The h-EXIN CCA for bearing fault diagnosis

G. Cirrincione
Senior Member, IEEE
Department of Electrical Engineering, University of Picardie (UPJV), LTI, 7 Rue du Moulin Neuf, 80000 Amiens, France

M. Delgado
Member, IEEE
Department of Electronic Engineering, Technical University of Catalonia (UPC), MCIA, Rbla. San Nebridi 22, 08222 Terrassa, Spain

M. Cirrincione
Senior Member, IEEE
School of Engineering & Physics, University of South Pacific (USP), Laucala Campus, Suva, Fiji Islands

***Abstract*— This paper presents the hierarchical EXIN CCA, which represents a novel and reliable approach to complex pattern recognition problems. The methodology is based on the EXIN CCA, which is an extension of the Curvilinear Component Analysis, for data reduction, and neural networks for data classification. The effectiveness of this condition monitoring scheme is verified in a demanding bearing fault diagnostic scenario.**

Keywords- Ball bearings, Classification algorithms, Curvilinear Component Analysis, Discriminant Analysis, Fault diagnosis, Neural Networks, Time domain analysis, Vibrations

I. INTRODUCTION

Classical data reduction approaches, based on linear techniques, such as Principal Component Analysis (PCA), have been discussed by many authors emphasizing its limitation about large data sets, because it look for a global structure of data. Concerning with this problem, non-linear techniques have being used. In this sense, manifold-based learning methods have been recently applied. Among them, Self-Organizing Feature Map is the most used. This paper introduces a novel approach: the hierarchical EXIN curvilinear component analysis (h-EXIN CCA), which is suitable for complex pattern recognition problems. The basic block is the EXIN CCA Neural Network [1]-[4]. EXIN CCA is a variant of the Curvilinear Component Analysis (CCA, [5]-[7]), which preserves as much as possible the original feature space distances without previous data knowledge. CCA is a neural network (NN) which uses a data training set for estimating the nonlinear projection from the data space to a lower dimensional space, its output, here called *latent space*. In the recall phase, by means of the same training algorithm, it outputs the projection of a data input. CCA is a better version of Sammon mapping [5], because the CCA error (*stress*) function uses weights depending on distances in latent space. This paper, which is focused on pattern recognition, considers the CCA algorithm weighted by a decreasing exponential. The corresponding error function is the *right Bregman divergence* [8]. Indeed, this function penalizes inconsistent long distances and its asymmetry allows a better unfolding of data.

CCA is only constrained by the preservation of distance topology and has two sources of randomness, namely the initial conditions and the index sequence of the fixed points in the stochastic gradient algorithm. As a consequence, if the training is repeated, a different projection is found, which results critical in case of classification. In [6] constraints are added to let the axis of maximum variance be horizontal: however, it is only a partial solution to the invariance problem. The EXIN variant of CCA extends the dimensionality reduction to the case of changing environment, avoids the variability of the projection (CCA only depends on the relative positions of points in data and latent space) and allows its extension to changing multidimensional data distributions. The h-EXIN CCA is a hierarchy of these networks, which matches the complexity of certain pattern recognition problems, which are hierarchical in nature. Nevertheless, unlike most common techniques, the pattern recognition problem is not solved by using simply a supervised projection technique, but exploits a classification hierarchy of multilayer perceptrons (MLP) to output class probability. As a consequence, the non-linear projection is non-constrained by levels (unsupervised), which allows a more realistic insight on data distribution [9].

Resuming, the classification is not uniquely based on projection, but is a specialized and supervised technique which is based on the power of the classification (MLP) step. This approach is tested on the classification of bearing faults, which is the most common failure problem in rotating machinery [10]-[11].

The continuous supervision of the bearing useful life represents an important issue for cost and maintenance savings in the industrial sector. A reliable condition monitoring scheme applied to rolling elements will allow not only the reduction of unscheduled stops, but a safer electrical drive operation. Bearing defects under normal operational conditions often occur because of material fatigue. The bearing faults start from small single point defects that grow during the bearing operation and finally become a generalized roughness failure [12]. Hence, the reliable detection of the earliest bearing fault stage represents a demanding problem. Single point defects are classified by the fault specific location

G. Cirrincione is with the department of Electrical Engineering, University of Picardie (UPJV), LTI 7 Rue du Moulin Neuf, 80000 Amiens, France (e-mail: giansalvo.cirrincione@u-picardie.fr).

M. Delgado is with the department of Electronic Engineering, Technical University of Catalonia (UPC), MCIA research center, Rbla. San Nebridi s/n, 08222 Terrassa, Spain (e-mail: miguel.delgado@mcia.upc.edu,).

M. Cirrincione is with the School of Engineering, University of the South Pacific, Laucala Campus, Suva, Fiji Islands (e-mail: m.cirrincione@ieee.org).

