

Evaluation of Machine Learning Techniques for Electro-Mechanical System Diagnosis

M. Delgado, A. García, J. C. Urresty, J.-R. Riba, J. A. Ortega
MCIA Center, Electronic Dept., Technical University of Catalonia
Rbla. San Nebridi s/n – TR14 GAIA - 0822
Terrassa, Spain
Tel.: +34 / (93) – 739 85 18
Fax: +34 / (93) – 739 89 72
E-Mail: miguel.delgado@upc.edu

Acknowledgments

This work was supported in part by the Spanish Ministry of Science and Technology under the TRA2010-21598-C02-01 Research Project.

Keywords

<< Diagnostics, Maintenance, Actuator, Permanent magnet motor >>

Abstract

The application of intelligent algorithms, in electro-mechanical diagnosis systems, is increasing in order to reach high reliability and performance ratios in critical and complex scenarios. In this context, different multidimensional intelligent diagnosis systems, based on different machine learning techniques, are presented and evaluated in an electro-mechanical actuator diagnosis scheme. The used diagnosis methodology includes the acquisition of different physical magnitudes from the system, such as machine vibrations and stator currents, to enhance the monitoring capabilities. The features calculation process is based on statistical time and frequency domains features, as well as time-frequency fault indicators. A features reduction stage is, additionally, included to compress the descriptive fault information in a reduced feature set. After, different classification algorithms such as Support Vector Machines, Neural Network, k -Nearest Neighbors and Classification Trees are implemented. Classification ratios over inputs corresponding to previously learnt classes, and generalization capabilities with inputs corresponding to learnt classes slightly modified are evaluated in an experimental test bench to analyze the suitability of each algorithm for this kind of application.

Introduction

It has been increased the use of new electrical machines, as Permanent Magnet Synchronous Motors (PMSM), during the last years. Although it is difficult to replace the classical Induction Motors in the most of industrial applications, the use of PMSM is growing in critical sectors such as automotive or aeronautical [1]. High performance, light weight and small size, are important characteristics which make the PMSM a good option for electrical traction or drive actuation tasks. In this context, PMSMs are implemented in these critical applications with a mandatory monitoring system. For that reason it has been necessary during the last years, the analysis of electro-mechanical systems PMSM based and fault diagnosis systems to cover any kind of condition.

A great deal of studies has been performed about PMSM behavior under different fault conditions [2]. However, in order to characterize the PMSM behavior, most of the works are focused on single fault detection, and even, under stationary speed and torque conditions. Due to the aforesaid critical applications, there is a demand of diagnosis methods able to detect different kinds of faults in an electro-mechanical actuator PMSM based. Moreover, it is necessary the PMSM analysis under a complete set of varying conditions, even if different faults appear at the same time combined in the system.

In order to deal with these high diagnosis requirements, a specific methodology based on four steps is being applied: data acquisition, features calculation, features reduction and classification.

If different faults are considered, it should be noticed that the effects reflected in a single physical magnitude from different sources of faults can be similar [3]. In order to overcome this limitation, the analysis of complementary physical magnitudes provides enhanced monitoring capabilities. Different faults can be observed from different physical magnitudes point of view, increasing the collected information of the faults effects. Classical vibrations signatures combined with motor current stator analysis, represents one of the most promising multidimensional strategies, since it has been proved their individual capabilities in a great deal of diagnosis applications.

The calculated features from the acquired data represent all the available information to describe the system condition. A reduced number of features cannot be useful and descriptive enough in order to cover different actuator faults. In order to obtain a complete features set, the calculation of a great deal of features is done. In this work, the data acquired is proposed to be analyzed in time, frequency and time-frequency domains to obtain a complete complementary information. Additionally to statistical features from time and frequency domain, Wavelet Packet Transform and Empirical Mode Decomposition are applied as advanced time-frequency analysis techniques, since it has been proved their feasibility in motor fault diagnosis under dynamic conditions [4-5].

Regarding the relation between the calculated features set and the performance of the classification algorithm, there is another issue to be taken into account. The known “*curse of dimensionality*” consists in the effect that a great deal of features can produce over the classifier convergence. The classification capability can be reduced if the number of features is too large, because there is not additional information in all of them, and this fact increase the classifier complexity and provokes classification confusions. The use of features reduction procedures is used between the feature calculation and classification stages. An optimized features set containing only the most informative features is obtained. Different features reduction techniques can be applied however, Partial Least Squares has been selected due to its proper results in different applications [6].

