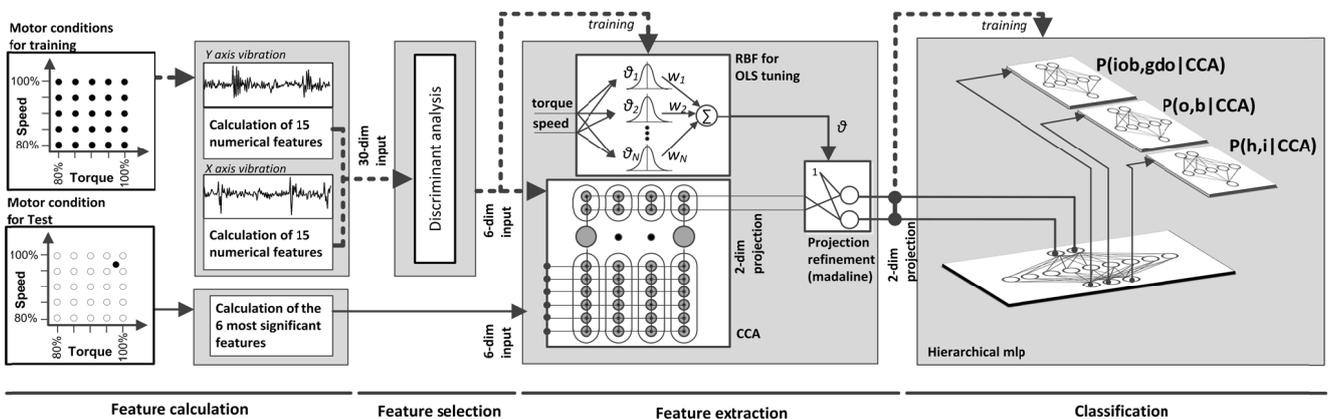


Fig. 5. Proposed diagnosis methodology scheme including feature calculation, feature selection, feature extraction and classification stages. The training set (dotted line) is used to select the most significant features during the feature selection stage, and train the RBFs, the CCAs and the hmlp. Then, the test set (solid line) is evaluated.

The initial learning rate (which decreases exponentially in time) is fixed to 0.5 for the selected rated CCA and 0.1 otherwise. Specifically, the rated CCA has been fixed as the one related with the rated working conditions (100% speed and 100% torque).

The refinement transformation follows. The RBF is trained with the 25 affine parameter vectors  $\theta$ , each one corresponding to each CCA. Therefore, the remaining variability in the rated CCA depends on the change of class position and shape under different values of the operating conditions, which is exploited in the next step (classification). In this sense the feature extraction transforms the original data base of six dimensional vectors into a data base of two-dimensional vectors, as it is shown in Fig. 6.

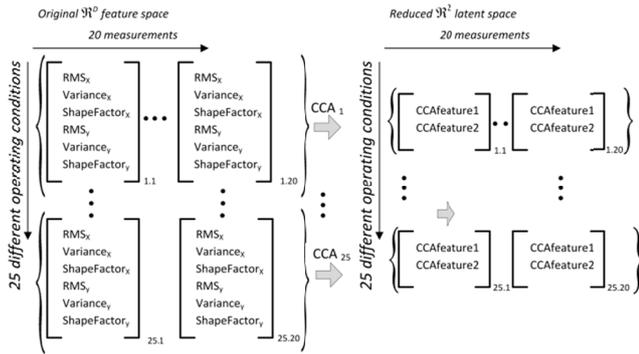


Fig. 6. Application of the feature reduction to the original feature space.

In the recall (test) phase, that is when data drawn under new working conditions have to be classified, the six-dimensional feature vector is fed to the corresponding  $k$  CCAs, and the corresponding torque and speed parameters are fed to the RBF in order to find the parameters of the affine transformation. A simple adaline neural network, composed by two adalines (madaline) [32], is used for transforming the new CCA projection to the rated CCA map. Then, the data is ready for classification.

#### D. Classification

Due to the different number of considered operating conditions (twenty-five), and the number of considered bearing fault scenarios (six), a two-level hierarchical neural network is applied to assure the optimum pattern recognition. Specifically, a hierarchy of multilayer perceptrons (hMLP) has been developed. This structure allows the classification in two steps: a first neural network classifies a two-dimensional feature vector (resulting from the CCA projection) between three predefined class pairs (in this application:  $h/i$ ,  $o/b$  and  $iob/gdio$ ), and enables three neural networks, each trained on a pair of classes. Once the input has been classified in the first neural network, the corresponding second neural network follows, and the bearing status diagnosis is obtained.

Additionally to the resulting class, the proposed classification structure offers also a diagnosis probability. As it is schematically represented in Fig. 7, the classification result in each neural network is related with a probability value following a sigmoid function.

