Chapter 5

Evaluation of Novelty Detection Methods for Condition Monitoring applied to an Electromechanical System

Miguel Delgado Prieto, Jesús A. Cariño Corrales,
Daniel Zurita Millán, Marta Millán Gonzalvez,
Juan A. Ortega Redondo and
René de J. Romero Troncoso

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/67531

Abstract

Dealing with industrial applications, the implementation of condition monitoring schemes must overcome a critical limitation, that is, the lack of a priori information about fault patterns of the system under analysis. Indeed, classical diagnosis schemes, in general, outdo the membership probability of a measure in regard to predefined operating scenarios. However, dealing with noncharacterized systems, the knowledge about faulty operating scenarios is limited and, consequently, the diagnosis performance is insufficient. In this context, the novelty detection framework plays an essential role for monitoring systems in which the information about different operating scenarios is initially unavailable or restricted. The novelty detection approach begins with the assumption that only data corresponding to the healthy operation of the system under analysis is available. Thus, the challenge is to detect and learn additional scenarios during the operation of the system in order to complement the information obtained by the diagnosis scheme. This work has two main objectives: first, the presentation of novelty detection as the current trend toward the new paradigm of industrial condition monitoring and, second, the introduction to its applicability by means of analyses of different novelty detection strategies over a real industrial system based on rotatory machinery.

Keywords: condition monitoring, electromechanical systems, failure diagnosis, feature reduction, industrial monitoring applications, novelty detection, open set recognition

1. Introduction

Currently, condition monitoring plays a key role in the reliability and safety strategies of most of the industrial applications [1]. Classical industrial condition monitoring methodologies imply the estimation of numerical features and their posterior processing in order to characterize the available physical magnitudes acquired during the operation of the system under analysis. Such numerical feature vectors are, then, presented to a classification algorithm in order to obtain a diagnosis outcome [2]. In this procedure, the algorithm of classification is previously trained with available data representative of different system conditions. Thus, during the regular operation of the condition monitoring scheme, each measurement acquired from the system will be transformed to a vector of numerical features, and its similarities with previous patterns will be evaluated in order to obtain the related probability. During the last decades, a great deal of studies has been done around different aspects of the electromechanical condition monitoring, that is, the potentiality of different physical magnitudes for fault detection, the analysis of time, frequency and time-frequency domains for numerical features' estimation, the effect of feature reduction techniques for patterns' characterization and dealing with data-based approaches and multiple classification strategies for diagnosis improvement [3, 4]. All of these works are, with no doubt, a major step forward to the study, research and development of enhanced condition monitoring schemes to be applied to electromechanical systems. However, currently, the scientific and industrial communities are working together toward more demanding industrial challenges in the frameworks of Industry 4.0 [5] and Zero-defect manufacturing [6]. Indeed, further capabilities are expected from the condition monitoring developments in order to face questions about their practical implementations, questions such as: *How must condition monitoring be managed in front of new operating scenarios not previously considered?*, *How to detect new operating scenarios?*, *Which numerical features should be used for unknown patterns' detection?*, *How to preserve the diagnosis reliability in the presence of new patterns?*, *Is it possible to automate these considerations or is the aid of an expert required?* In order to find answers to such questions, specific research is being gathered around the so-called novelty detection topic, which can be defined as the task of recognizing that the data under analysis differ, in some respect, from the initial available data.

Indeed, a priori characteristic fault patterns of specific rotatory machinery are not usually available and highly difficult to estimate through theoretical approaches. Thus, condition monitoring strategies capable of detecting novel operating conditions, alongside the classification of known conditions, represent the most convenient solutions [7]. This approach is known as the open set recognition problem, where only a reduced set of known operating scenarios are included in the initial dataset and used during the training stage, and, then, novel (unknown) scenarios may appear during the online diagnosis stage. In general terms, in order to deploy a novel detection strategy, a model must be trained with all the available data describing the initial-known scenarios of the machinery under monitoring. Thus, the model generates a threshold system that allows to discriminate between known scenarios and measurements corresponding to new cases, novelties. Different approaches differ in the way that the threshold system is generated [8].

http://dx.doi.org/10.5772/67531

In this chapter, first, a more comprehensive description of the novelty detection topic, including different approaches and their dependencies, is introduced. Later, the practical application of novelty detection applied over an industrial electromechanical system is described. The performances obtained with different novelty detection strategies, including the effect of feature reduction, are discussed finally.