Finally, arrays of features are used to supply an intelligent classification algorithm. The classifiers are able to recognize hidden patterns within features in order to relate them with different considered classes. However, the training method and learning capabilities are different depending on the used classification algorithm. Classifiers based on Support Vector Machines (SVM), Neural Networks (NN), k -Nearest Neighbor (k NN) and Classification Trees (CT) have been successfully applied by different authors to automated detection and diagnosis purposes. Due to its structural risk minimization principles, SVM exhibits a high accuracy and a good generalization capability able to offer better classification performance than other techniques. Different authors [7] have emphasized the SVM diagnosis capabilities, in motor diagnosis however, dealing with multiclass problems, there is not a clear SVM structure to use because SVM is intrinsically a two-class solver. NN represents one of the most classical classification algorithms used in machine diagnosis field [8]. NN offers a great deal of configuration parameters and different structures, which represent an important adaptation capability to different diagnosis problems, but also a complex classifier to obtain optimized results. One of the simplest classification algorithms is the k NN, which is totally based on the training input set [7]. A similarity measure is calculated over new inputs to establish the nearest class. This calculation simplicity represents an understandable classification process, but a rigorous method in diagnosis scenarios with non-well separated classes, as in some electro-mechanical fault detection systems. Other techniques as CT, base the classification process in a tree of binary partitioning *if..then* rules, to assign the corresponding class to new inputs [9]. Although it has been proved to be an efficient classifier, not all the feature are considered during the classification, and some useful descriptive features relations can be unfortunately discarded.

Therefore, the contribution of this work lies in the evaluation of different machine learning techniques for electro-mechanical system diagnosis purposes. Different classification algorithms are introduced in a fixed diagnosis structure formed by the implementation of data acquisition, features calculation and features reduction stages. Single and combined fault scenarios as well as stationary and non-stationary speed and torque requirements are considered to compare the diagnosis performances of each classifier. Moreover, the generalization capabilities are also analyzed by using slightly modified inputs, corresponding to similar classes learnt during the training process.

The organization of this paper is as follows. First, the used diagnosis scheme as well as an introduction of each used technique is presented. Second, the electro-mechanical experimental test-bench is shown. Finally, the experimental results are discussed, and the main conclusions are summarized.

Method and materials

The electro-mechanical system diagnosis methodology is shown in figure 1. The used techniques in each stage, with special attention to the classification algorithms, are explained next.

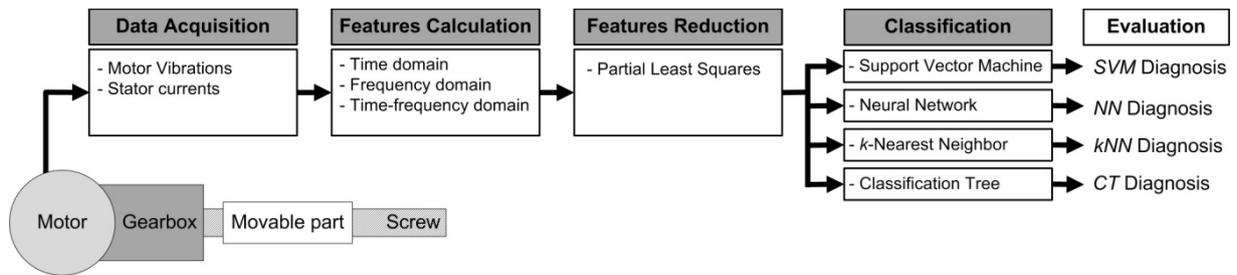


Fig. 1. Diagnosis methodology diagram: data acquisition, features calculation, features reduction and classification stages over an electro-mechanical system.

Data acquisition

The multidimensional acquisition, from a different physical magnitudes point of view, contributes to increase the observation capability. Machine vibrations and stator currents have been deeply analyzed in order to probe its individual diagnosis aptitudes. However, some studies [10] conclude that although there is not a clear superior acquisition signal than the other for diagnosis proposes, the effects of most of the faults are reflected within both signals. Therefore, it is possible to obtain two sources of information for the same fault or combination of faults, which implies an increase in pattern recognition capabilities.

Features Calculation

The fusion of features from different signal domains represents a valuable source of information. The system condition is not only analyzed from different physical magnitudes, but also analyzed in time, frequency and time-frequency domains.

Statistical parameters calculated from the acquired signals in time and frequency domains are able to describe some system condition patterns [8]. Some studies have been done in order to analyze the diagnosis capabilities of specific statistical features. However, the description capabilities of this kind of feature lie in the analysis of the relations between them. A total of 25 features are proposed for each acquired signal. Fourteen features from time: Mean, Maximum value, Root mean square, Square root mean, Standard deviation, Root mean square shape factor, Square root mean shape factor, Crest factor, Latitude factor, Impulse factor, Skewness, Kurtosis, Normalized 5th and 6th moment, and eleven features from frequency: Mean power spectrum, Standard deviation power spectrum, Skewness power spectrum, Kurtosis power spectrum, Frequency center, Standard Deviation frequency, Root Mean Square frequency, Stabilization factor, Coefficient of variability, Skewness frequency and Kurtosis frequency, are used in pattern recognition tasks by different authors [11-12].