Therefore, the posterior probability for each class is given by the product of the three-class neural network output with the corresponding two-class neural network output. For instance, define as  $h$  the event *healthy*, as  $hi$  the

event pair *healthy/inner* and as new vector  $y_{new}$ , the two dimensional input from CCA:

$$P(h|y_{new}) = P(h|hi, y_{new})P(hi|y_{new}) \quad (3)$$

where  $P(h|hi, y_{new})$  is the probability to obtain a  $h$  classification as output of the 2-class neural network, and  $P(hi|y_{new})$  is the probability to obtain a  $hi$  classification as output of the 3-class neural network.

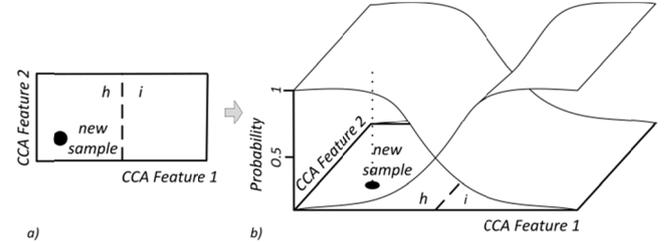


Fig. 7. Representation of the probability calculation in one neural network. (a) Neural network for classification between  $h$  and  $i$ , the dotted line represents the boundary between classes. (b) Sigmoidal functions for each of the classes ( $h$  and  $i$ ) to calculate the probability.

The building block of the classification setup is a multilayer perceptron (MLP). Each MLP has two layers and the hidden activation function is the hyperbolic tangent. For the two-class problem, the output activation function is the logistic sigmoid. Training uses the backpropagation rule for the gradient estimation and the scaled conjugate gradient as minimization technique [32]. All MLPs have 45 neurons.

#### IV. EXPERIMENTAL RESULTS

The experimental set-up is based on one induction motor and a controlled brake. They are connected by means of an additional shaft in which two bearing supports are mounted. The driving motor is controlled by an inverter. The drive is a 0.37 kW motor at 2780 rpm of rated speed.

A set of six identical bearings have been used covering the most important bearing fault scenarios: healthy ( $h$ ), inner race fault ( $i$ ), outer race fault ( $o$ ), ball fault ( $b$ ), inner-outer-ball ( $iob$ ) faults at the same time, and generalized degradation in inner and outer races ( $gdio$ ). The parameters of the bearings under test can be seen in Table I.

TABLE I  
BEARING PARAMETERS

Type	Outside diameter	Inside diameter	$N_b$	Bd	Pd	$\cos\phi$
SKF 6004	42mm	20mm	9	6.35mm	31mm	1

It is important to notice that this set of bearing conditions covers: three cyclic single point defects ( $i$ ,  $o$ ,  $b$ ), one non-cyclic generalization-roughness defect ( $gdio$ ), and a multiple cyclic single point defects ( $iob$ ), which is a rare case not usually studied. The bearing faults were carried out during the manufacture. A milling cutter was used to scratch the corresponding surfaces.

The experimental setup and the bearing set are shown in Fig. 8. Two monoaxial (orthogonal  $x$  and  $y$  axis) piezoelectric accelerometers are attached using screw mounting to one of the bearing supports, and its data were collected using an acquisition card sampling at 10 kS/s, 1 second for each measurement.

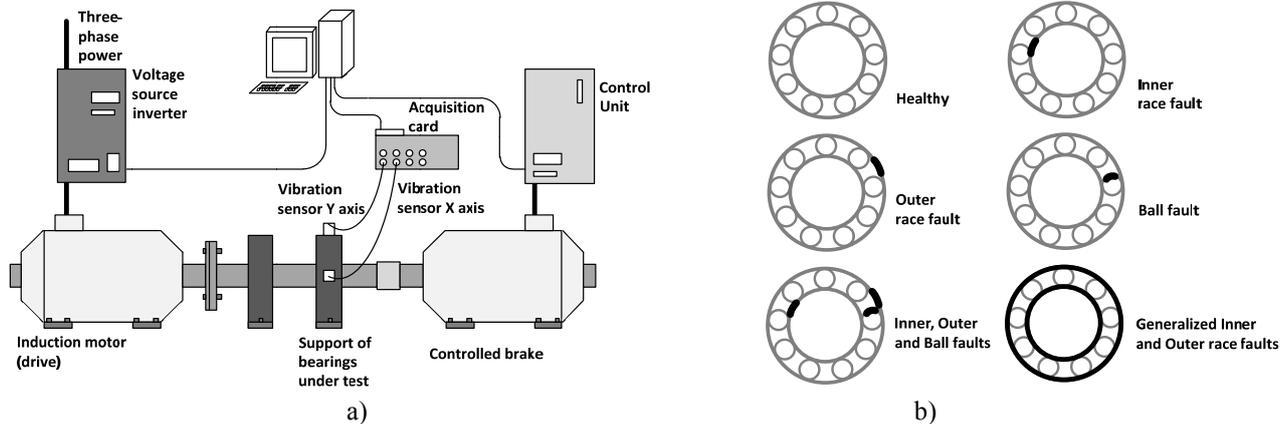


Fig. 8. Experimental arrangements. (a) Scheme of the experimental setup formed by a drive motor, a controlled break, two vibration sensors and an acquisition card. (b) Set of six bearings used, and schematic view of the faults.