2. Novelty detection

The introduction of novelty detection into the classical monitoring chain represents a previous condition to the diagnosis assessment. The classical step flow to implement novelty detection is shown in **Figure 1**. The procedure begins with the off-line processing of the available information (generally, the healthy behavior of the electromechanical chain). Such processing, in regard to the raw data acquired (stator currents, temperatures, etc.), consists of the definition of the same blocks as in classical diagnosis procedures, that is, feature estimation (calculation of a set of numerical features) and feature reduction (feature vector transformation for improved characterization). Once the available data is characterized by vectors of D features, the configuration of the novelty detection model follows. This part depends entirely on the nature of the novelty model (different approaches can be applied); however, the objective is the

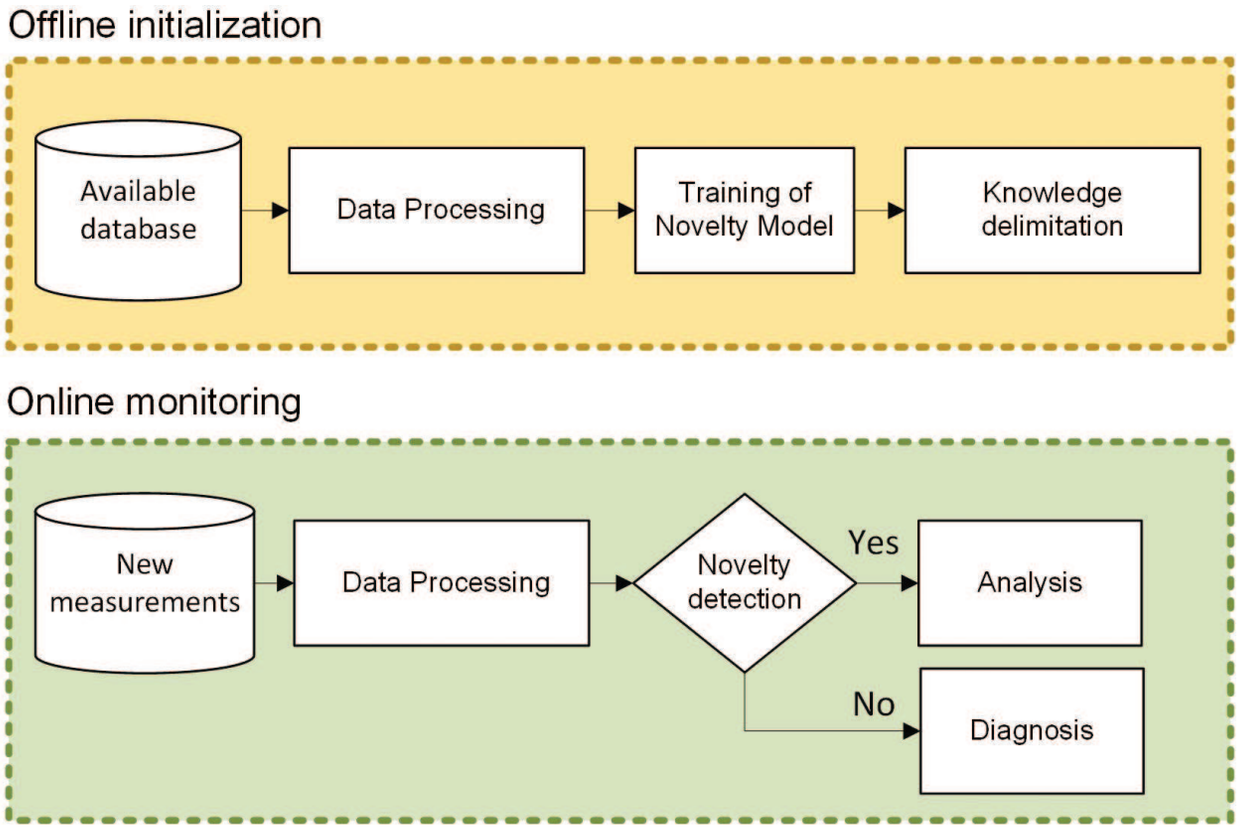

Figure 1. Implementation scheme of novelty detection, including the offline initialization for knowledge delimitation and the online monitoring for novelty evaluation.

delimitation of the available knowledge defining a set of mathematical descriptions in the *D*-dimensional feature space in which the available database is projected [9].

Thus, during the online monitoring, the novelty detection model will analyze the new acquisitions and will determine if the new data correspond to known operating scenarios previously learned or present different characteristics and can be considered novel representations. In case of known operating scenarios, the diagnosis follows. It must be noted that the diagnosis procedure could include different feature estimation and feature reduction stages because its objective is completely different. In this sense, the novelty detection does not require different labels since all data belongs to the class *knowledge* or *normal*. The diagnosis, however, requires to maintain different labels in order to allow the identification of the different operating scenarios. In case of unknown operating scenarios, the diagnosis cannot be carried out since the diagnosis reliability would be affected. In this case, the presence of novel data is reported and, after the supervision of an expert, the measurements are stored in order to upgrade the known operating scenarios, which will imply the retraining of the novelty model [10].

In order to illustrate the novelty detection operation, an example is shown in **Figure 2**. A *D* = 2-dimensional feature space is considered, in which a set of measurements representing the available data has led to the definition of the boundaries corresponding to the known conditions. When new measurements are acquired, the novelty model analyzes them and determines, in this example by means of their position in the 2-dimensional feature space, if they represent novelty or if the behavior is still considered known. If a significant amount of novel acquisitions with the same characteristics are detected, then, a novel operation mode is detected and, if validated, the data will be included as known behaviors.