978-1-4799-4775-1/14/$31.00 ©2014 IEEE

in: inner race, outer race and ball faults. Although different physical magnitudes such as stator currents or acoustic emissions have been considered to develop bearing fault diagnosis methodologies, the vibration analysis is the most popular in practice [13].

Most of the bearing monitoring schemes are based on the detection of some characteristic fault harmonic components [14]. However, this approach is not a simplistic matter since most of the developed diagnosis schemes leads to a delayed diagnosis until the characteristic fault frequencies have enough presence in the spectra to be clearly localized. Advanced signal processing techniques, such as probabilistic models [15], high-resolution frequency analysis [16] or enhanced wavelete decompositions [17]-[18], applied to the measured physical magnitude have been also used to obtain reliable fault indicators. However, most of these approaches do not deal with an earlier single bearing fault identification.

II. Curvilinear Component Analysis

One of the most recent and powerful strategies for nonlinear feature reduction is based on distance preservation algorithms. The curvilinear component analysis is a self-organizing neural network which performs the quantization of a data training set for estimating the corresponding nonlinear projection from the data space to a lower dimensional space (latent space). For every pair of different features vectors in the original feature space (data space), a between-point distance D_{ij}, is computed, $D_{ij} = \|x_i - x_j\|$. 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. In order to face this problem the CCA technique defines a distance function threshold, λ, 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 basic procedure of the CCA is shown schematically in Fig. 1.

The EXIN CCA right Bregman divergence [8] is given by:

$$E_{Bregman}(p_j) = \lambda^2 p_j \left[e^{-\frac{D_{ij}}{\lambda}} - e^{\frac{L_{ji}}{\lambda}} + \left(D_{ij} - L_{ij}\right) \frac{e^{-\frac{L_{ij}}{\lambda}}}{\lambda} \right] \qquad (1)$$

being p_j the *j-th* sample to project. The stochastic gradient algorithm for minimizing (1) is then:

$$p_j \leftarrow p_j - \alpha \frac{L_{ij} - D_{ij}}{L_{ij}} e^{-\frac{L_{ij}}{\lambda}} \qquad (2)$$

where α (learning rate) and λ are scalars. The learning rate uses to be fixed at 0.5 to allow a substantial initial step size, while λ, which determines the radius of influence, used to be defined as three times the standard deviation of the data set. Indeed, this function penalizes long distances and its asymmetry allows a better unfolding of data. 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. In the recall phase, by means of the same training algorithm, it outputs the projection of a data input.

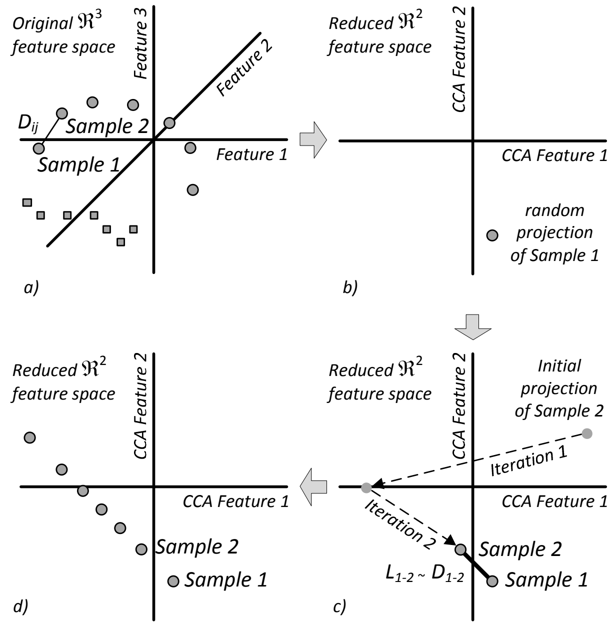


Fig. 1. CCA operation scheme sequence. (a) Seven feature vectors for each of the two classes (circles and squares) represented in a three-dimensional data space. (b) CCA projection of the first feature vector of one operating condition (circle) in the latent space. (c) CCA projection of the second feature vector of the same operating condition (circle) 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 (circle).

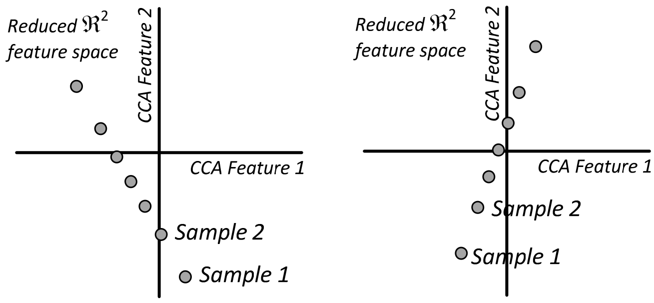


Fig. 2. Representation of two CCA executions over a same original set of feature vectors. The resulting two-dimensional projections are identical, but the global latent space position changes.

