

# Vibration-based adaptive novelty detection method for monitoring faults in a kinematic chain

Jesus A. Carino <sup>a</sup>, Juan J. Saucedo-Dorantes <sup>b</sup>, Daniel Zurita <sup>a</sup>, Miguel Delgado <sup>a</sup>, Juan A. Ortega <sup>a</sup>, Roque Osornio-Rios <sup>b</sup> and Rene Romero-Troncoso <sup>c†</sup>.

<sup>a</sup> MCIA Research Center, Department of Electronic Engineering, Technical University of Catalonia (UPC). Rbla. San Nebridi n° 22, Gaia Research Building, 08222 Terrassa, Barcelona, Spain.

<sup>b</sup> CA Mecatronica, Facultad de Ingenieria, Campus San Juan del Rio, Universidad Autonoma de Queretaro, Rio Moctezuma 249, Col. San Cayetano, 76807 San Juan del Rio, Queretaro, Mexico.

<sup>c</sup> CA Procesamiento Digital de Señales, CA Telematica, Division de Ingenierias, Campus Irapuato-Salamanca, Universidad de Guanajuato, Carretera Salamanca-Valle km 3.5+1.8, Comunidad de Palo Blanco, 36700 Salamanca, Guanajuato, Mexico.

[jesus.carino@mcia.upc.edu](mailto:jesus.carino@mcia.upc.edu), [miguel.delgado@mcia.upc.edu](mailto:miguel.delgado@mcia.upc.edu), [jsaucedo@hspdigital.org](mailto:jsaucedo@hspdigital.org),  
[daniel.zurita@mcia.upc.edu](mailto:daniel.zurita@mcia.upc.edu), [juan.antonio.ortega@mcia.upc.edu](mailto:juan.antonio.ortega@mcia.upc.edu), [raosornio@hspdigital.org](mailto:raosornio@hspdigital.org)

† Corresponding author: [troncoso@hspdigital.org](mailto:troncoso@hspdigital.org)

## Abstract

This paper presents an adaptive novelty detection methodology applied to a kinematic chain for the monitoring of faults. The proposed approach has the premise that only information of the healthy operation of the machine is initially available and fault scenarios will eventually develop. This approach aims to cover some of the challenges presented when condition monitoring is applied under a continuous learning framework. The structure of the method is divided in two recursive stages. First, an offline stage for initialization and retraining of the feature reduction and novelty detection modules; and second, an online monitoring stage to continuously assess the condition of the machine. Contrary to classical static feature-reduction approaches, the proposed method reformulates the features by employing first a Laplacian Score ranking and then the Fisher Score ranking for retraining. The proposed methodology is validated experimentally by monitoring the vibration measurements of a kinematic chain driven by an induction motor. Two faults are induced in the motor to validate the method performance to detect anomalies and adapt the feature reduction and novelty detection modules to the new information. The obtained results show the advantages of employing an adaptive approach for novelty detection and feature reduction making the proposed method suitable for industrial machinery diagnosis applications.

**Keywords:** Artificial intelligence; novelty detection; feature-reduction; vibration analysis; condition monitoring.

## 1. Introduction

Condition-based maintenance has become accepted by the industrial manufacturing sector as a key factor to achieve the currently required high-competitiveness levels. In the field of electromechanical systems monitoring, an increased number of approaches of health monitoring schemes have been proposed during the last decades[1]–[3]. These methodologies have demonstrated to be a reliable option as fault detection techniques when data of the possible faults is available.

Nevertheless, in order to overcome the practical requirements of an industrial implementation, some considerations must to be taken into account. There is a necessity to develop methodologies capable of monitoring complicated industrial machinery, that is, diagnosis approaches capable of consider complete electromechanical kinematic chains including electric motors, gearboxes and additional mechanical elements. It is important to mention that information regarding the possible faults that could occur on complicated machines is not usually available due to the high cost and the time consuming it could imply obtaining this information.

Unexpected events in the form of faulty scenarios, not initially considered, or deviations over the normal operation take place during the useful life of the machinery. These situations must be

identified in order to avoid misclassifications and the consequent incorrect maintenance action. To cope with these challenging scenarios, the condition-based maintenance methods must be able to work under the assumption that only information of the healthy condition of the machine is available [4]. Another relevant characteristic is the capacity to incorporate new knowledge. Since new faults could appear during monitoring, the condition-based maintenance methods must also be able to adapt and include new scenarios to their initial knowledge [5], [6].

These challenging scenarios are not always taken into consideration in condition monitoring methodologies; therefore, the application of classical methods on an industrial environment is limited or non-viable.

In the field of electromechanical system monitoring, a great deal of health monitoring schemes based on vibrations have been proposed during the last years [7]–[10]. Mainly, these proposals can be divided in two categories regarding the available information of the monitored machine and the classification objective.

The first category, multi-fault detection, consists in the development of a monitoring system capable of identifying multiple faults previously known [1], which implies that the information available to train the monitoring system include measurements of the healthy state of the machine but also measurements of the machine working under different faulty conditions. Over the last years, instead of looking for highly specific features for a component fault, a trend towards calculation of more general statistical features and the fusion of information to enhance the performance of the diagnosis system has been carried out [11]. Yet, this approach is mostly carried out at the laboratory scale with successful results where controlled faults are provoked on the machine.

However, the information of the possible faults that could appear on a machine is not always available; consequently, a different approach must be considered if only information of healthy condition of the machine is available. The second category corresponds to novelty detection, also known as anomaly detection, and it represents a solution to cope with the lack of information available of the monitored machine. The objective of novelty detection is to detect novel events that differ in some manner from the information of which the model was trained. Novelty detection framework has been successfully applied to network intrusion, medical diagnosis, image segmentation and handwritten digit recognition, fault detection and condition monitoring [4], [6], [12]–[14]. Special effort has been made to improve condition monitoring novelty detection applications based on vibrations. Ma et al., (2015), proposed a novelty detection approach to rotating machinery by means of monitoring the thresholds obtained from an extreme value distribution [15]. McBain and Timusk (2011), proposed a feature extraction methodology for novelty detection applied to a rotating mechanical system [16].

Most of the methodologies for novelty detection are limited to a static analysis and the incorporation of the novel information to the novelty detection system is not usually considered. An approach to include adaptability to the novelty framework, based on vibrations, is proposed by Wang et al., (2013) [17]. The proposed monitoring scheme include testing data on the boundary of the novelty model, based on Support Vector Data Description, and retrain the model with this information to gain robustness. Nevertheless, this approach does not take into consideration the possibility to include novel scenarios during the monitored process.

Other approaches to develop an adapting condition monitoring scheme were presented by Filev et al., (2010) where a practical framework for autonomous monitoring of industrial equipment based on novelty detection is analyzed [18]; and by Costa et al., (2015), where a two-stage algorithm for real-time fault detection and identification is presented [19]. Both approaches provide the opportunity to incorporate novel detected faults to the monitoring system, nevertheless, in both methods the incorporation is limited to update the known data base, but an adaptation of the numerical features analyzed is not considered.