In order to take into account different speed and torque combinations twenty measurements are performed for each fault and operating condition and twenty five operating conditions have been considered as it has been mentioned.

Previously to the evaluation of the proposed methodology, the characteristic fault harmonic components  $f_b, f_o, f_i$  [9] have been analyzed as well as RMS values, as classical indicators for bearing faults detection. For each measurement corresponding to each operating condition, these indicators are calculated and then the average value is obtained. It is shown in Table II these values corresponding to the rated operating condition, 100% torque and 100% speed. Although most of the indicators analyzed in Table II exhibit values bigger than the obtained under healthy conditions, it is difficult to fix thresholds values to distinguish between the six bearings conditions. Some of the indicators, specially the corresponding to the characteristic fault frequencies, are masked between them.

TABLE II  
CLASSICAL BEARING FAULTS INDICATORS  
ANALYZED UNDER RATED CONDITIONS

	$f_i, 266\text{Hz}$ [db ref.:1V RMS]		$f_o, 173\text{Hz}$ [db ref.:1V RMS]		$f_b, 112\text{Hz}$ [db ref.:1V RMS]		RMS [mV]	
	x	y	x	y	x	y	x	y
<i>h</i>	-47.7	-56.3	-55.3	-57.1	-54.7	-56	182	155
<i>i</i>	-45.6	-48.4	-46.1	-46.6	-46.1	-48.3	766	494
<i>o</i>	-40.5	-42	-41.2	-41.3	-43.3	-38.7	1005	702
<i>b</i>	-44.3	-46	-58.7	-52.2	-48	-48.4	823	517
<i>iob</i>	-44.1	-44.2	-44.5	-47.2	-45.4	43.4	1598	1116
<i>gdio</i>	-37.2	-39.2	-40.1	-43.5	-38.3	-43.2	2542	1720

It should be noticed that these results regard only one of the considered operating points, which implies that the analysis of these features in the complete range of working conditions will be more complex. However, the RMS value shows good discrimination capabilities and enough dynamic range to distinguish between the considered conditions. This fact confirms the inclusion of this parameter in the proposed statistical-time features set.

#### A. Experimental validation of the proposed methodology

Regarding the proposed methodology, as it has been mentioned, twenty-five CCAs are executed, one for each operating condition considered. Indeed, the projection of

the whole data set, considering the six bearing conditions and the 25 operating conditions at the same time, results in a saturation of the projection capabilities of the CCA. It can be seen, in Fig. 9(a), the CCA projection for the whole data set by a unique CCA. The corresponding  $dy-dx$  diagram, Fig. 9(b) relates the distances of the samples in the data space ( $dx$ ) with the distances in the latent space ( $dy$ ). It can be seen that a great deal of samples are out of the bisector which implies a poor projection quality. Therefore the obtained latent space, Fig. 9(b) presents inconsistencies, as the overlapped region between classes *h/i*. However, following the proposed methodology, by means a distributed CCA, operation the projection quality is increased. It can be seen in Fig. 10(a) the resulting CCA projection for the feature vectors drawn under rated operation conditions. This is the similar for each operating condition. It can be seen that most points lie on the bisector. Indeed, the small distances are well represented. However, there is a thickening of points around the bisector for bigger distances (distances between classes): detected as *h/i*, *b/o* and *iob/gdio*.

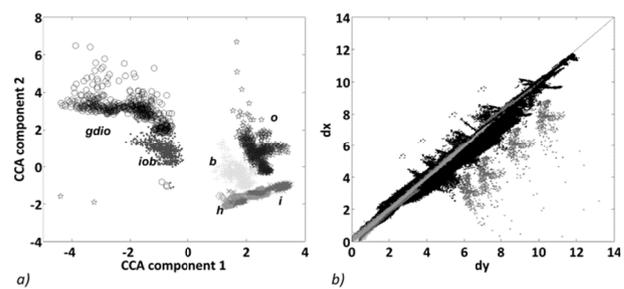


Fig. 9. Unique CCA projection for the whole data set, 20 samples per class,  $\alpha=0.5$ , 10 iterations. (a) CCA projection. (b)  $dy-dx$  diagram.