III. Diagnosis Methodology

The proposed diagnosis methodology is composed of four steps:

1) Feature estimation from the vibration signal.
2) Feature selection (only the most significant features are selected during the training process).

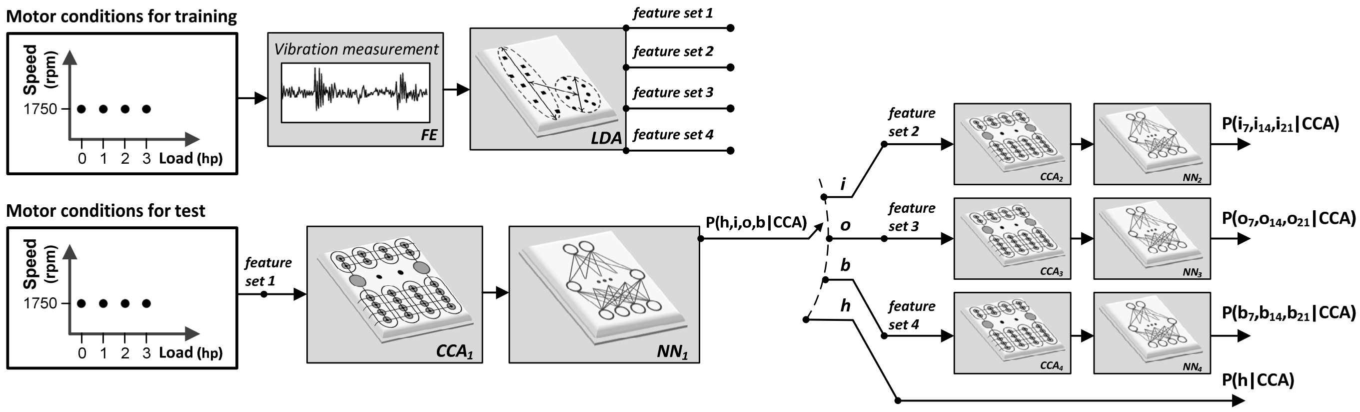


Fig. 3. Proposed diagnosis methodology scheme including feature estimation (FE), feature selection (LDA), feature reduction (CCA) and classification (NN).

3) Feature reduction by projecting to the intrinsic data dimension.
4) Classification based on a neural approach (fault diagnosis).

The complete methodology is represented in Fig. 3. The training set is used to select the most significant features during the feature estimation stage, and train the hierarchical CCA and MLP. Then, the test set is evaluated. In this work four bearing conditions have been considered, namely: healthy (*h*), inner race fault (*i*), outer race fault (*o*) and ball fault (*b*). Moreover, three different severity levels (sub-classes) have been considered for each fault ranging from small (i_7, o_7 and b_7) to medium (i_{14}, o_{14} and b_{14}) and to big single-point defect (i_{21}, o_{21} and b_{21}).

Four different steady state operating conditions have been considered. For each combination of bearing scenario, operating condition, and severity level a set of vibration measurements have been acquired.

A. Feature estimation

From each acquired vibration signal measurement, a set of statistical-time features is computed. This kind of features allows the characterization of the acquired measurements. A total of 15 features from time-domain are proposed: mean, maximum value, root mean square (rms), square root mean (srm), 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. These features exhibit most of the characteristic statistical-time information contained in the measurements [19].

B. Feature selection

The proposed features contain a large portion of the information contained in the vibration signal; however only some of them are really significant. These ones, in turn, depend on the considered bearing defects, the appearance of additional sources of vibration, and the bearing location. In this sense, the most significant features may be different depending on the scenario.

Different techniques can be applied to analyze the feature relevance with regard to the considered diagnosis scenario. Linear Discriminant Analysis (LDA) [20] is one of the classical techniques for feature selection. LDA quantitatively evaluates the discriminant capabilities of the proposed features with regard to classes. This analysis shows how each feature contributes to a proper representation of the measurements in the data space by estimating how well classes are delimited and separated. Every two and three combinations of the estimated features, as well as their individual capabilities have been evaluated. In the end an ordered list with the most significance features is obtained.

However, although some features exhibit better discrimination capabilities between classes than others, specific features sets are proposed in this study depending on the analyzed set of measurements. In this sense, four sets of features are proposed in order to do a hierarchical discrimination: first classification between *h*, *i*, *b* and *o*, second classification between i_7, i_{14} and i_{21}, third classification between b_7, b_{14} and b_{21} and finally, a fourth between o_7, o_{14} and o_{21}.

The most significant features to discriminate the considered classes/severities are shown in Table I.

TABLE I SETS OF THE MOST SIGNIFICANT FEATURES DEPENDING ON THE CLASSIFICATION SCENARIO.