The performance of a novelty detection system is highly dependent on the numerical features considered. When there is no previous information of the possible faults that can occur, the application of a suitable numerical features analysis strategy represents a critical challenge [4]. Considering a continuous monitoring framework where the initial information available is the healthy operating condition and, later on, different faults are identified progressively when the

machine condition deteriorates, all the approaches previously discussed do not modify the initial set of features when new information of faults is incorporated. This static approach have the advantage of providing a most adequate situation for on-line adaptation, nevertheless, analyzing the information of the faults detected during the monitoring process could improve the identification of a most adequate set of features to discriminate the possible upcoming or already detected fault scenarios.

To improve the performance of the novelty detection task a dynamic approach for reselection of the features including information of new faults detected is proposed on this work. To the extent of the authors' knowledge, the study regarding the dynamic change of the feature space by considering new operating scenarios is novel in the applications to electromechanical systems for fault diagnosis.

Thereby, the contribution of this study is to provide an adaptive methodology for novelty detection where the information of identified faults during the monitoring process is exploited to improve the anomaly detection task. The methodology is divided in two recursive stages. First, an offline stage for initialization and retraining of the feature reduction and novelty detection modules; second, an online monitoring stage to continuously assess the condition of the machine. The novelty model employed is a one-class Support Vector machine (OC-SVM) [20], which is a domain based model successfully employed on vibration signals monitoring [4]. Contrary to classical static feature reduction approaches, the proposed method reformulates the features by employing first a Laplacian Score ranking and then the Fisher Score ranking for retraining.

The proposed methodology is tested using real data from a laboratory based kinematic chain, where faults are induced, and the vibration signals from two accelerometers are monitored to assess the system condition. To highlight the benefits of the methodology proposed, a comparison is performed of classical feature reduction approaches to the adaptive feature reduction scheme proposed.

This paper is organized as follows. In Section 2, theoretical aspects of the proposed method are described. Section 3 describes the method presented. The experimental setup employed for assessment and the results obtained are presented and discussed in section 4. And finally, conclusions and future work are summarized in Section 5.

## 2. Theoretical Background

### 2.1. One-Class Support Vector Machine

As mentioned, the objective of novelty detection is to detect abnormal events that differ in some manner from the information of which the model was trained. The one-class Support Vector machine (OC-SVM), which is a domain based model successfully employed in condition monitoring applications, encloses the training data by means of a boundary or threshold which represents the limit of normality. The OC-SVM presents the possibility to include a kernel formulation, providing it with enough flexibility to adapt to the distribution of the training data. To assess the condition of the machine under test, new measurements are analyzed regarding their position to the boundary created in the training phase. The test measurements underlying inside the boundary are considered “normal” or “known” while the measurements underlying outside the boundary are considered “outlier” or “novel”. Compared to other novelty detection techniques, the OC-SVM is robust to outliers during the training phase, but has the disadvantage of presenting several configuration parameters, which are described later in this section.

The OC-SVM was proposed by Schölkopf et al. (1999), for estimating the support of a high-dimensional distribution [20]. The OC-SVM classification objective is to separate one class of target samples from all other class samples. In this type of problem one class is characterized properly, called target class; while for the other class, usually, no measurements are available. Considering  $X = [x_1, \dots, x_N]^T \in R^{N \times M}$ , which denotes the normal data set, and  $x_i$ ,  $i=1..N$ , denotes training samples (available measurements), characterized by  $M$  numerical features, then, in order to obtain the boundary, an optimization model is considered as follows

$$\min \left\{ \frac{\|\mathbf{w}\|^2}{2} + \frac{1}{N\nu} \sum_{i=1}^N \xi_i - \rho \right\} \quad (1)$$

Subject to

$$\mathbf{w} \cdot \Phi(x) \geq \rho - \xi_i, \quad \xi_i \geq 0$$

where  $\nu$  is a regularization parameter and  $\xi_i$  is the slack variable for the point  $x_i$ . The constants  $\mathbf{w}$  and  $\rho$  are the normal vector and offset of the hyperplane, respectively. Thus, the decision boundary can be formulated as

$$f(x) = \mathbf{w} \cdot \Phi(x) - \rho \quad (2)$$

where  $x \in R^M$ , and  $\Phi$  is a higher dimensional projection vector. For the classification problem of two categories, the data sets are not always linearly separable in the original space, then,  $\Phi$  projects the original data sets into a higher dimensional space, the so-called feature space, where the data sets can be linearly separable. However,  $\Phi$  is inexplicit in the practical application, and only the dot product from  $\Phi(x_i) \cdot \Phi(x_j)$  is necessary to be known.  $K$  represents the kernel function  $\Phi(x_i) \cdot \Phi(x_j)$ . The most commonly used kernel functions is the Gaussian

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \quad (3)$$

In order to solve the optimization problem (1), Lagrange multipliers  $a_i \geq 0$  and  $\beta_i \geq 0$  ( $i=1, \dots, N$ ) are introduced and the Lagrange equation is formed as

$$\begin{aligned} L(\mathbf{w}, \boldsymbol{\xi}, \rho, \mathbf{a}, \boldsymbol{\beta}) &= \frac{\|\mathbf{w}\|^2}{2} + \frac{1}{N\nu} \sum_{i=1}^N \xi_i - \rho \\ &- \sum_{i=1}^N a_i (\mathbf{w} \cdot \Phi(x_i) - \rho + \xi_i) - \sum_{i=1}^N \beta_i \xi_i \end{aligned} \quad (4)$$

The partial derivatives of the Lagrangian equation with respect to  $\mathbf{w}$ ,  $\boldsymbol{\xi}$  and  $\rho$  are set to zero. Then,  $\mathbf{w}$  and  $a_i$  can be formulated as

$$\mathbf{w} = \sum_{i=1}^N a_i \Phi(x_i) \quad (5)$$

$$a_i = \frac{1}{N\nu} - \beta_i \quad \sum_{i=1}^N a_i = 1 \quad (6)$$

Substitute (5)-(6) into Lagrangian equation (4) and its dual form is presented as

$$\min \mathbf{a}^T \mathbf{H} \mathbf{a} \quad (7)$$

$$\text{subject to } 0 \leq a_i \leq \frac{1}{N_V} \quad \sum_{i=1}^N a_i = 1$$

where  $\mathbf{a} = [a_1 \dots a_N]^T$ , and  $H$  is the kernel matrix and the factor of  $H$ , i.e.  $H_{ij}$ , which can be expressed as:

$$H_{ij} = K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \quad (8)$$

Solve the optimization problem (7) to get  $\mathbf{a}$  and then  $\rho$  can be given as

$$\rho = \frac{1}{n_s} \sum_{i=1}^{n_s} \sum_{j=1}^N a_j K(x_i, x_j) \mathbf{a} \quad (9)$$

where  $n_s$  is the number of support vectors.