The time-frequency analysis is proposed to be based on use two advanced signal processing techniques by signal decomposition: Wavelet Packet Transform (WPT) and Empirical Mode Decomposition (EMD). These techniques are able to emphasize the non-stationary information from the raw data, without necessity to analyze and track all the possible characteristics fault harmonics in a time-frequency distribution. In both techniques the signal is continuously decomposed in embedded sub-signals containing different components of the original. Detailed information of these techniques is shown in different studies [4-5]. In this work, the original signal is decomposed by WPT in three levels using a Daubechies wavelet 'db10', which mean that fifteen wavelet frequencies bands will be obtained. Regarding the EMD analysis, the first eight IMFs are used. In order to quantify all this

information, Shannon entropy of every wavelet frequency band and IMF is proposed as time-frequency features [13].

Features Reduction

The Partial Least Squares (PLS) represents a feature reduction technique used to extract a set of linear combinations from the original input features set [6]. Therefore, a new reduced features set can be generated. PLS analyzes directions that have high variance in the feature space and high correlation with the response. In this sense, unlike other techniques as Principal Component Analysis or Multiple Linear Regression, the PLS analyzes both the features set and the observed response values.

PLS build a linear model specifying the linear relationships between dependent responses Y (classes) and a set of predictor variables, the feature set X. The PLS model is represented as (1).

$$Y_{n \times m} = X_{n \times p} \times B_{p \times m} + E_{n \times m} \quad (1)$$

Where Y is an n observations by m response variables matrix, X is an n observations by p predictor variables features matrix, B is a p by m PLS components regression matrix, and E is a noise and error matrix with the same dimensions that Y. This model is done by PLS producing a weight matrix W, which is applied over the X matrix in order to generate a factor score matrix $T=WX$ which maximizes the covariance between responses Y and the factor score matrix. Different algorithms are used to analyze the regression between Y and T producing the matrix Q which corresponds to the Y weights such that $Y=TQ+E$. Taking into account that $B=WQ$, the model is then complete. The SIMPLS algorithm is one of the most used algorithms for computing PLS regression components [14], and it has been proposed to be used in this work.

Classification

Following the diagnosis methodology, the reduced feature set is used to supply the classification stage. As has been mentioned before, four classifiers are implemented and evaluated: Support Vector Machines, Neural Network, k -Nearest Neighbor and Classification Trees.

Support Vector Machines

Some classifiers designs seek to construct a complete model from the training features set able to solve properly every particularity. This procedure becomes too rigid and complex the classifier due to the overfitting effect. New data with slight differences can be not properly classified. The SVM structure risk minimization principle, however, is based on a compromise between the fitness of the model regarding the presented inputs, and the classification success over the training data [7].

Another important SVM characteristic is the classification strategy, which is approached as a two-class problem whose main objective is to separate the two classes by a line or hyperplane. As it is shown in figure 2a, the samples of two classes in a n -dimensional space can be separated by different lines (H1, H2, H3, H4, and others). However, from all the possible boundary lines, there is only one that leads to largest margin on both sides. This line is the optimal separating hyperplane (H2 in figure 2a). The boundary is then placed in the middle of this margin between two classes. The nearest sample points are used to define the margin (H+, H-) and are known as support vectors. Once the support vectors are selected, the rest of the feature set can be discarded. During the SVM design, kernel function is an important parameter to be selected, because projects the input set into a high-dimensional feature space, providing more possibilities to construct the optimal separating hyperplane.

Due to the binary nature of SVM, the multi-class SVM is a non-totally solved research problem. Currently, there are several methods that have been proposed for multi-class classification [15], such as *one-versus-one*, *one-versus-rest* or others methods based on *tree-structured-graph*. However, *one-versus-rest* is the simplest strategy to obtain multi-class SVM classifier. It constructs a chain of two-class SVMs sequentially executed. Each SVM is trained with all training samples but specialized only in one considered class. The training and test stages of this method are usually very slow, because all the samples are used in every SVM however, the application time is low.

A great deal of SVM applications has been reported. In [7] the authors use a multivariable SVM formed by *one-versus rest* strategy to determine the number of short-circuited turns in a PMSM. A total accuracy of 100% is achieved under controlled conditions.

Neural Network

This kind of structures is based on individual neurons, which evaluate the input set of values and generate an output value depending on the associated transfer function. The neurons are grouped in layers, which provide the capability of relation between the input features vectors and intermediate values by means of weights, in order to finish with a proper class assignment. The combination of different neuron functions in different neuron layers provides a wide variety of classifier configurations. Different NN designs have been presented such as recurrent networks or adaptive neuro-fuzzy approaches. However, a feed-forward Neural Network with an input, two hidden and an output layers, offers enough mathematical capabilities to cover a wide range of nonlinear classes distribution [16]. Therefore it is proposed in the present work the use of a feed-forward neural network structure as it is shown in figure 2b. In this work, a first hidden layer with 60 neurons of linear transfer functions, and a second layer with 30 neurons of hyperbolic tangent sigmoid transfer functions have been used. The used training technique has been the backpropagation algorithm, which allows training the network with a representative set of input-targets reaching good classification results [16].