Indeed, the detection of novel events is an important ability of any condition monitoring scheme. Considering the fact that it cannot be trained within a machine learning system with all possible systems' variability, it becomes important to include the differentiation capability between *known* and *unknown* object information during the system's monitoring. However, it has been considered in practice by several studies that the novelty detection is an extremely

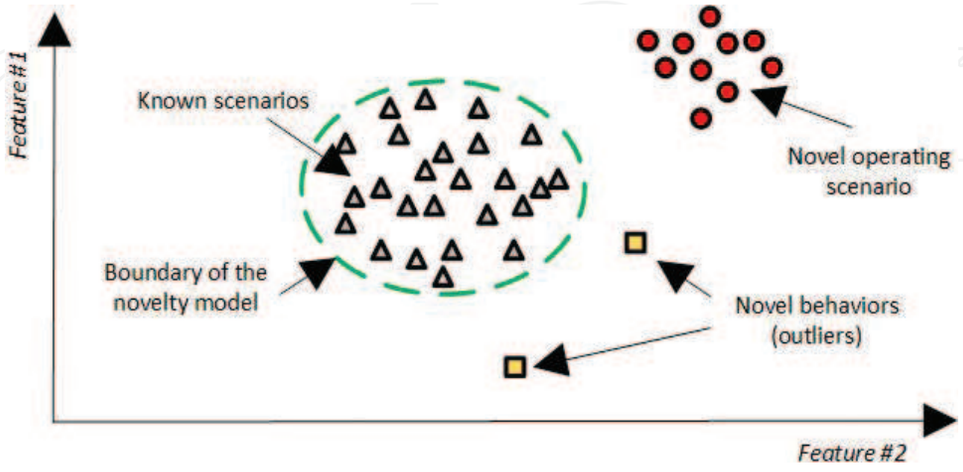

Figure 2. Example of a novelty detection basic operation, including available data, Δ, mathematical description of the available data, --, novel behaviors detected as outliers, □ and novel behaviors detected as the novel operating scenario, O.

http://dx.doi.org/10.5772/67531

challenging task. It is for this reason that there exist different approaches of novelty detection that have been demonstrated to perform well under different applications [11, 12]. Unfortunately, it is clearly evident that there is no single best model for novelty detection, and the success depends not only on the type of the method used but also on the statistical properties of the available data. Next, the three basic novelty detection approaches are described, including probabilistic, domain-based and distance-based methods.

2.1. Probabilistic methods

Probabilistic approaches to novelty detection are based on estimating the probability density function (PDF) of the available data. The resulting distribution may then be thresholded to define the boundaries of *normality* in the feature space and assess whether a new measurement belongs to the same distribution or not. The training data is assumed to be generated from some underlying probability distribution. This estimation usually represents the novelty model, and a novelty threshold can be set over such estimation. The estimation of the underlying data density from a multivariate training dataset is a well-established topic [10].

Probabilistic methods are divided in parametric and nonparametric approaches. Parametric approaches impose a restrictive model on the data, which results in a large bias when the model does not fit the data. Nonparametric approaches set up a very flexible model by making fewer assumptions over the data: The model grows in size to accommodate the complexity of the data, but it requires a large sample size for a reliable fit out of all the free parameters. The opinion in the scientific literature is divided as to whether various techniques should be classified as parametric or nonparametric. For the purposes of providing probabilistic estimators, Gaussian Mixture Model (GMM) and Kernel Density Estimator (KDE) have proven popular. The GMM is typically classified as a parametric technique [11] because of the assumption that the data is generated from a weighted mixture of Gaussian distributions. The KDE is typically classified as a nonparametric technique [13] as it is closely related to histogram methods, one of the earliest forms of nonparametric density estimation approaches.

2.1.1. Gaussian-mixture model

The GMM is a parametric probability density function represented as a weighted sum of Gaussian component densities. The GMM parameters are estimated from the available training data using, for example, the iterative expectation-maximization algorithm or the maximum a posteriori estimation. Thus, a GMM is a weighted sum of M component Gaussian densities, mathematically described as,

$$p(x|\lambda) = \sum_{i=1}^{M} w_i g\left(x|\mu_i, \sum i\right) \tag{1}$$

where x is a D-dimensional vector, w_i, i=1,..M are the mixture weights and $g(x|u_i, \sum i)$, i=1,..,M are the component Gaussian densities. Each component density is a D-variate Gaussian function of the form,

$$g(x|\mu_i, \sum i) = \frac{1}{(2\pi)^{D/2}|\sum i|^{1/2}} exp\left\{-\frac{1}{2}(x-\mu_i)' \sum_i^{-1}(x-\mu_i)\right\} \quad (2)$$

with mean vector μ_i and covariance matrix Σ_i. The mixture weights satisfy the constraint that $\sum_{i=1}^{M} w_i = 1$. The complete Gaussian mixture model is parameterized by the mean vectors, covariance matrices and mixture weights from all component densities. These parameters are collectively represented by the notation $\lambda = \{w_i, \mu_i, \Sigma_i\},\ i = 1, .., M$. There are several variants on the GMM. The covariance matrices, Σ_i, can be in full rank or constrained to be diagonal. Additionally, parameters can be shared, or tied, among the Gaussian components, such as having a common covariance matrix for all components. The choice of model configuration is often determined by the amount of data available for estimating the GMM parameters and how the GMM is used in a particular application [14]. In fact, GMM can suffer from the requirement of large numbers of training examples to estimate model parameters. A further limitation of parametric techniques is that the chosen functional form for the data distribution may not be a good model of the distribution that generates the data.