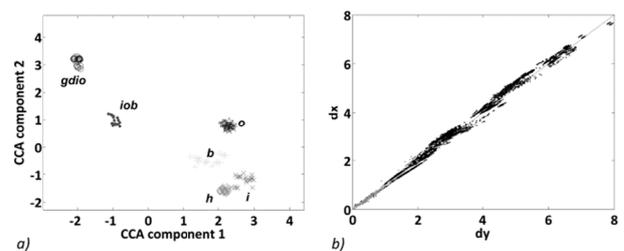


Fig. 10. CCA projection of feature vectors corresponding to 100% speed and 100% torque, 20 samples per class,  $\alpha=0.5$ , 10 iterations. (a) CCA projection. (b)  $dy-dx$  diagram.

This analysis reveals that the selected and compressed features represent the considered bearing faults as a set of disconnected manifolds. This fact implies that the use of common reduction techniques as PCA, would not be capable to characterize the considered faults.

The resulting rated CCA projection map, that will be used for classification, is shown in Fig. 11, formed by the projection of the 25 CCAs. The class pairs are well separated, although there is an overlapping in *h/i*. This figure shows the real bearing conditions behavior and how the working conditions influence on them. This resulting global CCA can be compared with the analysis of the data by PCA shown in Fig. 12.

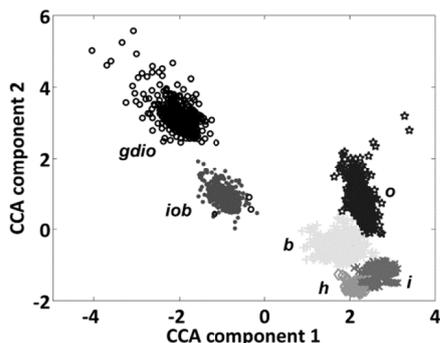


Fig. 11. Resulting global rated CCA projection.

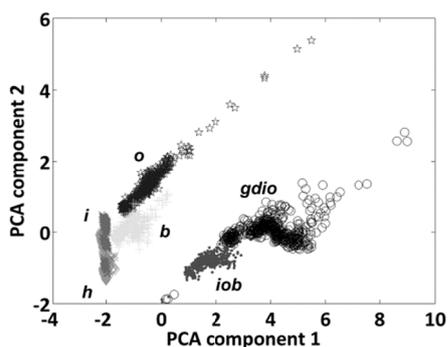


Fig. 12. PCA projection of the whole data set.

For checking the generalization properties of the proposed methodology, a test set for the recall phase has been considered. The test data base is formed by 20 vectors of features for each of the six considered bearing conditions, and calculated from the measurement of vibration signals under 98% rated speed and 93% rated torque.

TABLE III  
CONFUSION MATRIX RESULTING FROM THE  
EVALUATION OF THE WHOLE DATA SET

	<i>h</i>	<i>i</i>	<i>o</i>	<i>b</i>	<i>iob</i>	<i>gdio</i>
<i>h</i>	20	0	0	0	0	0
<i>i</i>	0	20	0	0	0	0
<i>o</i>	0	0	17	3	0	0
<i>b</i>	0	0	0	20	0	0
<i>iob</i>	0	0	0	0	16	4
<i>gdio</i>	0	0	0	0	0	20

The projected data is projected by the weighting procedure of  $k$  closest CCAs to the test operating conditions, with  $k=4$ . The classification ratio for the test set is 95% approximately. The hMLP decision regions are shown in Fig. 13. It can be seen that all points

corresponding to healthy machine are correctly classified, and only some samples between some clusters, *o/b* and *iob/gdio*, are misclassified.

It should be noticed that the evaluation of the proposed methodology has been done taken into account a wide range of working conditions (a 10% of the whole nominal speed/torque motor map). The methodology, under these diagnosis requirements, exhibits a good performance.

Moreover, and due to the distributed processing of the data in different CCAs, depending on working conditions, it suggest that the same procedure can be increase for any range of operation conditions.