Tax and Duin (1999), proposed another form of OC-SVM which is called Support Vector Data Description [21]. The basic idea of SVDD is to construct a minimum-volume hypersphere in a high dimensional feature space to enclose as much as normal data points. Both of these two forms of SVM have equivalent solution if the diagonal entries of kernel matrix  $H$  are equal to a constant.

## 2. 2. Dimensionality Reduction Techniques

Working with high dimensional datasets could complicate the learning part of novelty detection methods, not only because of possible presence of noise and redundancy in the data but for other reasons as well [22]. The empty space phenomenon states that to cover the whole space it is needed a number of samples that grows exponentially with dimensionality. The curse of dimensionality implies that in order to learn successfully, it is needed a number of training examples that also grows exponentially with the dimensionality. The ‘‘concentration of measure’’ phenomenon seems to render distance measures not relevant to whatever concept is to be learnt as the dimension of the data increased. For these reasons, among others, there is a necessity to apply dimensionality reduction techniques in novelty detection applications. Dimensionality reduction strategies differ in the question of whether the learning process is supervised or unsupervised. The difference between both learning processes is the availability of labels to distinguish the different classes.

Principal Component Analysis (PCA) is one of the most commonly used technique for unsupervised dimensionality reduction [23]. It aims to find the linear projections that best capture the variability of the data.

Another well-known dimensionality reduction technique is the Laplacian Score (LS), where the merit of each feature is measured according to its locality preservation power [24]. A nearest neighbor based graph is constructed from the training set, and analyzed to rank each feature individually according a weighting approach selected for the graph edges. If the labels of the classes of the analyzed dataset are given, the method changes to a supervised approach and the graph is created based on these labels, instead of the nearest neighbor approach. Two variations for weighting the edges can be applied, Heat Kernel and Simple minded. To rank each feature, its Laplacian Score is computed, which is a measure of which extent the analyzed feature preserves the structure present in the graph divided by the variance of the feature. For a feature to be selected it must have a low LS, which implies high variance and locality.

Linear discriminant analysis (LDA) is one of the most well-known supervised techniques for linear dimensionality reduction in multi-class problems [25]. LDA searches attempts to maximize the linear separation between data points belonging to different classes. In contrast to most other dimensionality reduction techniques, LDA, as a feature extraction technique, finds a linear mapping that maximizes the linear class separation in the low-dimensional representation of the data. The

criteria that are used to formulate linear class separation in LDA are the within-class scatter and the between-class scatter.

Another variant based on the Fisher Coefficient is a feature dimensionality reduction approach called Fisher Score (FS), where the objective is to find a subset of features which maximize the Fisher coefficient [26]. The main difference between LDA and Fisher Score is that LDA is a feature extraction approach and Fisher Score is a feature selection approach; then, instead of extracting new features from the subspace obtained, FS ranks the features and select the reduced subspace of features selected which maximizes the dimensionality reduction criterion, in this case, the Fisher criterion.

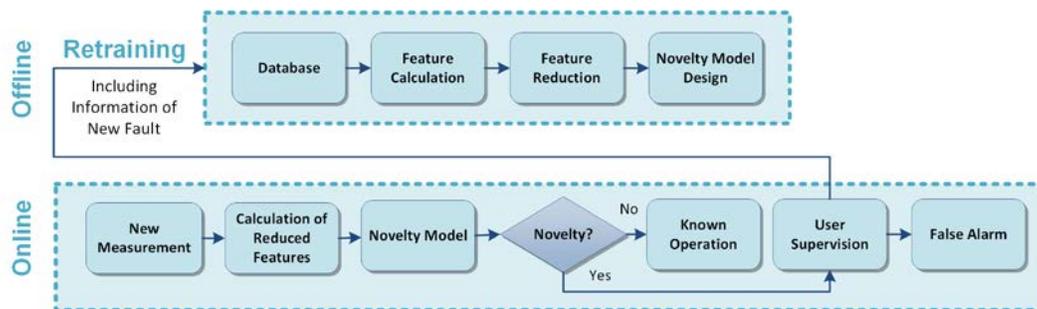
The aforementioned linear feature reduction techniques have a different combination of objective (variance preservation, topology preservation, discrimination) and method of employment (unsupervised or supervised and extraction or selection), and have been widely used in the literature with successful results[16], [22]; yet, they have been applied under a static framework where the set of features is reduced initially and no possible adaptation during the monitoring phase is considered.

### 3. Methodology

Given that the data initially available is usually related to the healthy condition of the machine under analysis, condition-based monitoring schemes must be designed to overcome two main challenges:

- (i) The identification of significant features to deal with the characterization of the known conditions of the machine, under the consideration that the occurrence of additional unknown faults must be detected.
- (ii) The adaptation of the condition-based monitoring scheme to update the considered data base of the machine, once unknown fault scenarios have been detected.

Such requirements are addressed on this work by means of the proposed adaptive novelty detection method shown in Fig. 1. This methodology represents an important step to the introduction of adaptive novelty detection schemes to the development of electromechanical system diagnosis procedures.



**Fig. 1.** Proposed methodology for the novelty detection approach. The monitoring method is composed by an offline stage for initialization and retraining, and an online stage for continuous monitoring.

The proposed method is composed by two stages: an offline stage and an online monitoring stage. The main objectives of the offline stage are, first, the analysis of the information available of the monitored machine to find a reduced set of numerical features to characterize the known machine conditions and, second, the design of the novelty model by means of the selection of the configuration parameters and training.

Once a reduced set of characteristic features is obtained and the novelty model is designed, the online stage is carried out. During the online stage, new measurements are continuously compared with the normality threshold,  $T_n$ , defined during the novelty model training in the offline stage. Thus, if a novel scenario is detected, the supervision of an expert user is proposed in order to confirm and label the new condition of the machine; consequently the monitoring system is retrained to include

the characteristics of the novel scenario. Detailed information of each stage and the retraining process is described in the following subsections.

### 3.1. Offline Stage

During the initialization, it is assumed that only information of the machine operating under healthy condition is available in the database. The first step is the calculation of numerical features from the vibration measurements obtained during the machine operation. Since the information of the possible faults of the monitored machine is not available yet during this initialization, a generic set of statistical time-based numerical features is proposed to be extracted from each available vibration axis measurement. The proposed set of potential features is shown in Table 1. These features have been successfully employed for fault detection on the last years [1], [27].