One of the major issues in novelty detection is the selection of a suitable novelty threshold. Within a probabilistic approach, novelty scores can be defined using the unconditional probability distribution $z(x) = p(x)$ and a typical approach to setting a novelty threshold k is to threshold this value, that is, $p(x) = k$. However, because $p(x)$ is a probability density function, a threshold on $p(x)$ has no direct probabilistic interpretation. Some studies have interpreted the model output $p(x)$ probabilistically, by considering the cumulative probability P associated with $p(x)$, that is, determining the probability mass obtained by numerically estimating the integral of $p(x)$ over a region R for which the value of $p(x)$ is above the novelty threshold k [15]. For unimodal distributions, one can integrate from the mode of the probability density function to the probability contour defined by the novelty threshold $p(x) = k$, which can be achieved in a closed form for most regular distributions.

2.1.2. Kernel density estimator

Nonparametric approaches do not assume that the structure of a model is fixed, that is, the model grows in size necessary to fit the data and accommodates the complexity of the data. The simplest nonparametric statistical technique is the use of histograms. The algorithm typically defines a distance measure between a new test data point and the histogram-based model of normality to determine if it is an outlier or not [10]. For multivariate data, attribute-wise histograms are constructed and an overall novelty score for a test data point is obtained by aggregating the novelty scores from each attribute. However, when a histogram is defined, it is necessary to consider the width of the bins (equal subintervals in which the whole data interval is divided) and the end points of the bins (where each of the bins starts). In consequence, the histograms present a nonsmooth behavior. In order to alleviate this deficiency, the kernel estimators were proposed.

It must be considered that observations are being drawn from some unknown probability density function $p(x)$ in a Euclidian D-dimensional feature space. Thus, considering a region R containing the D-dimensional measurement x, the probability mass associated with this

http://dx.doi.org/10.5772/67531

region is given by $P = \int_R p(x)dx$. Taking into account a dataset comprising N observations drawn from $p(x)$, each data point has a probability P of falling within R, and the total number K of points that lie inside R will be distributed according to the binomial distribution $Bin(K|N,P) = \frac{N!}{K!(N-K)!}P^K(1-P)^{1-k}$. The mean fraction of points falling inside the region is $E[K|N] = P$, and the variance around this mean is $var[K|N] = P(1-P)/N$. For large N, this distribution will sharply peak around the mean and so $K \cong NP$. If, however, it is assumed that the region R is sufficiently small that the probability density $p(x)$ is roughly constant over the region, then $P \cong p(x)V$, where V is the volume of R. Thus, density estimate is obtained in the form,

$$P(x) = \frac{K}{NV} \tag{3}$$

Note that the validity of Eq. (3) depends on two contradictory assumptions, namely that the region R is sufficiently small that the density is approximately constant over the region and yet sufficiently large (in relation to the value of that density) that the number K of points falling inside the region is sufficient for the binomial distribution to sharply peak. The resultant Eq. (3) can be exploited in two different ways. Either it can be fixed K and the value of V can be determined from the data, which gives rise to the K-nearest-neighbor technique that will be presented later, or it can be fixed V and K can be determined from the data, giving rise to the kernel approach. It can be shown that both the K-nearest-neighbor density estimator and the kernel density estimator converge to the true probability density in the limit $N \to \infty$, provided V shrinks suitably with N, and K grows with N [13]. Thus, considering the region R as a small hypercube centered on the point x at which is desired to determine the probability density, the number K of points falling within region is defined as follows,

$$K(u) = \begin{cases} 1, & |u_i| \le \frac{1}{2},\ i = 1,..,D \\ 0, & otherwise \end{cases} \tag{4}$$

which represents a unit cube centered on the origin. The function $K(u)$ is an example of a kernel function and in this context is also called a Parzen window. From Eq. (4), the quantity $K\left(\frac{x-x_i}{h}\right)$ will be one if the data point x_i lies inside a cube of side h centered on x and zero otherwise. The total number of data points lying inside this cube will therefore be $K = \sum_{i=1}^{N} k\left(\frac{x-x_i}{h}\right)$. Substituting this expression in Eq. (3) gives the following result for the estimated density at x,

$$p(x) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{h^D}k\left(\frac{x-x_i}{h}\right) \tag{5}$$

where $V = h^D$ for the volume of a hypercube of side h in D dimensions. Eq. (5) represents the kernel density estimator [16]. Even though Gaussian kernels are the most often used, there are various choices among kernels that can be found in the literature [17].