Classification scenario	*Most significant features*	
Healthy (h) / Inner race fault (i) / Outer race fault (o) / Ball fault (b)	*feature set 1* rms srm	 shape factor standard deviation
Inner race fault (i) severities: i_7, i_{14} and i_{21}	*feature set 2* rms srm	 maxim value variance
Outer race fault (o) severities: o_7, o_{14} and o_{21}	*feature set 3* rms standard deviation	 srm variance
Ball fault (b) severities: b_7, b_{14} and b_{21}	*feature set 4* crest factor rms	 impulse factor shape factor

C. Feature reduction

The first level of classification is more difficult to achieve than the second one because of the bigger complexity of the data set distribution.

Hence, the whole data set has been divided in several subsets in order to improve the CCA data reduction.

The hierarchical nature of the classification (classes and subclasses) can be exploited by using a corresponding hierarchy of CCA's : the data reduction problem is so simplified because only one CCA is needed for classifying the fault and other three CCA's are needed for estimating the corresponding severity level.

During the recall phase, the first four-dimensional feature vector (feature set 1) is fed to CCA_1. After this step (bearing fault classification), if a fault bearing is detected, the corresponding feature vector (feature set 2, feature set 3 or feature set 4) is fed to the corresponding CCA (CCA_2, CCA_3 or CCA_4), to assess the associated neural network.

D. Classification

Due to the different number of considered operating conditions (four), and the number of considered bearing fault scenarios (three and healthy) and severity levels (three for each fault scenario), a two-level hierarchical neural network is applied. The building block of this architecture is the multilayer perceptron, MLP [20]. The choice of this neural network has been dictated both from the simplicity and flexibility, and, above all, for the possibility of outputting the class conditional probabilities thanks to the use of the soft max activation function (this is not possible with other neural networks) and the cross-entropy error function [20]. Each MLP has two layers and the hidden activation function is the hyperbolic tangent. Its training uses the backpropagation rule for the gradient estimation and the scaled conjugate gradient as minimization technique. All MLP's have 45 neurons. These blocks are hierarchically organized (h-MLP). This structure allows the classification in two steps: a first neural network classifies a two-dimensional feature vector (resulting from the CCA_1 projection) between four predefined classes (in this application: h, i, o and b). If the classification result is different from healthy, specific degradation assessment classifiers (three) are placed in a second level, one for each fault scenario. Then, once the input has been classified in the first neural network, the corresponding second neural network is recalled, and the bearing status and its severity level are obtained. The neural network does not only output the class membership but also its probability, as cited previously. This additional information is fundamental both for assessing the level of confidence of the classification and if risk analysis is required. At our knowledge, this feature, which justifies the use of MLP, has never been exploited for bearing fault diagnostics.

The posterior probability for each class is given by the product of the four-class neural network output with the corresponding three-class neural network output.

For instance, define as i the event *inner race fault* and define the new feature vector to be classified as y_{new}.

$$P(i_7 \mid y_{new}) = P(i_7 \mid i, y_{new}) P(i \mid y_{new}) \tag{3}$$

where $P(i \mid y_{new})$ is the probability of obtaining the event i as output of the first four-class classifier, and $P(i_7 \mid i, y_{new})$ is the probability of obtaining event i_7 as output of the corresponding three-i-class classifier.

IV. EXPERIMENTAL RESULTS

The experimental data come from the bearing data center [21], which provides access to bearing test data for normal and faulty conditions. Experiments were conducted using a 2 horsepower Reliance Electric motor. The accelerometer was mounted on the drive end of the motor housing near to the motor bearings. Motor bearings were seeded with faults using electro-discharge machining. Faults ranging from 0.007 inches in diameter to 0.021 inches in diameter were introduced separately at the inner raceway, rolling element and outer raceway. Faulted bearings were reinstalled into the test motor and vibration data were recorded for motor loads of 0 to 3 hp (motor speeds of 1797 to 1720 rpm, respectively). Data used in this study correspond to the case of normal bearings and single-point drive end defects. Data were collected at 12,000 samples/second for all the experiments. The bearing fault conditions of data and the parameters of the bearings under test are shown in Table II and III, respectively.

TABLE II

CLASSICAL BEARING FAULT INDICATORS ANALYZED UNDER RATED CONDITIONS

Bearing condition		*Fault specifications*	
		Diameter [inches]	*Depth [inches]*
Healthy (h)		-	-
Inner race fault (i)	i_7	0.007	0.0011
	i_{14}	0.014	0.0011
	i_{21}	0.021	0.0011
Outer race fault (o)	o_7	0.007	0.0011
	o_{14}	0.014	0.0011
	o_{21}	0.021	0.0011
Ball fault (b)	b_7	0.007	0.0011
	b_{14}	0.014	0.0011
	b_{21}	0.021	0.0011

TABLE III

BEARING PARAMETERS

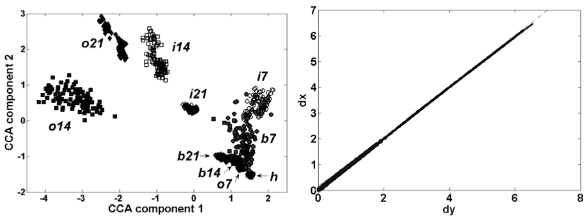
Type	Outside diameter	Inside diameter	N_b	Bd	Pd	$\cos\phi$
SKF6205	2.04 in	0.098 in	9	7.95 in	1.53 in	0.9

Thirty measurements are performed for each kind of fault, severity level and operating condition, twenty-five measurements are used for training and five for test purposes.