**Table 1**  
Statistical Time-Frequency Features

Max Value	$\bar{x}_k = \max(x)$	(10)
Root Mean Square (RMS)	$RMS = \sqrt{\frac{1}{n} \cdot \sum_{k=1}^n (x_k)^2}$	(11)
Crest Factor (CF)	$CF = \frac{\bar{x}_k}{RMS}$	(12)
Kurtosis	$k = \frac{E[(x_k - \bar{x})^4]}{\sigma^4}$	(13)
Shape Factor	$SF = \frac{RMS}{\frac{1}{n} \cdot \sum_{k=1}^n  x_k }$	(14)

The resulting number of numerical features is proportional to the number of available vibration axis collected during the acquisition. However, in order to allow the compression and visualization of the data, a feature reduction module is implemented. During such offline stage initialization, an unsupervised feature reduction approach must be used, a Laplacian Score Ranking is proposed in this work as a good trade-off between simplicity and performance [24], to rank the features according the topology preservation capabilities. The two or three first ranked features in terms of Laplacian score are selected.

Next, the novelty model is designed. There is a significant number of novelty models proposed on the literature [4], each one demonstrated to be a capable option under certain circumstances. An increasing amount of works imply that domain based novelty detection models present promising results [19], [28]. In this work a standard OC-SVM with Gaussian kernel is used. The design of the novelty model includes the selection of the parameters for configuration, and training employing the known scenarios stored at the database. Then, the initialization of the offline stage finalizes with the design of the OC-SVM.

### 3.2. Online Stage

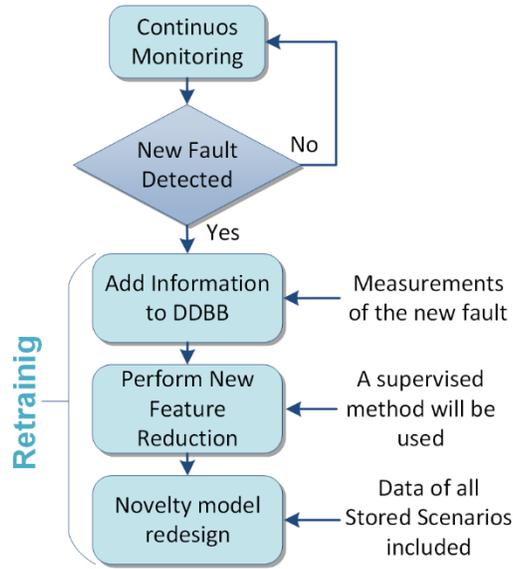
This stage continuously monitors the condition of the machine to detect if an anomaly is present. To accomplish this, new measurements of the machine are acquired each certain amount of predefined time. Each measurement is segmented and a set of features is calculated from each segment. The set of numerical features calculated on this stage are reduced in the offline stage by means of Laplacian Score.

Thus, each new measurement characterized by the numerical features is analyzed by the novelty model. In case of no novelty detection, it is assumed that the machine is working under known conditions. However, if the analyzed measurement is detected as a novelty, an alarm is triggered in

order to consider the user assessment. Then, if the occurrence of a new scenario is confirmed the corresponding measurements are stored at the database and a retraining procedure is performed.

### 3.3. Retraining

Once the retraining is triggered, the feature reduction and the novelty model design modules are modified at the offline stage. A diagram of the retraining procedure is presented in Fig. 2.



**Fig. 2.** Proposed retraining approach. First, measurements characterizing the fault are stored, and then the feature reduction module and the novelty design module are modified to incorporate the new scenario encountered.

It must be noticed that two important contributive aspects are proposed on this retraining approach, first, the reconsideration of the feature reduction module, second, the incorporation of fault scenarios to the novelty model. As it has been explained, the feature reduction module, during its initialization, is supported by an unsupervised Laplacian score ranking due to the lack of additional scenarios; yet, once the information of a fault is available, the feature reduction module could improve its performance by employing supervised methods.

Then, the Laplacian approach is replaced by a Fisher Score ranking approach in this work, where the features are sorted according the Fisher coefficient calculated from each feature. It is important to mention that Laplacian Score can be configured to work also under a supervised framework and could obtain similar results than employing a Fisher approach, similarities between the two approaches are discussed in [24]. Nevertheless, Fisher Score is used on this methodology to search for a possible discriminative scenario to increase the detection of novel scenarios. A comparison of results obtained from both features reduction approaches are presented in the results of this work.

The consideration of a faulty scenario in the novelty model may contradict the principle of anomaly detection, where the objective is to detect healthy behaviors from the rest. Nevertheless, the aim of an adaptive condition monitoring system should be to learn from all the identified conditions to subsequently detect them if they are presented again by a fault detection module. Indeed, some works present a parallel structure for fault detection and novelty detection modules [18], [19], where the novelty module learn the known faults along with the healthy operation because its objective is to detect only new scenarios, while the objective of the fault detection is to identify the condition of the machine, including known faults. This work is based on such parallel structure approach, where the fault scenarios must be taken into consideration to be included in the novelty model, and then they could be identified by a complementary fault detection module with a high confidence level

## 4. Experimental Results

### 4.1. Description of the experimental platform

The test bench used for testing the kinematic chain with different faults and the acquisition system used to capture the vibration signals is shown in Fig. 3. The test bench consists in a 1492-W three-phase induction motor (WEG 00236ET3E145T-W22), with its rotational speed controlled through a variable frequency drive-VFD (WEG CFW08), the operating speed is fixed to 60 Hz for all experiments. It also consist of a 4:1 ratio gearbox (BALDOR GCF4X01AA) and a DC generator (BALDOR CDP3604), which is used as the mechanical load comprising around 20% of the nominal load. The vibration signals are acquired using a tri-axial accelerometer (LIS3L02AS4), mounted on a board with the signal conditioning and anti-alias filtering. Two 12-bit 4-channel serial-output sampling analog-to-digital converter (ADS7841), are used on-board data acquisition system (DAS). Vibration signal from the perpendicular plane ( $y$ - $z$ ) to the motor shaft is acquired using a tri-axial accelerometer, LIS3L02AS4, mounted on a board with the signal conditioning and anti-aliasing filtering. Sampling frequency is set to 3 kHz for vibration acquisition. The data retrieved by the DAS is stored in a regular computer (PC).