2.2. Domain-based method

Domain-based method requires a boundary to be created based on the structure of the training dataset. These methods are typically insensitive to the specific sampling and the density of the target class because they describe the target class boundary, or the domain, and not the class density. Class membership of unknown data is then determined by their location with respect to the boundary. Domain-based novelty detection is approached with the two-class problem in terms of Support Vector Machine (SVM), where the location of the novelty boundary is determined using only those data that lie closest to it (in a kernel-based transformed space), by means of the support vectors. All other data from the training set (those that are not support vectors) are not considered when setting the novelty boundary. Hence, the distribution of data in the training set is not considered, which is seen as an easy novelty detection approach [7]. The original SVM is a network that is ideally suited for binary pattern classification of data that are linearly separable. Indeed, the SVM defines a hyperplane that maximizes the separating margin between two classes. Since the introduction of the original idea, several modifications and improvements have been made.

2.2.1. Support vector data description

A data domain description method, inspired by the support vector machine approach, called the Support Vector Data Description (SVDD), is used for novelty or outlier detection. The objective is the definition of a spherically shaped decision boundary around a set of measurements by a set of support vectors describing the hypersphere boundary. The method allows the possibility of transforming the data to new feature spaces, where the SVDD can obtain more flexible and more accurate data descriptions. The minimizing problem to delimitate the radius of the hypersphere is expressed as the Lagrangian, $L = \sum_i \alpha_i (x_i \cdot x_i) - \sum_{i,j} \alpha_i \alpha_j (x_i \cdot x_j)$, under the constraints of $0 \leq \alpha_i \leq C$ and $\sum_i \alpha_i = 1$, where $\alpha_{i,j}$ are the Lagrange multipliers, $x_{i,j}$ are the data training points, the variable C gives the trade-off between simplicity (or volume of the sphere) and the number of errors (number of target objects rejected). For those objects the coefficients $\alpha_{i,j}$ will be nonzero and are called the support objects. In order to determine whether a new measurement is within the hypersphere, the distance to the center of the sphere has to be calculated. A new measurement z is considered *known* when this distance is smaller than the radius,

$$(z \cdot z) - 2 \sum_i \alpha_i (z \cdot x_i) + \sum_{i,j} \alpha_i \alpha_j (x_i \cdot x_j) \leq r^2 \tag{6}$$

where a is the center of the sphere and r is the radius [18]. Kernels could be applied to soften the margins of the sphere, being applied over the measures and data descriptors.

2.2.2. One-class support vector machine

The One-Class SVM, OC-SVM, is based on the definition of the novelty boundary in the feature space corresponding to a kernel, by separating the transformed training data from the origin in the feature space, with the maximum margin. This approach requires fixing a priori

http://dx.doi.org/10.5772/67531

the percentage of positive data allowed to fall outside the description of the *normal* class. This makes the OC-SVM more tolerant to outliers in the *normal* training data. However, setting this parameter strongly influences the performance of this approach. The shape of the domain delimiting the boundaries depends on the kernel selected. Thus, the development of the algorithm is the classic SVM approach. The difference with the other domain-based method approach is that OC-SVM does not consider a specific structure (e.g., a hypersphere) to delimit the domain and therefore does not automatically optimize the model parameters by using artificially generated unlabeled data which are uniformly distributed. The detection of novelty is therefore delimited by,

$$f(x) = \sum_{i=1}^{N} \alpha_i K(x_i, x) - p \tag{7}$$

where p is an offset. The famous kernel trick is the procedure of using a kernel function in input space, $K(x_i, x)$, to replace the inner product of two vectors into a huge, or even infinite, dimensional feature space. Some drawbacks of these methods are found in literature reviews [7], and it turns out to be surprisingly sensitive to specific choices of representations and kernels in ways which are not very transparent. In addition, the proper choice of a kernel is dependent on the number of features in the binary vector. Since the difference in performance is very dramatic based on these choices, this means that the method is not robust without a deeper understanding of these representation issues.

2.3. Distance-based method

Distance-based methods represent a novelty detection approach similar to that of estimating the PDF of data. Distance-based methods such as nearest neighbors or clustering are based on well-defined distance metrics to compute distance, as the similarity criterion, among data points.

2.3.1. Nearest neighbor

The main idea that rears this technique is that the *normal* data is projected near their neighborhoods, while novelties will be projected far from their neighbors. That is, considering an unknown data point x, this point is accepted as normal if the distance to its nearest neighbor y, in the training set, is less than or equal to the distance from y to the nearest neighbor of y in the training set. Otherwise, x is considered as novelty. Euclidian distance is the most popular choice for univariate and multivariate continuous attributes,

$$\|x - y\| = \sqrt{\sum_{i=1}^{D} (x_i - y_i)^2} \tag{8}$$

Several well-defined distance metrics to compute the distance (or the similarity measure) between two data points can be used, which can broadly be divided into distance-based methods, such as the distance to the kth nearest neighbor and local density-based methods in which the distance to the average of the k's nearest neighbors is considered [11].