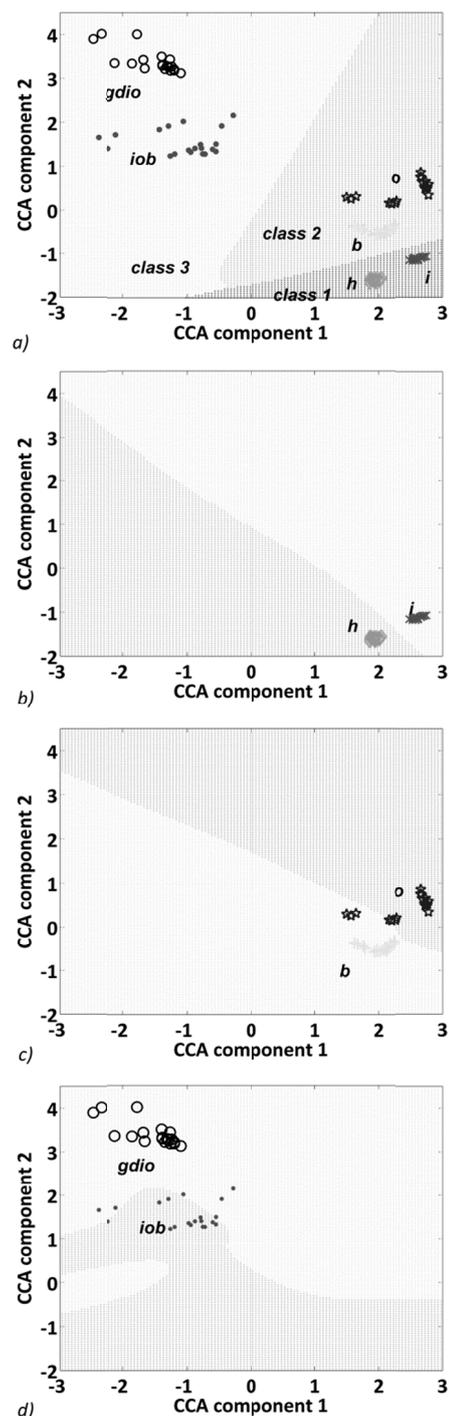


Fig. 13. Decision regions for the four MLP's of the hierarchical MLP. (a) Decision regions between *h/i*, *o/b* and *iob/gdio*. (b) Decision regions between *h* and *i*. (c) Decision regions between *o* and *b*. (d) Decision regions between *iob* and *gdio*.

### B. Experimental validation of the proposed methodology applied to a lower range of operating conditions.

The proposed methodology is, initially, presented to be feasible under a wide range of operating conditions. That is, from 80% to 100% of rated speed, and from 80% to 100% of rated torque. In this sense, if the motor is working within the specified range, the proposed bearings diagnosis method could be executed obtaining high levels of reliability. However, if a small range of operating conditions is considered, some simplifications can be applied to the proposed methodology to facilitate the implementation. Under this scenario, the use of a unique CCA projection to obtain the global CCA projection map is affordable. In order to evaluate this simplification of the proposed methodology only four operating conditions are considered: 95% and 100% of rated speed, and 90% and 95% of rated torque. The resulting CCA projection map by the application of a unique CCA is shown in Fig. 14(a). As it can be seen in Fig. 14(b), the projection preserves most of the original information.

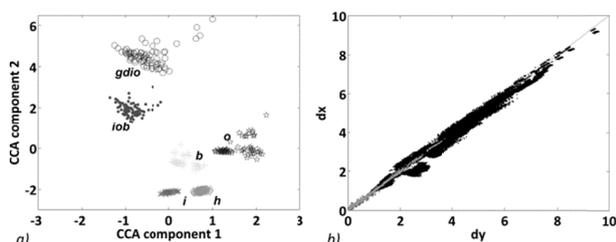


Fig. 14. CCA projection of four working conditions, 20 samples per class,  $\alpha=0.5$ , 10 iterations. (a) CCA projection. (b)  $dy-dx$  diagram.

The CCA projection map is used for classification. Due to the reduced number of working conditions, the complexity of the data distribution has decreased, and only one neural network is proposed to manage the classification. In order to evaluate this results the same test data base used previously is applied. It can be seen in Fig. 15 the classification results.

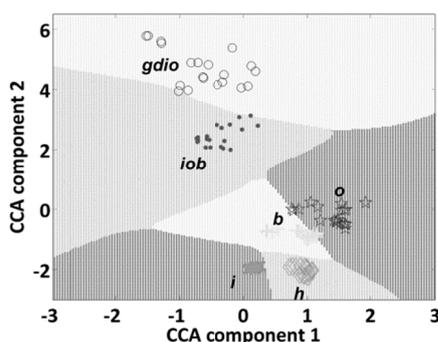


Fig. 15. Decision regions for the MLP.

TABLE IV  
CONFUSION MATRIX RESULTING FROM THE  
EVALUATION OF A LOWER RANGE OF OPERATING CONDITIONS

	<i>h</i>	<i>l</i>	<i>o</i>	<i>b</i>	<i>iob</i>	<i>gdio</i>
<i>h</i>	20	0	0	0	0	0
<i>i</i>	0	20	0	0	0	0
<i>o</i>	0	0	17	1	0	0
<i>b</i>	0	0	4	16	0	0
<i>iob</i>	0	0	0	0	20	0
<i>gdio</i>	0	0	0	0	1	19

The classification ratio for the test set is 95%. It can be seen in Table IV that all points corresponding to healthy machine are correctly classified.