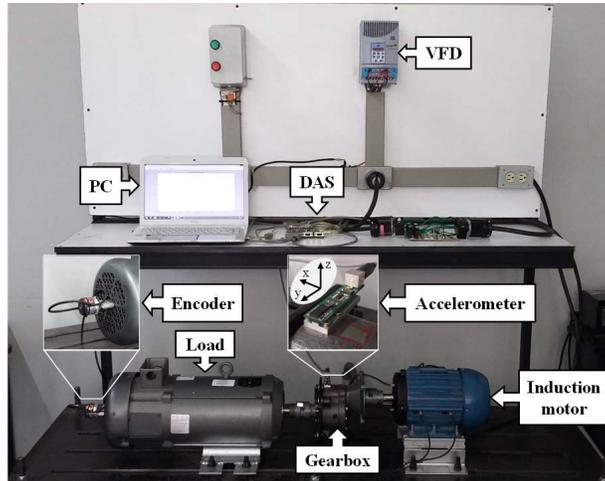
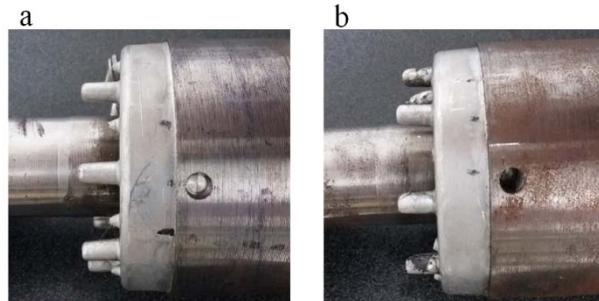


Fig. 3. Test bench used for experimentation.

Three scenarios are considered to verify the performance of the proposed method, the first one,  $H$ , is the kinematic chain working under healthy condition and the other two,  $F_1$  and  $F_2$ , represent the kinematic chain working under faulty conditions. For  $F_1$  the motor is working with a half broken bar, and for  $F_2$  the motor is working with a fully-broken bar. The detail of the failures is shown in Fig. 4. The half broken bar failure is artificially produced by drilling a 6 mm hole with a depth of 3 mm that corresponds mostly to the 22% of the section of the rotor bar, and the full broken bar is produced by a through-hole with a diameter of 6 mm and a depth of 14 mm, which corresponds to the complete section of the rotor bar.



**Fig. 4.** Detail of the faults produced in the test bench. (a) Corresponds to the  $\frac{1}{2}$  broken rotor bar, and (b) to one broken rotor bar

The information stored from the kinematic chain consist of an acquisition of 60 seconds of the machine working under the three scenarios mentioned; each acquisition is segmented in 30 parts of 2 seconds and a set of features is calculated from the 30 segmented measurements. Since two axes are taken into consideration, a total of ten features are calculated from each segmented acquisition of the machine working under the different scenarios mentioned. The first step of the methodology is the offline stage, where a reduced set of features is obtained and the novelty model is designed.

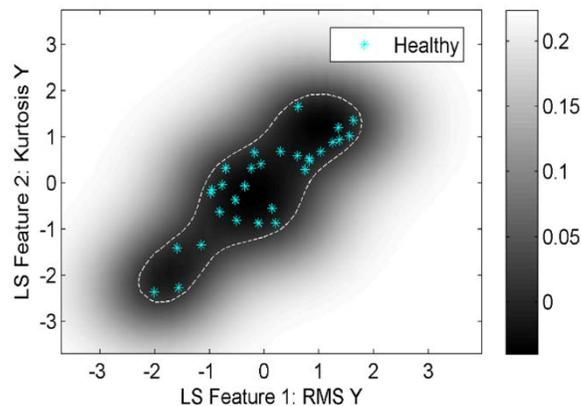
#### 4.2. Parameter Selection

Regarding the Laplacian Score configuration, a simple approach is followed for parameter tuning, that is, a value of  $k=3$  is used for constructing the adjacency graph and a “simple minded” weighting approach is followed. Since the proposed approach is be compared to different feature reduction modules, selecting generic parameters settings is useful for the purpose of evaluation, but ignores that there may be dependencies between the feature reduction model and the novelty model. Regarding the design of the novelty model, the kernel used is the *Gaussian* and the value of the configuration parameter is tuned to minimize the error in the validation. In all experiments 80% of the samples are used for training and 20% for validation. To train and adjust the parameter the novelty model a five-fold cross-validation is used.

#### 4.3 Results and discussions

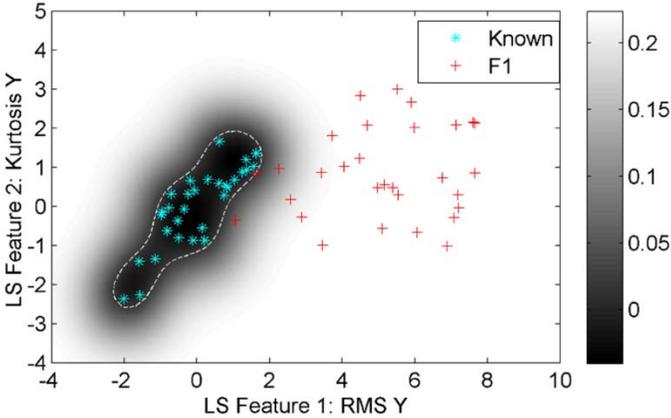
In order to highlight the contributions and motivation of this work, the outline of the results is presented as follows: first, a test is performed by a classical approach, then the proposed methodology is applied and the results are compared. The classical static approach imply conserving the reduced features set obtained at the offline initialization stage, meanwhile the proposed dynamic approach implies a possible reformulation of the reduced features set during the retraining stage.

During the initialization stage, the reduced set of features selected by means of the Laplacian Score ranking is composed by the RMS of the Y axis and the Kurtosis of the Y axis. The OC-SVM model is trained employing healthy condition data. The resulting OC-SVM during training is shown in Fig. 5. The marks, \*, in the figure, represent measurements of the machine used to train the model, on this case and on the subsequence figures only 1 fold of the five-fold cross validation is displayed. The dotted line represents the novelty threshold value, all data inside the boundaries of the dotted line is considered normal. The contour plot represents the novelty score evaluation over different regions of the feature map.



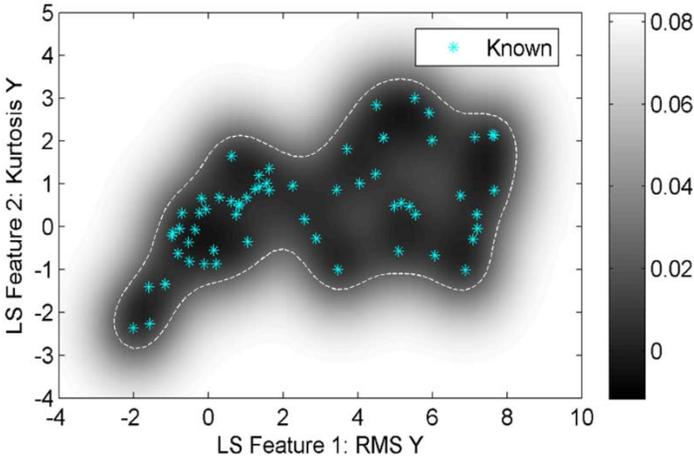
**Fig. 5.** Initial novelty model representation. Limit of the novelty threshold, --, and measurements used to train the model, \*.

Once the offline initialization stage is finished, the online stage follows, that is, new measurements are obtained to assess the condition of the monitored machine. To give robustness to the novelty detection and avoid false alarms, a batch consisting of 30 measurements are evaluated, if 75% of the analyzed measurements are evaluated as novelty then the alarm is triggered. Next, the  $F_1$  scenario is presented to test the performance of the model. The plot of the scenario and the novelty threshold obtained during training is presented in Fig. 6.