In conclusion, novelty detection approaches differ on the assumptions made about the nature of the available data. Each approach exhibits its own advantages and disadvantages and faces different challenges for complex datasets. **Table 1** collects the main characteristics of the considered methods. Thus, probabilistic methods make use of the distribution of the training data to determine the location of the novelty boundary. Domain-based methods determine the location of the novelty boundary using only those data that lie closest to it and do not make any assumptions about the data distribution. Distance-based methods require the definition of an appropriate distance measure for the given data.

Method	Advantages	Disadvantages
Domain-based *i.e. One-class SVM*	Robust to labeled outliers in training by forcing them to lie outside the description. Robust to unlabeled outliers in training.	Several configuration parameters. Sensitive to the scaling of the feature values. Requires a minimum number of training.
Probabilistic parametric *i.e. Gaussian mixture models*	Great advantage when a good probability distribution is assumed. Provides a more flexible density method.	Requires a large number of training samples to overcome the curse of dimensionality. The distribution of the data is assumed. Unlabeled outliers in training affect the estimation of the covariance matrix.
Probabilistic nonparametric *i.e. Kernel density estimator*	Flexible density model. Possible configuration of the kernel width h on each feature direction. Low computational cost for training. The density estimation is only influenced locally.	Requires a large number of training samples to overcome the curse of dimensionality. Expensive computational cost for testing. Limited applicability of the method when there is a large dataset in high-dimensional feature spaces.
Distance-based i.e. k-NN	Rejects parts of the feature space which are within the target distribution. Lack of configuration parameters, besides k; therefore, it relies completely on the training samples.	Scale sensitive due to the use of distances in the evaluation of test objects. Performance affected when unlabeled outliers are presented in training. Sensitive to noise.

Table 1. Summary of the main characteristics of the novelty detection approaches.

3. Case study

In order to illustrate the practical implementation of novelty detection in an industrial application, an interesting case study is proposed next. Indeed, as it has been mentioned, currently, due to the worldwide market situation, the industrial sector is being subjected to a high degree of competitiveness. Critical sectors as the automotive industry are investing in higher levels of quality and safety assessment procedures in order to reduce costs without compromising the attributes of their mechanical manufactured assets. In regard to the automotive rotatory mechanical components, such as the electrical-assisted power steering columns (EPS), end-of-line tests (EOLs) are carried out to analyze their performances. The EPS column is rotated by a test machine in order to quantify the required torque to perform a complete revolution of the EPS column without the influence of any external load. Thus, if the recorded torque is compared with a reference pattern for decision support purposes, then, the EOL test is complete. However, the condition monitoring of the EOL machines, as the represented in **Figure 3**, has