The NN diagnosis capabilities have been widely proved. In previous works as [8], simple NN based on multilayer perceptron is used to detect four fault conditions of three phase induction motor. Various learning rules and transfer functions are investigated for different number of hidden layers obtaining reasonable classification ratios of 98% approximately.

k-Nearest Neighbor

The *k*NN leads to classify input data according to the most similar class. The *k*NN takes as a reference a set of points θ in a *n*-dimensional space, where *n* implies the number of features describing each point in θ . An input data *y* is then evaluated by a distance function in order to analyze the nearest neighbors in θ . The class which collects a specified *k* number of nearest neighbors will be the assigned class for input *y* as it is represented in figure 2c. The *k* value is completely up to user. Generally, *k*=10 gives proper results in the classification process.

Any distance measurement can be used for distance calculation of new inputs *y*. The most known and used is the Euclidian distance, that is able to evaluate the closeness of two instances θ_i and *y*. All the features f_n which describe the instances are considered to obtain the value $dst(\theta_i, y)$ as in (2).

$$dst(\theta_i, y) = \sqrt{\sum_{n=1}^N (f_n(\theta_i) - f_n(y))^2} \quad (2)$$

The *k*NN algorithm has been used in different diagnosis works, as in [7] where previously to the application of SVM to distinguish the severity, *k*NN is used to determine the presence of the fault.

Classification Trees

The classification by CT consists in classify an input vector *y* of *n* features through a chain of binary decisions. Starting at the root node and proceeding down the tree, nodes of the form $f_n(y) < T_{hr}$ are evaluated to determine if the input goes in one or other direction. Each node is called a splitting rule, and $f_n(y)$ and T_{hr} are the associated feature and threshold value respectively. Figure 2d shows a CT representation dividing the feature space in three classes by three splitting rules. During the training process of the CT, the tree is implemented by recursively finding splitting rules. These rules split the data into two parts with maximum homogeneity, until all terminal nodes represent a class or cannot be split further. The feature and the threshold at a node are determined by optimizing an impurity function. One of the most node impurity functions used in CT is based on *Gini's* criterion used also in this work [17].

CT has been used for classification core implementation in different diagnosis systems as in [9], where classification ratios between 89% and 100% are obtained in a system based on induction motor with multiple faults using statistical time-domain features.