Regarding the proposed methodology, as it has been mentioned, four CCAs are executed, one for each classification scenario. Following the proposed methodology, by means of a distributed CCA operation, the projection performance is respected. However, the use of different four-dimensional feature sets, depending on the classification scenario, allows an increase of the discrimination capabilities. In this context, this CCA-based methodology allows the data visualization and an enhanced interpretation of the underlying physical phenomenon.

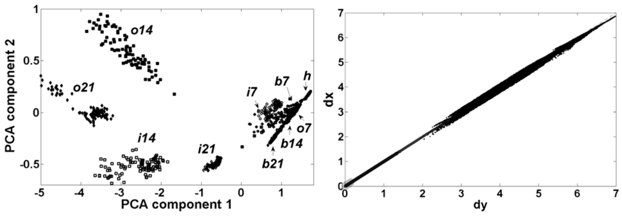
It can be seen in Fig. 4(a) the resulting CCA_1 projection for the feature vectors drawn for the whole data base. The corresponding *dy-dx* diagram, Fig. 4(b), relates the distances of the samples in the data space (*dx*) with the distances in the latent space (*dy*). It can be seen that most points lie on the bisector. This analysis reveals that the selected and compressed features represent the considered bearing faults as a set of disconnected manifolds, which are well detected and represented because all points remain very close to the bisector. The use of common reduction techniques as Principal Component Analysis [20], which is the most used linear approach, are not be capable to characterize the considered faults. Indeed, the resulting CCA_1 projection can be compared with the analysis of the same data by PCA shown in Fig. 5. The classical PCA projection, although maintaining most of the data variance, exhibits (see Fig. 5(a)) a lower projection performance, as it is evident in the highly *i/o/b* overlapped region. Also, Fig. 5(b) shows a big number of large distances (the cluster around the central and right part of the bisector), which correspond to the inter-cluster distances. It reveals the presence of bad allocations of clusters because of this projection: more compact groupings imply more intra-cluster and less inter-cluster distances. Indeed, this is what is shown in Fig. 5(b).

The resulting rated CCA1 projection map, used for the first classification level, is shown in Fig. 6(a). The four classes are well separated, although there is an overlapping between case *o*/*b* (case *o* and case *b*). This figure shows the real bearing conditions behavior and how the working conditions and severities degrees influence them. The same procedure has been carried out with the CCA's projections and MLP classifiers in the second layer.



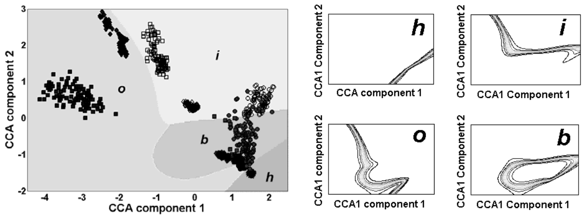
a) b)

Fig. 4. CCA_1 projection of feature vectors corresponding to the whole data base characterized by the feature set 1, α=0.5, λ=1, 50 iterations. (a) CCA_1 projection. (b) *dy-dx* diagram.



a) b)

Fig. 5. PCA projection of feature vectors corresponding to the whole data base. (a) PCA projection. (b) *dy-dx* diagram.



a) b)

Fig. 6. Decision regions for the first MLP's of the hierarchical MLP. (a) *h*, *i*, *o* and *b* classification regions. (b) Probability curves related with each region.

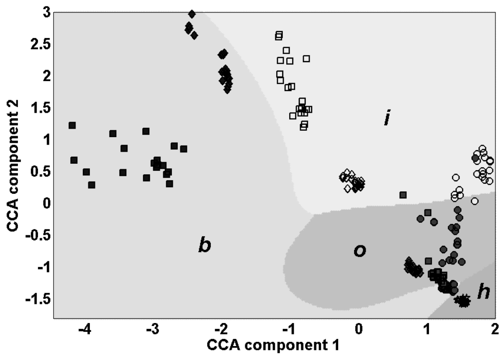
For checking the generalization properties of the proposed methodology, a test set for the recall phase has been considered. The test data base is formed by five vectors of features for each of the four considered bearing scenarios, severity levels and operation conditions.