http://dx.doi.org/10.5772/67531

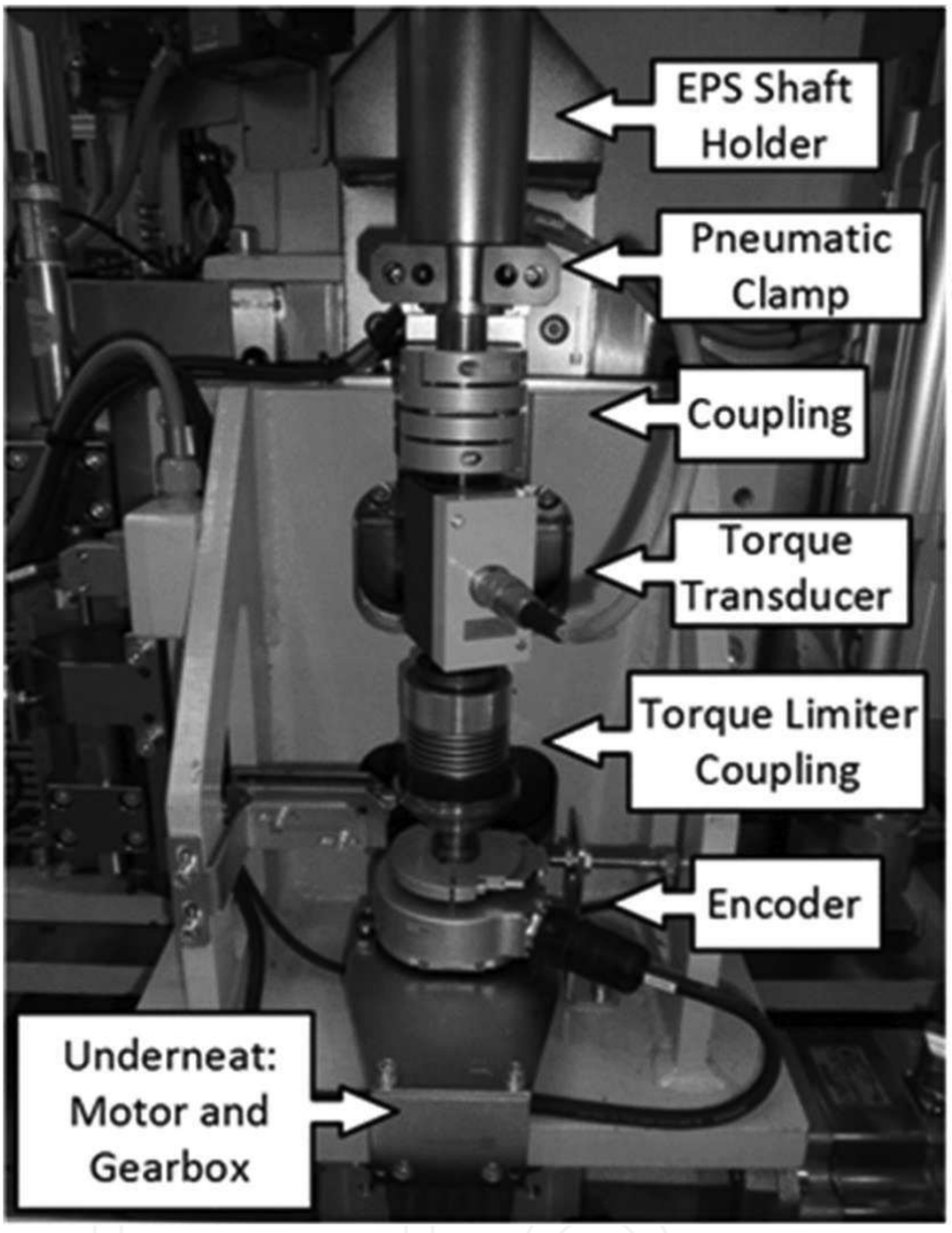

Figure 3. End-of-test machine, composed by a servomotor, a gearbox, an encoder, a torque transducer and a pneumatic clamp to hold the intermediate shaft of the EPS column.

not being attended classically. The maintenance program is limited to a preventive approach, leading to torque response deviations due to EOL machine degradation that are not detected by the machine operator until an evident malfunction. In this regard, the detection and identification of EOL malfunctions during its operation becomes an impactful contribution to the sector and is considered a challenging condition-based monitoring scenario.

In this work, a specific end-of-line test machinery is analyzed. The system under monitoring is based on an electrical drive, where a 1.48 kW at 3000 rpm servomotor connected to a 60:1 reduction gearbox emulates the input torque of the steering wheel to perform a 180° turn in order to evaluate the mechanical performances of power-assisted steering systems. The

measurement equipment is focused on the acquisition of the torque signal and the rotatory shaft position from the encoder. Data acquisition is done at 1 kHz of sampling frequency by means of a NI cDAQ-9188, composed by the modules NI-9411 and NI 9215.

The torque induced by the drive is expected to follow a specific predefined set point pattern. However, these test systems present two main limitations: first, if the test machine does not generate the input of the torque set point correctly, an inaccurate result is obtained during the assessment of the power-assisted steering system, leading to the nonvalidation of the components under test and second, the lack of malfunctions' characterization over the testing machine, since the faults' variability and appearance in the torque generation test are unpredictable. Thus, this work presents an electromechanical system novelty detection approach, based on the temporal torque signal characterization by statistical time features and the evaluation of different novelty detection algorithms (probabilistic, domain-based and distance-based), for novelty assessment.

In order to analyze the performance of the proposed methodology, some faulty conditions have been induced in the machine to provoke different severity degrees of a common fault scenario. Three operating scenarios are considered, that is, healthy, H, a coupling low wear, C_{LW}, and a coupling high wear, C_{HW}. The coupling wear fault is emulated by employing two different intermediate elastomers in the torque limiter coupling, each one with different dynamic torsional stiffness (DTS). The values of the DTS of the pieces under test are all lower than the standard used in the machine in order to emulate classical wear, thus, 2580 Nm/rad corresponds to C_{LW} and 2540 Nm/rad to C_{HW}.

3.1. Method

During the test, the assisting motor of the EPS is not powered. The test starts smoothly in a clockwise direction for the first 45° until a speed set point is reached. The acceleration time depends on the drive capability. During the next 360°, the speed is fixed at the set point, in this case 15 rpm. The last additional 45° is for a mild brake of the EPS column under test. Then, the same procedure is employed to return to the original start point in the opposite direction. The drive is applied to the steering shaft of the EPS. Then, the torque signal analysis is carried out during the stationary speed set point corresponding to a 360° turn of the EPS column. It is expected that malfunctions and anomalies could appear during segments of the revolution of the EPS column; therefore, the segmentation represents a viable strategy to gain resolution during the characterization. That is, the 4-second torque signal (time taken to perform the 360° turn) is segmented in four parts of 1 second. A set of five statistical time-domain features is calculated from each segment of the torque signal. The proposed features are listed in **Table 2**. These features have been successfully employed in different studies for electromechanical systems' fault detection [19]. Therefore, a total of 20 features are calculated from each torque signal measurement.