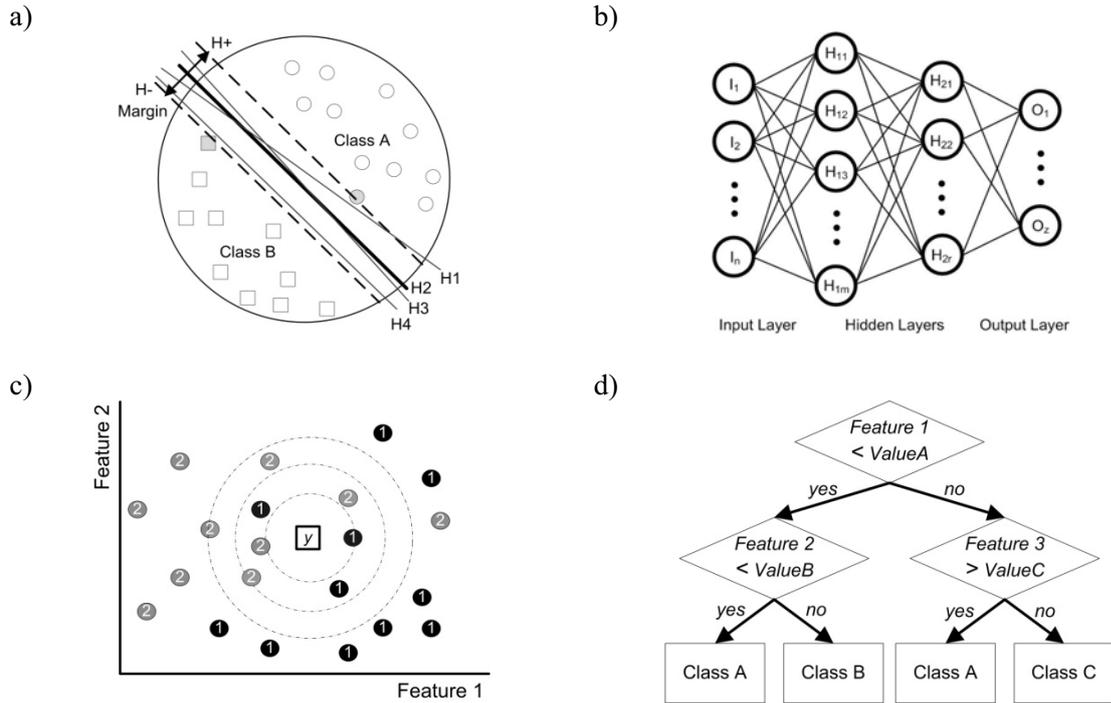


Fig. 2. a) Training procedure in a SVM classifier, b) Feed-forward Neural Network structure, c) k NN representation between two class features vectors, with a selected k value equal to 5 the y input is assigned to class 2, d) Binary decision tree based on three splitting rules for three classes problem.

Experimental Test-Bench

In order to evaluate the behavior of the difference classifiers proposed in this work, experimental data has been obtained from a test bench based on an electro-mechanical actuator. The experimental motor bench is based on two identical featured face to face motors, the motor under test and the motor that acts as a load. These motors are connected by means of a screw and a gear box, constituting the test bench. A motor runs the input axle of the gear box. The output axle of the gear box runs the screw, which in turns, displace the movable part. The experimental arrangement diagram is shown in figure 3. A set of six fault scenarios, between healthy, single and combined faults, have been considered to complete the experimental arrangements. First of them, a complete healthy electro-mechanical actuator have been tested. Second, partially demagnetized motor was developed during the manufacture with a 20% of nominal flux reduction in one pair of poles. Third, degraded bearings have been mounted. The non-end bearing inner as well as outer races have been scraped thoroughly in order to cause a generalized rough defect. Fourth, two gear teeth have been smoothed out to impose a degradation degree on the reduction box. Fifth, a combined fault scenario including the aforesaid motor demagnetization and gearbox teeth defects have been implemented at the same time in the test bench. Finally, the sixth fault scenario has been proposed combining the motor bearing and the gearbox teeth defect. Each of these fault scenarios have been experimentally reproduced following two different speed patterns (a ramp from 1500 to 4500 rpm with a slope of 12 rpm/ms, as well as a sinusoidal speed pattern from 3000rpm to 6000rpm at 0.5Hz), combined with five stationary and non-stationary torque patterns (0%, 50%, 100% and 110% of rated torque and a ramp with a slope of 0.01Nm/ms). Therefore, ten different conditions have been tested for each fault scenario considered. For each of the experimental cases, sixty complete acquisitions were made. As it is shown in figure 4, the first 40 acquisitions are used for classifiers development and training, and the last 20 for test purposes. Additionally, in order to test the generalization capabilities of the classification algorithm, it has been removed from the training and test sets the experimental conditions related with 0% and 110% of rated torque, in order to be used as new input conditions once the classifier is completely built.

Regarding the SVM, NN, k NN and CT classifiers as well as the PLS reduction algorithm have been carried out in a Matlab environment applying different Statistical toolbox packages.

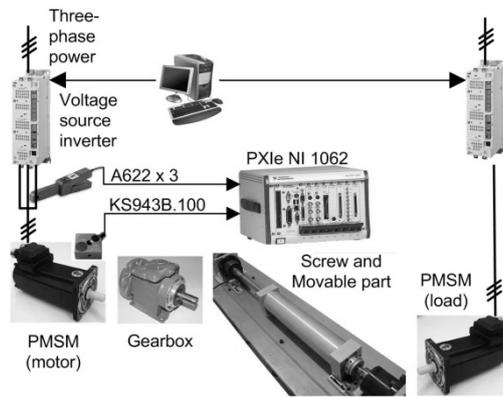


Fig. 3. Experimental arrangement diagram including inverters, acquisition system, sensors and electro-mechanical test bench.

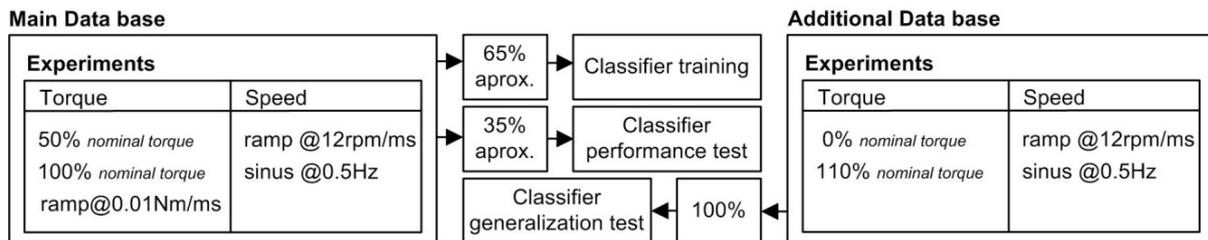


Fig. 4. Experiments and data base organization: first, the classifier is trained with a part of the main data base. The classifier performance is obtained by similar inputs to the learnt ones. After, generalization performance is obtained by specific experiments with some modifications regarding the used to train the classifier.