This simplification of the proposed methodology is speed and torque independent as long as it can be assured the operation of the electromechanical system in the fixed range of operating conditions.

### V. CONCLUSIONS

This paper introduces a novel diagnosis methodology applied to bearings faults using information in time-domain from the vibration data.

Six different bearing scenarios have been considered, these include single points defects, combined single points defects and generalized degradation. These scenarios have been analyzed over 25 operating conditions, i.e speed and torque.

From the acquired vibration signals, a feature calculation process is performed. Then the selection process is applied, resulting in a selected set of the most significant features to maximize the discrimination between the considered faults.

Afterwards, the reduction process is carried out by means of the CCA, wherein a new set of improvements to apply this technique in fault diagnosis structures are presented. By this way, all the operating conditions as well as the different fault scenarios and its evolution, can be shown in easy and understandable two dimensional space, so that a physical interpretation of the phenomena can be performed.

Lastly, the fault identification is obtained classifying the information coming from the reduction process by means of a hierarchical neural network.

This fault diagnosis system has been trained with 25 different operating conditions, from 80%-100% rated speed and 80%-100% rated torque, every variation of 5% of each parameter. The methodology carries out the diagnosis for whatever operating condition within the aforesaid range, contrary as it could be though. This is so, the new developed improvements based on the CCA interpolation capabilities.

It should be noted that this methodology can be applied in whatever electromechanical structure where the bearings are present. The potential of the selection and extraction stages has the processing capability to extract the information coming from the defects in the bearings themselves that from other machine vibration sources.

This work results in a high performance and advanced classification structure, to the development of electromechanical system diagnosis procedures.

### VI. REFERENCES

- [1] R. J. Romero-Troncoso, R. Saucedo-Gallaga, E. Cabal-Yepez, A. Garcia-Perez, R. A. Osornio-Rios, R. Alvarez-Salas, H. Miranda-Vidales and N. Huber, "FPGA-Based Online Detection of Multiple Combined Faults in Induction Motors Through Information Entropy and Fuzzy Inference," *IEEE Trans. Ind. Electron.*, vol. 58, no. 11, pp. 5263-5270, Nov. 2011.
- [2] A. Yazidi, H. Henao, G.-A. Capolino, F. Betin and F. Filippetti, "A Web-Based Remote Laboratory for Monitoring and Diagnosis of AC Electrical Machines," *IEEE Trans. Ind. Electron.*, vol. 58, no. 10, pp. 4950-4959, Oct. 2011.
- [3] O. Ondel, E. Boutleux and G. Clerc, "Diagnosis by pattern recognition for PMSM used in more electric aircraft," in *Proc. IEEE IECON*, pp.3452-3458, Nov. 2011.