TABLE IV

CONFUSION MATRIX RESULTING FROM THE EVALUATION OF THE h-MLP

	h	i_7	i_{14}	i_{21}	o_7	o_{14}	o_{21}	b_7	b_{14}	b_{21}
h	60	0	0	0	0	0	0	0	0	0
i_7	0	19	0	0	1	0	0	0	0	0
i_{14}	0	0	20	0	0	0	0	0	0	0
i_{21}	0	0	0	20	0	0	0	0	0	0
o_7	0	0	0	0	18	0	0	2	0	0
o_{14}	0	0	0	2	0	18	0	0	0	0
o_{21}	0	0	0	0	0	0	20	0	0	0
b_7	0	0	0	0	2	0	0	18	0	0
b_{14}	0	0	0	0	0	0	0	0	20	0
b_{21}	0	0	0	0	0	0	0	0	0	20

The classification ratio for the test set is 97% approximately. The h-MLP decision regions are shown in Fig. 7. It can be seen that all points corresponding to the healthy machine are correctly classified, and only some samples between clusters *o*/*b* (case *o* and case *b*) are misclassified.



a)

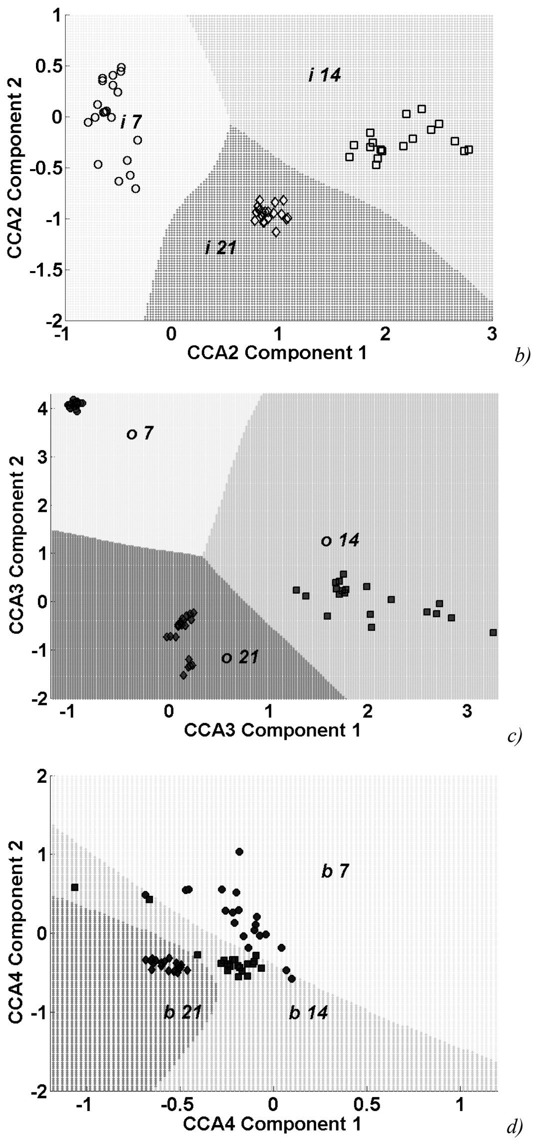


Fig. 7. Decision regions for the four MLP's of the hierarchical MLP and the corresponding test sets. (a) Decision regions between healthy, inner, outer and ball cases (case *h*/*i*/o/b). (b) Decision regions between the three inner fault severity levels (case $i_7/i_{14}/i_{21}$). (c) Decision regions between outer fault severity levels (case $o_7/o_{14}/o_{21}$). (d) Decision regions between ball fault severity levels (case $b_7/b_{14}/b_{21}$).

V. GENERAL DISCUSSION

In the proposed work, the bearing diagnosis is approached as a pattern recognition problem. A signal processing based methodology is proposed to, firstly, identification of the most significant statistical-time features with regard to the diagnosis problem, as an adaptive procedure to the application; secondly, the data visualization and an enhanced interpretation of the underlying physical phenomenon in a 2-dimensional space; thirdly, the class membership and its probability, fundamental both for assessing the level of confidence of the classification and if risk analysis.

Moreover, the potential of the selection and reduction stages has the processing capability to extract the information coming from the defects in the bearings themselves that from other machine vibration sources. Hence, the proposed methodology includes the filtering of external perturbations such as mechanical or electrical noise.

The analysis of the regular characteristic fault frequencies is not a simplistic matter, and basic diagnosis schemes may lead to a delayed diagnosis. The bearing defects could be detected after a severe level of bearing degradation, in which the characteristic fault frequencies have enough presence to be clearly localized. The absence of clear characteristic fault frequencies should not be interpreted as a completely healthy condition of the bearing. The characteristic bearing fault spectral indicators are usually masked between them due to additional vibration modes produced by the rest of the mechanical interactions.