Experimental Results

Once the physical magnitudes are acquired, and the aforesaid specific features from time, frequency and time-frequency domains are calculated, a matrix formed by features vectors describing the considered output classes is obtained. In this point, the feature reduction process is executed from a general point of view, considering all the features and their descriptive relations regarding all the considered classes.

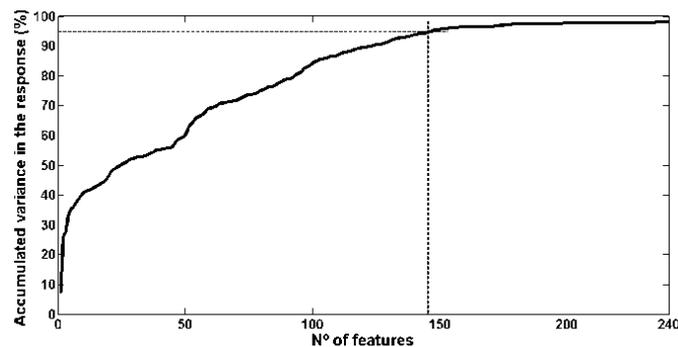


Figure 5. Number of features (PLS components) and variance in the training response significance.

It can be seen in figure 5 that analyzing with PLS the original descriptive vectors, composed by 240 features, the first 145 PLS components show nearly 95% of the variance in the response. Therefore, most of the information calculated from the electro-mechanical actuator can be compressed in a reduced features set of 145 components. This analysis it should not be interpreted in the sense that the whole system can be described with only 145 calculated features. The 145 PLS components are a compressed representation of the whole features set. That is, an addition or extraction of a feature in the original features set will lead to a modification in the number and/or magnitude of the reduced features set. By means PLS, the original features are mathematically treated during the feature reduction process. The physical meaning that the features have in the initial set is totally lost in the

final reduced set. Once a more informative and manageable features set is obtained, the classification algorithms can be designed and trained. The more important configuration parameters for each of the proposed classifiers is collected in Table 1.

Table I. Classifiers configuration parameters.

Classifier	Main configuration parameters	
Support Vector Machine	Multivariable structure Kernel function Parameter C Parameter γ	<i>One versus rest</i> <i>Radial Basis Function</i> 1 2
Neural Network	Neural Network structure Training algorithm Number of hidden layers Number of neurons in hidden layer 1 Neuron functions in hidden layer 1 Number of neurons in hidden layer 2 Neuron functions in hidden layer 2	<i>Feedforward</i> <i>Backpropagation</i> 2 60 <i>linear</i> 30 <i>hyperbolic tangent sigmoid</i>
k -Nearest Neighbor	Distance function Parameter k	<i>Euclidian</i> 10
Classification Tree	Cost function Minimum observation number for split	<i>SIMPLS</i> 20

In order to increase the available information of the classifiers, the v -fold-cross-validation has been used during the training process [18]. Instead the training of the classifiers by training and test sets, in v -fold-cross-validation, the training set is divided in v subsets of equal size (containing the same number of inputs vectors for each class). Sequentially, each subset is tested using the classifier trained on the remaining ($v-1$) subsets. Thus, each instance of the whole training set is predicted once so the cross-validation accuracy is the mean percentage of data that have been correctly classified. Assuming a balanced training data base, the v -fold-cross-validation shows the classifier stability since it is v times trained and tested. After the classifiers cross-validation, with a proposed $v = 10$ in this work, the classifier is trained again with the whole training set.

In figure 6 it can be seen the classification accuracies for each of the classification algorithms as well as the v -fold-cross-validation results.