- [4] B. Akin, C. Seungdeog, U. Orguner and H. A. Toliyat, "A Simple Real-Time Fault Signature Monitoring Tool for Motor-Drive-Embedded Fault Diagnosis Systems," *IEEE Trans. on Ind. Electron.*, vol.58, no.5, pp.1990-2001, May 2011.
- [5] A. Bellini, F. Filippetti, C. Tassoni and G. A. Capolino, "Advances in Diagnostic Techniques for Induction Machines," *IEEE Trans. Ind. Electron.*, vol.55, n°12, pp. 4109-4126, Dec. 2008.
- [6] A. Garcia-Perez, R. de Jesus Romero-Troncoso, E. Cabal-Yepez and R. A. Osornio-Rios, "The Application of High-Resolution Spectral Analysis for Identifying Multiple Combined Faults in Induction Motors," *IEEE Trans. Ind. Electron.*, vol.58, no.5, pp.2002-2010, May 2011.
- [7] B. Zhang, C. Sconyers, C. Byington, R. Patrick, M. E. Orchard and G. Vachtsevanos, "A Probabilistic Fault Detection Approach: Application to Bearing Fault Detection," *IEEE Trans. Ind. Electron.*, vol. 58, no. 5, pp. 2011-2018, May 2011.
- [8] L. Frosini and E. Bassi, "Stator Current and Motor Efficiency as Indicators for Different Types of Bearing Faults in Induction Motors," *IEEE Trans. Ind. Electron.*, vol. 57, no. 1, pp. 244-251, Jan. 2010.
- [9] J. R. Stack, T. G. Habetler and R. G. Harley, "Fault classification and fault signature production for rolling element bearings in electric machines," *IEEE Trans. Ind. Appl.*, vol. 40, no. 3, pp. 735- 739, 2004.
- [10] P. Zhang, Y. Du, T. G. Habetler and B. Lu, "A Survey of Condition Monitoring and Protection Methods for Medium-Voltage Induction Motors," *IEEE Trans. Industry Applications*, vol. 47, no. 1, pp. 34-46, Feb. 2011.
- [11] F. Immovilli, M. Cocconcelli, A. Bellini and R. Rubini, "Detection of Generalized-Roughness Bearing Fault by Spectral-Kurtosis Energy of Vibration or Current Signals," *IEEE Trans. Ind. Electron.*, vol. 56, no. 11, pp. 4710-4717, Nov. 2009.
- [12] F. Immovilli, A. Bellini, R. Rubini and C. Tassoni, "Diagnosis of Bearing Faults in Induction Machines by Vibration or Current Signals: A Critical Comparison," *IEEE Trans. Ind. Appl.*, vol. 46, no. 4, pp. 1350-1359, Aug. 2010.
- [13] C. Bianchini, F. Immovilli, M. Cocconcelli, R. Rubini, A. Bellini, "Fault Detection of Linear Bearings in Brushless AC Linear Motors by Vibration Analysis", *IEEE Trans. Ind. Electron.*, vol.58, no.5, pp.1684-1694, May 2011.
- [14] F. Immovilli, C. Bianchini, M. Cocconcelli, A. Bellini and R. Rubini, "Currents and vibrations in asynchronous motor with externally induced vibration" in *Proc. IEEE SDEMPED*, pp. 580-584, Sept. 2011.
- [15] W. Q. Lim, D. H. Zhang, J. H. Zhou, P. H. Belgi and H. L. Chan, "Vibration-based fault diagnostic platform for rotary machines," in *Proc. IEEE IECON*, pp. 1404-1409, Nov. 2010.
- [16] K. W. Wilson, "Probabilistic inter-disturbance interval estimation for bearing fault diagnosis", in *Proc. IEEE SDEMPED*, pp.1-6, Aug. 2009.
- [17] J.-H. Zhou, "Reinforced Morlet wavelet transform for bearing fault diagnosis" in *Proc. IEEE IECON*, pp. 1179-1184, Nov. 2010.
- [18] E. C. C. Lau and H. W. Ngan, "Detection of Motor Bearing Outer Raceway Defect by Wavelet Packet Transformed Motor Current Signature Analysis," *IEEE Trans. Instrum. and Measur.*, vol. 59, no. 10, pp. 2683-2690, Oct. 2010.
- [19] L. Hou, "Induction motor fault diagnosis using industrial wireless sensor networks and Dempster-Shafer classifier fusion", in *Proc. IEEE IECON*, pp. 2992-2997, Nov. 2011.
- [20] J. Yu, "Local and Nonlocal Preserving Projection for Bearing Defect Classification and Performance Assessment," *IEEE Trans Ind. Electron.*, vol. 59, no. 5, pp. 2363-2376, May 2012.
- [21] C. Chen, B. Zhang, G. Vachtsevanos and M. Orchard, "Machine Condition Prediction Based on Adaptive Neuro-Fuzzy and High-Order Particle Filtering," *IEEE Trans.Ind. Electron.*, vol. 58, no. 9, pp. 4353-4364, Sept. 2011.
- [22] V. N. Ghatge and S. V. Dudul, "Cascade Neural-Network-Based Fault Classifier for Three-Phase Induction Motor," *IEEE Trans. on Ind. Electron.*, vol.58, no.5, pp.1555-1563, May 2011.
- [23] A. K. Mahamad and T. Hiyama, "Improving Elman Network using genetic algorithm for bearing failure diagnosis of induction motor," in *Proc. IEEE SDEMPED*, pp.1-6, Aug. 2009.
- [24] C. Chen, B. Zhang, G. Vachtsevanos and M. Orchard, "Machine Condition Prediction Based on Adaptive Neuro-Fuzzy and High-Order Particle Filtering," *IEEE Trans.Ind. Electron.*, vol. 58, no. 9, pp. 4353-4364, Sept. 2011.
- [25] V. Choqueuse, M. E. H. Benbouzid, Y. Amirat and S. Turri, S., "Diagnosis of Three-Phase Electrical Machines Using Multidimensional Demodulation Techniques," *IEEE Trans. on Ind. Electron.*, vol.59, no.4, pp.2014-2023, April 2012.
- [26] L. A. O Martins, F. L. C. Pádua and P. E. M. Almeida, "Automatic detection of surface defects on rolled steel using Computer Vision and Artificial Neural Networks," in *Proc. IEEE IECON*, pp. 1081-1086, Nov. 2010.
- [27] M. I. Chacon-Murguia and S. Gonzalez-Duarte, "An Adaptive Neural-Fuzzy Approach for Object Detection in Dynamic Backgrounds for Surveillance Systems," *IEEE Trans Ind. Electron.*, vol. 59, no. 8, pp. 3286-3298, Aug. 2012.
- [28] P. Demartines and J. Hérault, "Curvilinear component analysis: a self-organizing neural network for nonlinear mapping of data sets," *IEEE Trans. on Neural Networks*, vol.8, no.1, pp.148-154, Jan 1997.
- [29] G. Chicco, R. Napoli and F. Piglion, "Comparisons among clustering techniques for electricity customer classification," *IEEE Trans. on Power Syst.*, vol.21, no.2, pp. 933- 940, May 2006.
- [30] M. Cirrincione, M. Pucci, G. Cirrincione and G.-A. Capolino, "A new experimental application of least-squares techniques for the estimation of the induction motor parameters," *IEEE Trans. on Ind. Appl.*, vol.39, no.5, pp. 1247- 1256, Sept.-Oct. 2003.
- [31] J. Sun, M. Crowe and C. Fyfe, "Extending metric multidimensional scaling with Bregman divergences". *Pattern Recognition*, vol. 44, n.5, pp. 1137-1154, 2011.
- [32] R. B. W. Heng and M. J. M. Nor, "Statistical analysis of sound and vibration signals for monitoring rolling-element bearing condition," *Appl. Acoust.*, vol. 53, pp. 211–226, 1998.