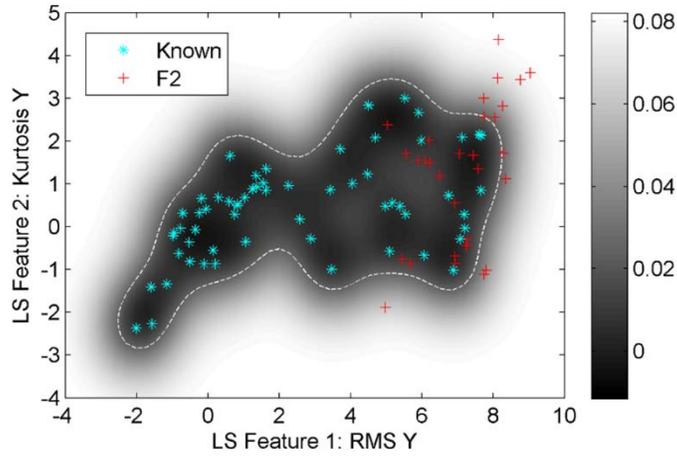
**Fig. 6.** Evaluation of the fault scenario  $F_1$ . The novelty model is trained employing data from healthy operation condition.

As can be appreciated, the new scenario lays outside the novelty threshold so it is successfully detected as novelty. Once a novel scenario is detected and identified as a fault by the user, it is incorporated to the database to consider it as part of the known scenarios and the novelty model is retrained, without changing the features, to include this information. Fig.7 shows the feature space after the novelty model is trained using Healthy and  $F_1$  data as part of the known scenarios.



**Fig. 7.** Contour plot of the novelty model after including  $F_1$ .

Once the model is re-trained, the third scenario,  $F_2$  is introduced. The visual representation of the test is presented in Fig. 8, this scenario is not detected as novel because only 50% of the measurements are labeled as novel by the model.



**Fig. 8.** Evaluation of the fault scenario  $F_2$ . The novelty model is trained employing data from healthy and  $F_1$  scenarios.

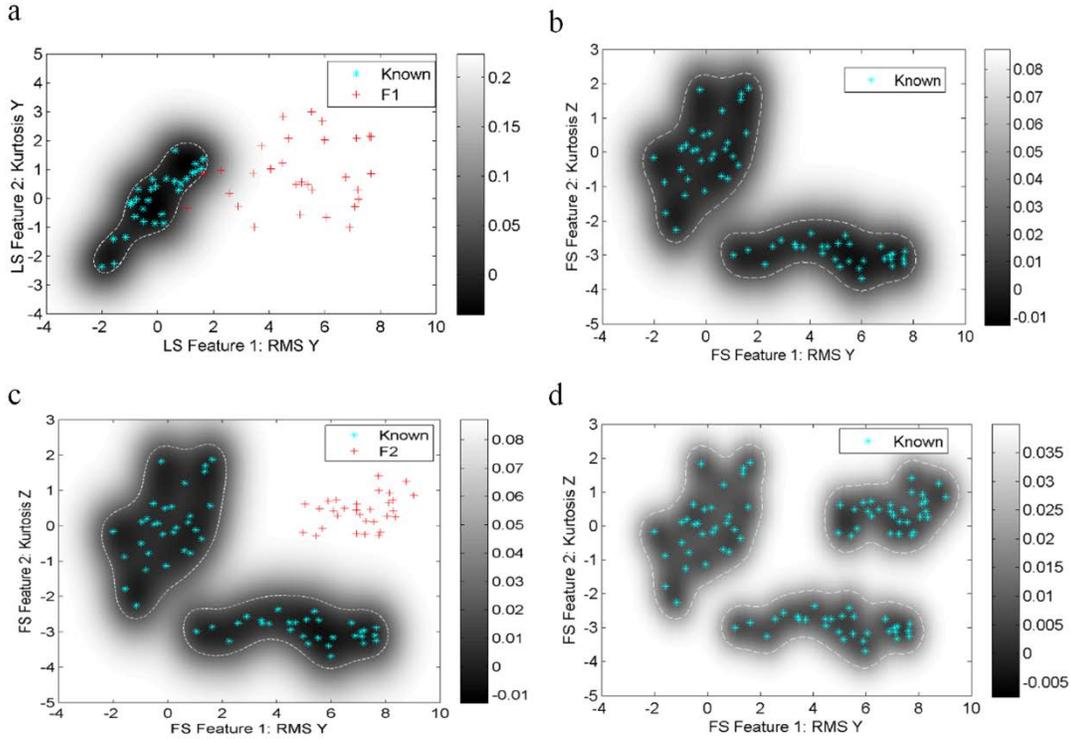
A similar result is obtained when the novelty model is trained using Healthy and  $F_2$  data and is tested with data of  $F_1$ . A summary of the results of novelty detection maintaining the same features obtained during initialization is shown in Table 2.

**Table 2.** Performance of the novelty detection using only healthy class data to reduce the number of features, where D.R. stands for dimensionality reduction. Different scenarios are included according the information available to train and test the novelty model.

D. R.	Performance Using OC-SVM (%)							
	Known		Test		Known		Test	
	H	F1 + F2	H + F1	F2	H + F2	F1		
LS	95.4 ( $\pm 1.1$ )		57.6 ( $\pm 6.7$ )		43.5 ( $\pm 6.4$ )			

As it can be seen in Table 2, using the reduced set obtained during initialization is easy to detect the novel scenarios when there is only information of the healthy condition, but when a novel scenario is included in the database and the novelty model is retrained, the reduced set of features initially obtained does not necessarily provide a good representation to detect a new scenario during test.

To improve these results, the methodology presented on this work proposes to evaluate again the feature reduction module each time a retraining is applied. Following the outline presented for results and parting from the first retrain where the scenario  $F_1$  is included in the database, the feature reduction module is applied again but this time including information from Healthy condition and  $F_1$  scenario. Since two scenarios are taken into consideration and the labels are known, a supervised approach can be applied, in this case a Fisher Score ranking for feature reduction is employed. The novelty model using the initial set of features and the novelty model after retraining with the new reduced set of features are shown in Fig. 9.



**Fig. 9.** Process of evaluation and retraining employing the methodology proposed. (a) Evaluation of the fault scenario  $F_1$ . (b) Retraining of the novelty model and reformulation of the reduced set of features including  $F_1$ . (c) Evaluation of the fault scenario  $F_2$ . (d) Retraining of the novelty model and reformulation of the reduced set of features including  $F_2$ .