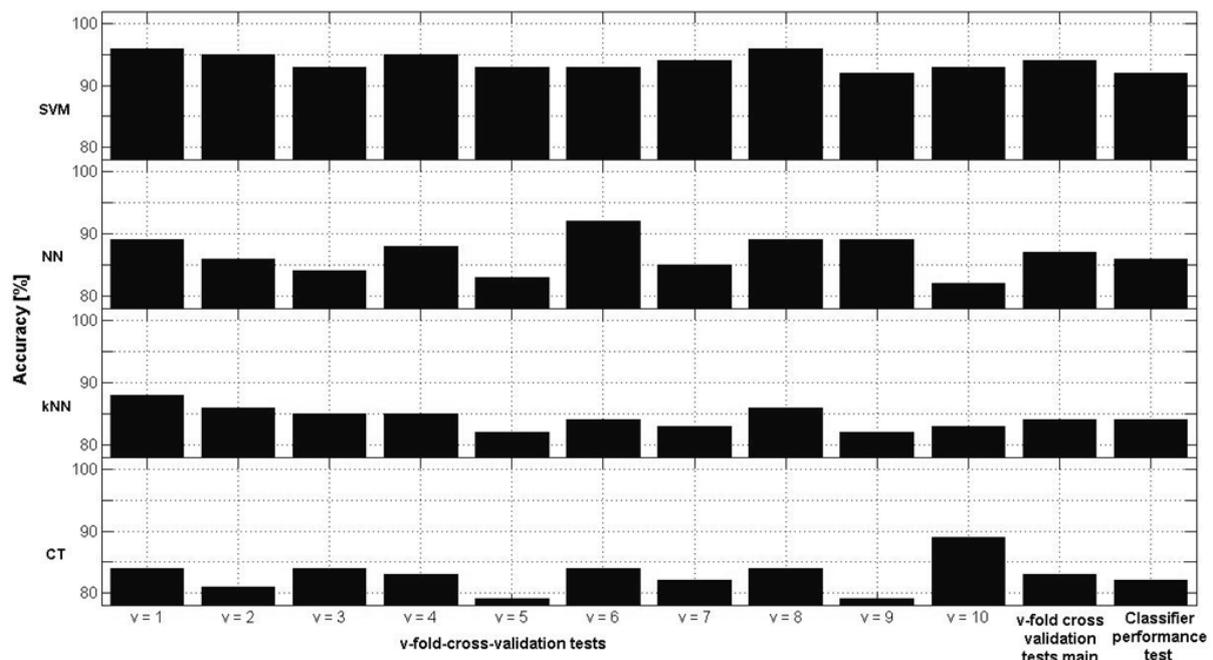


Figure 6. Classification accuracies of SVM, NN, k NN and CT during the v -fold-cross validation and the performance test.

It can be seen in figure 6, that all the classifiers considered exhibit a constant accuracy during the cross-validation process. Variations of the accuracy in the range of $\pm 8\%$ during the folds tests, regarding the cross validation mean, imply a proper stability of the classifiers and the use of a well-distributed data base. Moreover, the classifiers performances tests, formed by a larger set of inputs, show similar accuracies than the cross validation tests for each classifier, which reinforce the classifiers robustness observed during the cross validations tests. Regarding the individual behaviors, the SVM exhibits the biggest performance with a 94% approximately of correct classifications. The NN classifier is the second classifier with best results, an 86% of accuracy is reached. The CT and the k NN classifiers are both under the 85% of accuracy.

In order to complete the experimental results, each classifier has been tested with another set of inputs describing learnt classes but under lightly differences. The generalization set consist in 60 samples of each fault scenario under the known two speed patterns combined with two new torque patterns (0% and 110% of rated torque). The obtained results for this generalization tests have been: 61% for SVM classifiers, 39% with NN, 54% in the k NN case and 48% using CT classifier. Taking into account the dissimilarities between the training and the generalization sets, since more similar cases can be chosen, the generalization capabilities of the classifiers are quite significance. However, specific classifiers adjustments should be done in this direction to achieve proper generalization ratios. Analyzing the results, it can be seen that some of the classifiers, as NN, suffer from overfitting, decreasing the generalization capabilities.

In table II it has been collected the main observations regarding the classifiers and their performance and generalization capabilities.

Table II. Main observations of the classifiers.

Classifier	Observations
Support Vector Machine	<ul style="list-style-type: none"> - The binary classification nature together with the risk minimization principles allows a comfortable class analysis. More singularities can be considered without overfitting. - In electromechanical faults diagnosis systems, some classes can exhibit similar patterns. Due to the SVM multiclass structures, the first classes analyzed have more possibilities to be selected.
Neural Network	<ul style="list-style-type: none"> - A great deal of configurations possibilities which makes difficult to achieve an optimized algorithm. However, standard configurations reach important adaptability being able to assume different classes. - The classification understanding is lost.
k -Nearest Neighbor	<ul style="list-style-type: none"> - It represents the best tradeoff between simplicity-understanding and classification accuracy. - Regarding electromechanical systems, it represents the simplest classifier algorithm to be implemented in a digital processor for on-line monitoring.
Classification Tree	<ul style="list-style-type: none"> - It is the most feature reduction dependent classification algorithm. - Although the training process is usually done in stand-alone stations, its implementation in a digital processor is simple.

Conclusions

Different machine learning algorithms have been presented and implemented in a complete diagnosis scheme formed with data acquisition, features calculation, features reduction and classification stages. An experimental electro-mechanical actuator has been used for evaluation proposals. High classification requirements have been imposed, since not only multiple classes are considered but also different variations (speed and torque patterns) are included. Although NN, CT and k NN have proper classification results, SVM exhibit the best classification and generalization capabilities. The classification algorithms have been adjusted by classical configurations; however optimized classifiers designs, adjusted to the applications requirements, could improve the general performance. One of the main classifiers weak points is the magnitude of classification space. For that reason, the specialization and collaboration between classifiers will improve the general performance of diagnosis systems, and is one of the future trends in classification that it should be explored.