In order to exploit completely the proposed methodology, a set of representative bearing conditions is needed to obtain a useful data base of measurements. However, once the competency of the method has been demonstrated in the manuscript, the methodology can be adapted to any kind of diagnosis requirements. In industrial applications it could be difficult to perform training process taking into account all the failure modes. Nevertheless, in a real application, it could be always possible to perform the calibration process considering two different states: healthy and not healthy. The vibration sensors are mounted near the bearing under test and the most significant features selected in the manuscript are calculated. Since this moment, the healthy region can be defined. As it has been observed during the experimental validation in the manuscript, if a measurement corresponds to a point in the reduced feature space outside the healthy region, some kind of fault is taking place.

The training procedure, as in most of the pattern recognition based methods is carried out off-line. That is, a representative data base is obtained from the system under test, and the different stages are configured. The time needed for this first stage is totally dependent of the data base characteristics and the user expertise. The second stage, once the methodology is already calibrated, takes into account: the acquisition of the physical magnitudes, directly the calculation of the selected features, the direct curvilinear component analysis projection and, finally, the hierarchical neural network application. This process can be executed in a few seconds by regular digital processors. The motor diagnosis processes must be available to be executed during a normal system operation. However, the diagnosis does not require generally to be executed in real time mode. Most of the mechanical faults are based on a low degradation time constant. This fact implies that once the signal is acquired, the diagnosis algorithms execution can be carried out during the next minutes.

VI. CONCLUSION

A novel architecture, tailored on a difficult diagnosis problem, as the bearing fault recognition, has been presented. It comprises a hierarchy of curvilinear component projections (h-CCA) and a hierarchy of multilayer perceptrons (h-MLP). The output is given by the class membership probability, which is suitable for further analysis (e.g. risk). Unlike neural and non-neural techniques in the literature, the task is shared by two different hierarchies. This is the peculiarity and the power of the proposed method, which allow its application to other complex pattern recognition problems. Future work will deal both with novel applications (e.g.

use of stator current for machine diagnostics) and with a deeper analysis of the hierarchies and their possible improvements.

ACKNOWLEDGMENT

The authors would like to express their thanks to the Case Western Reserve University (CWS) for their disinterested contribution providing access to the bearing data center analyzed in this paper. Furthermore, the authors appreciate the support of the Professor Laurens van der Maaten, from the Delft University of Technology, for his disinterested distribution of different feature reduction technique's codes.

VII. REFERENCES

[1] G. Cirrincione, M. Delgado, J. A. Ortega and H. Henao, “Bearing fault diagnosis by EXIN CCA,” in *Proc.* IEEE *International Joint Conference on Neural Networks*, IJCNN, Brisbane, Australia, June 2012.

[2] M. Delgado, G. Cirrincione, A. García, J. A. Ortega and H. Henao, “EXIN CCA for Visualization and Classification,” in *Proc.* IEEE *International Conference on Electrical Machines*, ICEM, Marseille, France, September 2012.

[3] M. Delgado, G. Cirrincione, A. García, J. A. Ortega and H. Henao, “Accurate Bearing Faults Classification based on Statistical-Time Features, Curvilinear Component Analysis and Neural Networks,” in *Proc.* IEEE *Industrial Electronics Conference*, IECON, Montréal, Canada, October 2012.

[4] M. Delgado, G. Cirrincione, A. García, J. A. Ortega and H. Henao, “Bearing Faults Detection by a Novel Condition Monitoring Scheme based on Statistical-Time Features and Neural Networks,” *IEEE Transactions. on Industrial Electronics*, to be published.

[5] P. Demartines and J. Hérault., “Curvilinear Component Analysis : a Self-Organising Neural Network for Non-Linear Mapping of Data Sets,” *IEEE Trans. on Neural Networks*, vol 8, n°1, pp.148-154, 1997.

[6] J. Hérault, C. Jausions-Picaud and A. Guérin-Dugué, “Curvilinear component analysis for high-dimensional data representation: I. Theoretical aspects and practical use in the presence of noise,” in *Proc.* IEEE *International Work Conference on Artificial Neural Networks*, IWANN, 1999, Alicante, Spain, 1999, pp. 625-634.

[7] A. Guérin-Dugué, P. Teissier, G. Delso-Gafaro. and J. Hérault, Curvilinear Component Analysis for High-dimensional Data Representation : II. Examples of introducing additional mapping constraints for specific applications.” in *Proc.* IEEE *International Work Conference on Artificial Neural Networks*, IWANN, Alicante, Spain, 1999, pp. 635-644.

[8] J. Sun, M. Crowe and C. Fyfe, "Extending metric multidimensional scaling with Bregman divergences," *Pattern Recognition Journal*, vol 44, n°5, pp. 1137-1154, 2011.

[9] J. A. Leeand M. Verleysen, “Unsupervised dimensionality reduction: Overview and recent advances,” in *Proc.* IEEE *International Joint Conference on Neural Networks*, IJCNN, Barcelona, Spain, July 2010.