**Miguel Delgado Prieto** (S'08) received the B.S. and M.S. degrees in electronics engineering from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain in 2004 and 2007 respectively. In 2007, he joined the UPC Department of Electronic Engineering as an Assistant Professor. Currently, he is a PhD student in the UPC, in the Motion and Industrial Control Group (MCIA) in Terrassa, Barcelona, Spain. His research interests include fault diagnosis in electric machines, fault detection algorithms, machine learning, signal processing methods and embedded systems.



**Giansalvo Cirrincione** (M'04) received the "Laurea" degree in electrical engineering from the Politecnico di Torino, Turin, Italy, in 1991, and the Ph.D. degree (with the congratulations of the jury) from the Laboratoire d'Informatique et Signaux de l'Institut National Polytechnique de Grenoble, Grenoble, France, in 1998. In 1999, he was on a post-doc scholarship with the Department SISTA, Leuven University, Leuven, Belgium. Since 2000, he has been an Assistant Professor with the Department of Electrical Engineering, University of Picardie "Jules Verne," Amiens, France. His current research interests include neural networks, data analysis, computer vision, brain models, and system identification.



**Antonio Garcia Espinosa** (M'05) received the M.S. degree in electrical engineering and the Ph.D. degree from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, in 2000 and 2005 respectively. In 2000, he joined the Electric Engineering Department of the UPC, where he is currently a Lecturer. He belongs to the Motion and Industrial Control Group (MCIA). His research interests include electromagnetic devices, electric machines, variable-speed drive systems, and fault-detection algorithms.



**Juan Antonio Ortega** received the M.S. Telecommunication Engineer and Ph.D. degrees in Electronics from the Technical University of Catalonia (UPC) in 1994 and 1997, respectively. In 1994, he joined the UPC Department of Electronic Engineering as a full time Associate Lecturer. In 1998, he obtained a tenured position as an Associate Professor. Since 1994 he has taught courses of microprocessors and signal processing. From 1994 to 2001 he was with

Sensor Systems Group working in the areas of smart sensors, embedded systems, and signal conditioning, acquisition and processing. Since 2001 he belongs to the Motion Control and Industrial Applications research group working in the area of motor current signature analysis. His current research activities include: motor diagnosis, signal acquisition, smart sensors, embedded systems and remote labs. In the last years, he has participated in several Spanish and European funded research projects about these items.



**Humberto Henao** (M'95–SM'05) received the M.Sc. degree in electrical engineering from the Technological University of Pereira, Pereira, Colombia, in 1983, the M.Sc. degree in power system planning from the Universidad de los Andes, Bogotá, Colombia, in 1986, and the Ph.D. degree in electrical engineering from the Institut National Polytechnique de Grenoble, Grenoble, France, in 1990. From 1987 to 1994, he was a Consultant for companies such as Schneider

Industries and GEC Alstom in the Modeling and Control Systems Laboratory, Mediterranean Institute of Technology, Marseille, France. In 1994, he joined the Ecole Supérieure d'Ingénieurs en Electrotechnique et Electronique, Amiens, France, as an Associate Professor. In 1995, he joined the Department of Electrical Engineering, University of Picardie "Jules Verne," Amiens, as an Associate Professor, where he has been a Full Professor since 2010. He is currently the Department Representative for international programs and exchanges (SOCRATES). He also leads the research activities in the field of condition monitoring and diagnosis for power electrical engineering. His main research interests are modeling, simulation, monitoring, and diagnosis of electrical machines and drives.