References

- [1] Romeral L., Rosero J.A., Espinosa A.G., Cusido J. and Ortega J. A.: Electrical monitoring for fault detection in an EMA, *IEEE Aerospace and Electronic Systems Magazine*, Vol. 25, no. 3, pp. 4-9, March 2010.
- [2] Urresty J., Riba J.-R.; Ortega, J. A., Cardenas, J.: Stator short circuits detection in PMSM by means of Zhao-Atlas-Marks distribution and energy calculation, *EPE*, pp. 1-9, Sept. 2009
- [3] Garcia-Perez A., Romero-Troncoso R. J., Cabal-Yepez E. and Osornio-Rios R. A.: 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.
- [4] Khan M. and Rahman M.A., Development and Implementation of a Novel Fault Diagnostic and Protection Technique for IPM Motor Drives, *IEEE Trans. Ind. Electron.*, Vol. 56, no. 1, pp. 85-92, Jan. 2009.
- [5] Espinosa A.G., Rosero J.A., Cusido J., Romeral L. and Ortega J.A.: Fault Detection by Means of Hilbert–Huang Transform of the Stator Current in a PMSM With Demagnetization, *IEEE Trans. Energy Conversion*, Vol. 25, no. 2, pp. 312-318, June 2010.
- [6] Jianxiao G., Hongli W., Yarong G. and Zhiwen Z.: A New Data Mining Method of Iterative Dimensionality Reduction Derived from Partial Least-Squares Regression, *IITA*, Vol. 2, pp. 471-474, Nov. 2009.
- [7] Mahdi Ebrahimi B. and Faiz, J.: Feature Extraction for Short-Circuit Fault Detection in Permanent-Magnet Synchronous Motors Using Stator-Current Monitoring, *IEEE Trans. Ind. Power Electron.*, Vol. 25, no. 10, pp. 2673-2682, Oct. 2010.
- [8] Ghate V.N. and Dudul S.V.: Fault Diagnosis of Three Phase Induction Motor Using Neural Network Techniques, *ICETET*, pp. 922-928, Dec. 2009.
- [9] Ngoc-Tu Nguyen DD., Jeong-Min Kwon and Hong-Hee Lee: Fault diagnosis of induction motor using decision tree with an optimal feature selection, *ICPE*, pp. 729-732, Oct. 2007.
- [10] Immovilli F., Bellini A., Rubini R. and Tassoni C.: Diagnosis of Bearing Faults in Induction Machines by Vibration or Current Signals: A Critical Comparison, *IEEE Trans. On Instr. and Meas.*, Vol. 46, no. 4, pp. 1350-1359, 2010.
- [11] Xu Z., Xuan J., Shi T., Wu B. and Hu Y.: Application of a modified fuzzy ARTMAP with feature-weight learning for the fault diagnosis of bearing, *Expert Systems with Applications*, Vol. 6, no. 36, pp. 9961-9968, Aug. 2009.
- [12] Lei Y., He Z. and Zi Y.: Application of an intelligent classification method to mechanical fault diagnosis, *Expert Systems with Applications*, Vol. 36, Iss. 6, pp. 9941-9948, 2009.
- [13] Zheng-You H., Xiaoqing C. and Guoming L.: Wavelet Entropy Measure Definition and Its Application for Transmission Line Fault Detection and Identification; (Part I: Definition and Methodology), *ICPST*, pp. 1-6, Oct. 2006.
- [14] de Jong S.: SIMPLS: An alternative approach to partial least squares regression, *Chemometrics Intell. Lab. Syst.*, Vol. 18, no. 3, pp. 251–263, Mar. 1993.
- [15] Guo J., Chen Y., Zhu M., Wang S. and Liu X.: An Efficient Support Vector Machine Algorithm for Solving Multi-class Pattern Recognition Problems, *ICCMS*, Vol. 2, pp. 461-465, Jan. 2010.
- [16] Sadeghian A., Zhongming Y. and Wu B.: Online Detection of Broken Rotor Bars in Induction Motors by Wavelet Packet Decomposition and Artificial Neural Networks, *IEEE Trans. Instrum. Meas.*, Vol. 58, no. 7, pp. 2253-2263, July 2009.
- [17] Sakthivel N.R., Sugumaran V. and Babudevasenapati S.: Vibration based fault diagnosis of monoblock centrifugal pump using decision tree, *Expert Systems with Applications*, Vol. 37, Iss. 6, June 2010.
- [18] Hsu C. W. and Lin C. J.: A comparison of methods for multiclass support vector machines, *IEEE Trans. on Neural Network*, Vol. 13, Iss. 2, pp. 415-425, 2002.