[10] S. Nandi, H. A. Toliyat, and X Li, “Condition monitoring and fault diagnosis of electrical motors—A review,” *IEEE Trans. on Energy Conversion*, vol. 20, no. 4, pp. 719-729, 2005.

[11] A. Bellini, F. Filippetti, C. Tassoni, and G. A. Capolino, “Advances in diagnostic techniques for induction machines,” *IEEE Trans. on Industrial Electronics*, vol. 55, no. 12, pp. 4109-4126, 2008.

[12] 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. on Industrial Applications*, vol. 40, no. 3, pp. 735- 739, 2004.

[13] 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. on Industrial Applications,* vol. 46, no. 4, pp. 1350-1359, 2010.

[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 *International Symposium on Diagnostics for Electrical Machines, Power Electronics ¬ Drives, SDEMPED*, pp. 580-584, Bologna, Italy, Sept. 2011.

[15] K. W. Wilson, “Probabilistic inter-disturbance interval estimation for bearing fault diagnosis”, in *Proc.* IEEE *International Symposium on Diagnostics for Electrical Machines, Power Electronics ¬ Drives, SDEMPED*, pp. 1-6, Cargèse, France, Sept. 2009.

[16] 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. on Industrial Electronics*, vol.58, no.5, pp.2002-2010, 2011.

[17] J.-H. Zhou, “Reinforced Morlet wavelet transform for bearing fault diagnosis” in *Proc.* IEEE *Industrial Electronics Conference,* IECON, Phoenix, USA, pp. 1179-1184, November 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. on Instrumentation and Measurement*, vol. 59, no. 10, pp. 2683-2690, 2010.

[19] L. Yaguo, H. Zhengjia and Z. Yanyang, “A new approach to intelligent fault diagnosis of rotating machinery”, *Expert Systems with Applications*, vol. 35, no. 4, pp. 1593-1600, Nov. 2008.

[20] C. Bishop, *Pattern Recognition and Machine Learning*, Springer; 2007.

[21] K. A. Loparo, “Bearings vibration data set,” Case Western Reserve University http://www.eecs.cwru.edu/laboratory/bearing/download.htm

VIII. BIOGRAPHIES

Giansalvo Cirrincione, IEEE Senior Member (2011) received the “Laurea” degree in electrical engineering from the Politecnico di Turino, Turin, Italy, in 1991, and the Ph.D. degree in Cognitive Science from the National Polytechnical Institute of Grenoble, France, in 1998. From 1999 to 2000 he was a post-doc in applied mathematics at the Catholic University of Leuven (Belgium). Since 2000 he has been Associate Professor at the University of Picardie Jules Verne (UPJV), at the Department of Electrical Engineering and Informatics Engineering (GEII) at the Professional University Institute (IUP) and member of the Laboratory of Innovative Technologies (LTI), Amiens.
His current research interests are neural networks, pattern recognition, diagnosis of electrical drives, signal processing, system identification, linear regression techniques, power electronics, electrical machines and drives, linear algebra, computer vision, applied mathematics, artificial intelligence.
He is author of over 100 papers, 30 of which on high impact factor journals, and of one book.

Miguel Delgado Prieto (S'08, M'12) received the M.S. degree in Electronics Engineering and the Ph.D. degree in Electronics Engineering from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain in 2007 and 2012 respectively. From 2004 to 2008 he was a Teaching Assistant in the Electronic Engineering Department of the UPC. In 2008 he joined the Motion and Industrial Control Group (MCIA), where he is currently a Post-Doc Researcher. His research interests include fault detection algorithms, machine learning, signal processing methods and embedded systems.

Maurizio Cirrincione, IEEE Senior Member (2010) received the “Laurea” degree from the Politecnico di Turino, Turin, Italy, in 1991, and the Ph.D. degree from the University of Palermo, Palermo, Italy, in 1996, both in electrical engineering. From 1996 to 2005 he was a Researcher with the I.S.S.I.A.-C.N.R. Section of Palermo (Institute on Intelligent Systems for Automation), Palermo, Italy. Since2005 he has been Full Professor of Control Systems at the UTBM (University of Technology of Belfort-Montbéliard) in Belfort, France and member of the IRTES Laboratory and the FCLab (Fuel Cell laboratory) in Belfort, France. Since April 2014 he has been the Head of the “School of Engineering and Physics” of the University of the South Pacific in Suva, Fiji Islands. His current research interests are neural networks for modeling and control, system identification, intelligent control, power electronics, power quality, renewable energy systems, control of fuel cell systems, hybrid vehicles, and electrical machines and drives with rotating or linear AC motors. Dr. Cirrincione was awarded the 1997 “E.R.Caianiello” prize in 1997 for the best Italian Ph.D. thesis on neural networks. He is author of over 140 papers, 50 of which on high impact factor journals, and of two books.

.