The new reduced set of features is composed by the RMS of the  $Y$  axis and the Kurtosis of the  $Z$  axis, as can be appreciated, the new set of features present a more discriminative distribution of the three scenarios considered. At the last step, Fig. 9(d), when  $F_2$  scenario is included, the Fisher Score still ranked the same features of the last retraining, Fig. 9(b), as the highest; yet, it would be possible that a different set of features would be obtained. The results obtained demonstrates the advantages of including the Fisher Score reduction module to the retraining procedure, the new distribution of the scenarios avoided an overlapping of the new scenario under test, therefore the novelty model successfully detected as novelty all measurements corresponding to the  $F_2$  scenario.

It is worth mentioning that if initially the scenario  $F_2$  is used for training and the scenario  $F_1$  is used for testing, the new set of features obtained could be different than the aforementioned. The results achieved from both scenarios are shown in Table 3, which also includes a comparative of the results obtained employing PCA, LDA and Laplacian Score dimensionality reduction techniques on the retraining step instead of the Fisher Score proposed.

**Table 3.** Performance of the novelty detection employing a reduction of features during retraining. Different scenarios are included according the information available to train and test the novelty model.

D. R.	Performance Using OC-SVM (%)			
	Known		Test	
	H + F1	F2	H + F2	F1
PCA	59.0 (± 5.2)		64.2 (± 5.1)	
LDA	47.4 (± 6.1)		51.2 (± 6.1)	
LS	84.3 (± 2.4)		100 (± 0.0)	
Fisher S.	100 (± 0.0)		100 (± 0.0)	

Regarding the classical feature extraction techniques, PCA and LDA, the test scenario is not identified as novel in both cases. The Fisher Score and the Laplacian Score successfully identified the new scenarios as novel since the percentage obtained in both cases is higher than the 75% predefined threshold in the methodology to activate the alarm. Both techniques achieved high scores, but still the Fisher Score provided a more appropriate selection of features.

As mentioned in Section II, LDA is a feature extraction technique based on the Fisher discriminant coefficient, so similar results between Fisher Score ranking and LDA are expected; however, employing Fisher Score ranking score achieved a better result; this can be caused because the test consist of novel scenarios and LDA finds the directions on the feature space specialized for the two supervised scenarios employed during training, meanwhile Fisher Score ranking provides a more general approach by selecting features.

To test the robustness of the Laplacian Score and Fisher Score approaches, a comparative test is performed where the set of features is increased from 10 to 15 and varying the number of reduced features selected. The five features included to the original set are obtained from the axis X of the accelerometer monitored, where the features calculated are presented on Table 1; these features were discarded initially because they are not part of the perpendicular plane of the motor and does not contribute significantly to the monitoring, in fact it could affect the performance of the feature reduction and novelty detection modules. The results are presented in Table 4.

**Table 4.** Performance of the novelty detection increasing the number of initial features from 10 to 15 and varying the number of the reduced set of features.

D.R.	Performance varying the number of features reduced (%)		
	2	3	4
LS	59.0 ( $\pm 4.1$ )	64.3 ( $\pm 3.3$ )	88.6 ( $\pm 2.5$ )
Fisher S.	100 ( $\pm 0.0$ )	100 ( $\pm 0.0$ )	100 ( $\pm 0.0$ )

The features obtained from the Fisher Score still present a better distribution to detect the new scenarios. The Laplacian Score performance is affected when irrelevant features are included on the feature set but increases when more dimensions are taken into consideration. Since the objective of the Fisher Score is not topology preservation, contrary to LS, it is capable of discarding all the irrelevant features that were included and the performance is not affected.

## 5. Conclusions

This work proposes an adaptive novelty detection methodology based on vibration analysis for the condition monitoring and diagnosis of the components of a kinematic chain. The methodology is based on the acquisition of vibration signals that are generated in the kinematic chain; along with an adequate signal processing to extract features to characterize the components, and an adaptive novelty detection model to detect anomalies. The method is composed by two sequential stages, an offline stage to initialize and retrain the modules, and an online stage to continuously assess the condition of the machine. During initialization the model is trained employing only information from the machine working under healthy condition and two additional faulty scenarios are introduced to test the performance of the method under unknown operations.

The adaptive novelty detection approach successfully detected both novel scenarios and the model incorporated the information to avoid generating alarms if the same fault is detected. A comparison between the proposed method and classical dimensionality reduction approaches highlights the limitations of maintaining a static set of features during monitoring, instead of reformulating the feature reduction module once new information is available. On this particular study, employing features reduced by Laplacian Score and Fisher Score obtained similar results; nevertheless, it does not imply that a similar outcome will be present during the analysis of other faults. Fisher Score is encouraged to be employed on this methodology rather than maintaining the

Laplacian Score approach due to the similarity between the method objective and the objective function of Fisher Score, on both the ideal case is to find the features to maximize the distance between scenarios while maintaining compact clusters. A specific comparative of performance between LS and Fisher Score was also included, in which the Fisher Score obtained better results when irrelevant features are included to the original set of features and when the dimensionality of the reduced set is increased; this highlights the advantages and robustness of the feature selection approach by Fisher Score compared to the LS.

From an industrial perspective, two aspects should be mentioned. First, dealing with an identical operating scenario (i.e. healthy), the variation of the working conditions (speed and torque patterns), will affect the resulting value of the estimated numerical feature/s and, consequently, its representation in the feature space used for novelty model design. Thus, the introduction of the working condition as an additional degree of freedom in the novelty detection scheme implies the increase of the dimensional space that the novelty detection model is managing. This fact implies an increase of the risk of data overlapping (different scenarios under different operating conditions showing similar feature values), and also, an increase of complexity for boundaries definition during the novelty modelling (data spread). Thus, the proposed method is not limited to stationary conditions, but to the repetition of the pattern of working condition used to train the model. Indeed, in practical industrial application, where cyclostationary processes take place, the proposed method should be executed during the same period (same working conditions), in order to reduce data variability and increase novelty performance. In case of multiple working conditions consideration, parallel novelty model structures could be generated in order to activate each one depending on the corresponding working condition.

Second, the integration of additional physical magnitudes as stator currents, temperatures and others, represents a trendy approach in electromechanical fault diagnosis in order to improve fault diagnosis resolution and performance. The corresponding data fusion scheme implies a high-dimensional feature space, that is, multiple numerical features estimated from multiple physical magnitudes considered. In this point, it must be noted that the presented methodology proposes the re-evaluation of the feature reduction module each time a retraining is applied. Thus, the larger the available pool of features is, the higher is the potentiality of the proposed method to increase the performance during the data characterization.

Also, the proposed method can be extended and improved for further development, this improvement could include a diagnosis method to not only detect anomalies on the kinematic chain, but to identify the fault causing the abnormal behavior. In these sense, future work will include a specific comparison of broken bar fault detection capabilities between the proposed methodology and classical approaches.

## Acknowledgements

This research was partially supported by CONACyT scholarships 278033 and 313604 grant and also by the Spanish Ministry of Education, Culture, and Sport under the grant FPU13/00589. The authors would also like to thank a partial financial support by the Generalitat de Catalunya (GRC MCIA, Grant n° SGR 2014-101).

## References

- [1] H. Henaou, G.-A. Capolino, M. Fernandez-Cabanas, F. Filippetti, C. Bruzzese, E. Strangas, R. Pusca, J. Estima, M. Riera-Guasp, and S. Hedayati-Kia, "Trends in Fault Diagnosis for Electrical Machines: A Review of Diagnostic Techniques," *IEEE Industrial Electronics Magazine*, vol. 8, no. 2, pp. 31–42, 2014.
- [2] N. Bayar, S. Darmoul, S. Hajri-Gabouj, and H. Pierreal, "Fault detection, diagnosis and recovery using Artificial Immune Systems: A review," *Eng. Appl. Artif. Intell.*, vol. 46, pp. 43–57, Nov. 2015.
- [3] S. Wang, J. Wan, D. Zhang, D. Li, and C. Zhang, "Towards smart factory for Industry 4.0: A self-organized multi-agent system with big data based feedback and coordination," *Comput. Networks*, Jan. 2016.

- [4] M. a. F. Pimentel, D. a. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," *Signal Processing*, vol. 99, pp. 215–249, Jun. 2014.
- [5] R. Langone, O. Mauricio Agudelo, B. De Moor, and J. A. K. Suykens, "Incremental kernel spectral clustering for online learning of non-stationary data," *Neurocomputing*, vol. 139, pp. 246–260, Sep. 2014.
- [6] M.-H. Chen, P.-C. Chang, and J.-L. Wu, "A population-based incremental learning approach with artificial immune system for network intrusion detection," *Eng. Appl. Artif. Intell.*, Jan. 2016.
- [7] S. Khomfoi and L. M. Tolbert, "Fault diagnosis and reconfiguration for multilevel inverter drive using AI-based techniques," *IEEE Transactions on Industrial Electronics*, vol. 54, no. 6, pp. 2954–2968, 2007.
- [8] A. Bellini, F. Filippetti, C. Tassoni, and G. A. Capolino, "Advances in diagnostic techniques for induction machines," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 12, pp. 4109–4126, 2008.
- [9] V. T. Tran and B. S. Yang, "An intelligent condition-based maintenance platform for rotating machinery," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 2977–2988, Feb. 2012.
- [10] P. A. Delgado-Arredondo, A. Garcia-Perez, D. Morinigo-Sotelo, R. A. Osornio-Rios, J. G. Avina-Cervantes, H. Rostro-Gonzalez, and R. de J. Romero-Troncoso, "Comparative Study of Time-Frequency Decomposition Techniques for Fault Detection in Induction Motors Using Vibration Analysis during Startup Transient," *Shock Vib.*, vol. 2015, pp. 1–14, 2015.
- [11] M. D. Prieto, G. Cirrincione, A. G. Espinosa, J. A. Ortega, and H. Henao, "Bearing Fault Detection by a Novel Condition-Monitoring Scheme Based on Statistical-Time Features and Neural Networks," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 8, pp. 3398–3407, 2013.
- [12] P. Perner, "Concepts for novelty detection and handling based on a case-based reasoning process scheme," *Eng. Appl. Artif. Intell.*, vol. 22, no. 1, pp. 86–91, Feb. 2009.
- [13] C. Englund and A. Verikas, "A hybrid approach to outlier detection in the offset lithographic printing process," *Eng. Appl. Artif. Intell.*, vol. 18, no. 6, pp. 759–768, Sep. 2005.
- [14] H. Garces and D. Sbarbaro, "Outliers detection in environmental monitoring databases," *Eng. Appl. Artif. Intell.*, vol. 24, no. 2, pp. 341–349, Mar. 2011.
- [15] S. Ma, S. M. Li, and Y. P. Xiong, "Uncertainty reduced novelty detection approach applied to rotating machinery for condition monitoring," *Shock Vib.*, vol. 2015, p. 10, 2015.
- [16] J. McBain and M. Timusk, "Feature extraction for novelty detection as applied to fault detection in machinery," *Pattern Recognit. Lett.*, vol. 32, no. 7, pp. 1054–1061, May 2011.
- [17] S. Wang, J. Yub, E. Lapirac, and J. Lee, "A modified support vector data description based novelty detection approach for machinery components," *Appl. Soft Comput. J.*, vol. 13, no. 2, pp. 1193–1205, Feb. 2013.
- [18] D. P. Filev, R. B. Chinnam, F. Tseng, and P. Baruah, "An industrial strength novelty detection framework for autonomous equipment monitoring and diagnostics," *IEEE Transactions on Industrial Informatics*, vol. 6, no. 4, pp. 767–779, 2010.
- [19] B. S. J. Costa, P. P. Angelov, and L. A. Guedes, "Fully unsupervised fault detection and identification based on recursive density estimation and self-evolving cloud-based classifier," *Neurocomputing*, vol. 150, Part , no. Part A, pp. 289–303, Feb. 2015.
- [20] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. C. Platt, "Support Vector Method for Novelty Detection.," in *Nips*, 1999, vol. 12, pp. 582–588.
- [21] D. M. J. Tax and R. P. W. Duin, "Support vector domain description," *Pattern Recognit. Lett.*, vol. 20, no. 11–13, pp. 1191–1199, Nov. 1999.
- [22] S. D. Villalba and P. Cunningham, "An evaluation of dimension reduction techniques for one-class classification," *Artif. Intell. Rev.*, vol. 27, no. 4 SPEC. ISS., pp. 273–294, 2007.
- [23] I. Jolliffe, *Principal Component Analysis*, 2nd ed. Springer-Verlag New York, 1986.
- [24] X. He, D. Cai, and P. Niyogi, "Laplacian Score for Feature Selection," in *Advances in Neural Information Processing Systems 18*, 2005, pp. 507–514.
- [25] A. J. Izenman, "Linear Discriminant Analysis," A. J. Izenman, Ed. New York, NY: Springer New York, 2008, pp. 237–280.

- [26] Q. Gu, Z. Li, and J. Han, "Generalized Fisher Score for Feature Selection," *CoRR*, vol. abs/1202.3, no. August, pp. 327 – 330, 2012.
- [27] Z. Gao, C. Cecati, and S. X. Ding, "A Survey of Fault Diagnosis and Fault-Tolerant Techniques&#x2014;Part II: Fault Diagnosis With Knowledge-Based and Hybrid/Active Approaches," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 6. pp. 3768–3774, 2015.
- [28] A. E. Lazzaretti, D. M. J. Tax, H. Vieira Neto, and V. H. Ferreira, "Novelty detection and multi-class classification in power distribution voltage waveforms," *Expert Syst. Appl.*, vol. 45, pp. 322–330, Mar. 2016.