High-dimensional datasets complicate the learning task of novelty detection as well as multiclass classification methods, because of the possible presence of nonsignificant and redundant information in the data, compromising the proper convergence of the algorithms. Indeed, the empty space phenomenon states that to cover the whole space, it needs a number of samples that grows

http://dx.doi.org/10.5772/67531

Root mean square (RMS)	$\sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_k)^2}$	Variance	σ^2
Shape factor	$\frac{RMS}{\frac{1}{n}\sum_{k=1}^{n}\lvert x_k\rvert}$	Skewness	$\sum_{k=1}^{n}\frac{(x_k-\overline{x})^3}{n\sigma^3}$
Crest factor (CF)	$\frac{\max(x)}{RMS}$	Kurtosis	$\sum_{k=1}^{n}\frac{(x_k-\overline{x})^4}{n\sigma^4}$

Table 2. Statistical time-domain features used for torque signal characterization.

exponentially with dimensionality. Thus, the curse of dimensionality implies that in order to learn successfully, it needs 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 learned as the dimension of the data increased. For these reasons, there is a necessity to apply dimensionality reduction techniques in condition monitoring applications. Thus, in order to analyze the performance of different novelty detection approaches, two main dimensionality reduction approaches are applied over the 20-dimensional vectors, that is, Principal Component Analysis, PCA, and Laplacian Score, LS.

Indeed, the dimensionality reduction strategies differ in the criteria applied over the data in order to reach a reduced feature space. PCA is one of the most commonly used techniques for unsupervised dimensionality reduction. It aims to find the linear projections that best capture the variability of the data [13]. Another well-known technique is the LS, where the merit of each feature is measured according to its locality preservation power. A nearest neighbor-based graph is constructed from the training set and analyzed to rank each feature individually according to a weighting approach selected for the graph's edges. To rank each feature, its LS is computed, which is a measure of the extent to which 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 [20].

Finally, the necessity of evaluating the novelty detection performance is critical. The use of a particular score depends on multiple interests, and then, the analysis of complementary scores represents the most interesting solution. Next, the most useful and common scores in a discrete scenario are described in order to be used later during the analysis of the experimental results.

- Accuracy and classification error (1-accuracy): One of the most frequent scores used to evaluate discrete classification in electromechanical diagnosis is accuracy. This score is indicative of the classification error committed while evaluating, in our case, two classes,

$$Accuracy = \frac{FP + FN}{N} \tag{9}$$

where FP is the number of false positives, FN is the number of false negatives and N, the total number of analyzed measures. Two novelty detection approaches could exhibit the same accuracy but provide a different novelty ratio for each class (normal data and novelty data).

- True positive rate (recall or sensitivity): This measure provides a proportion of one kind of sample that was correctly assessed. But it only evaluates the positive cases,

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

where TP is the number of true positives.

- Precision: This performance metric evaluates the correct classification of the positive class,

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

- F-measure: This score can help to solve any contradiction that may appear between precision and recall scores. F-measure leaves out the TN performance. Several versions exist. The most common expression is,

$$F_1 = \frac{2[Precision * Recall]}{[2 * Precision + Recall]} \tag{12}$$

3.2. Experimental results

In order to expose the novelty detection performances, the outline of the experimental results is presented as follows: The initial database is characterized by the proposed set of features, then both feature reduction approaches are applied and, finally, over each reduced set of features, the three novelty detection approaches are applied. The application of the novelty detection is done sequentially, that is, first, the data corresponding to the healthy, H, operating scenario is characterized by the novelty model. Second, the first fault operating scenario, C_{LW}, is presented as well as additional measures of the H operating scenarios. At this point, the performance of the novelty detection model is analyzed. Third, the novel data identified is included in the upgraded version of the novelty model by retraining, and over this updated novelty model, the second fault operating scenario, C_{HW}, is presented as well as the additional measures of the C_{LW} and H operating scenarios. At this point, the performance of the novelty detection models is analyzed again. Finally, the novel data identified is included in an upgraded version of the novelty model by a new retraining.

Three novelty detection methods have been implemented, that is, the mixture of Gaussians as the probabilistic approach, one-class support vector machines as the domain-based novelty detection approach and, finally, k-nearest neighbors as the distance-based novelty detection approach. Next, the PCA variant of the novelty detection methodology is shown in **Figure 4**.

The proposed scores during the assessment of the novelty detection models in front of a new set of measurements are shown next. Thus, in **Table 3**, the scores in regard to the PCA feature reduction and the novelty detection models 'performance, dealing with the projection of new measurements corresponding to H and C_{LW} operating scenarios over the novelty models

http://dx.doi.org/10.5772/67531

trained with the *H* operating scenario, can be seen. It should be noted that, in regard to all the scores shown in the next tables, a 10-fold cross validation strategy has been considered, in which the mean and the dispersion ratio of the obtained scores is shown.

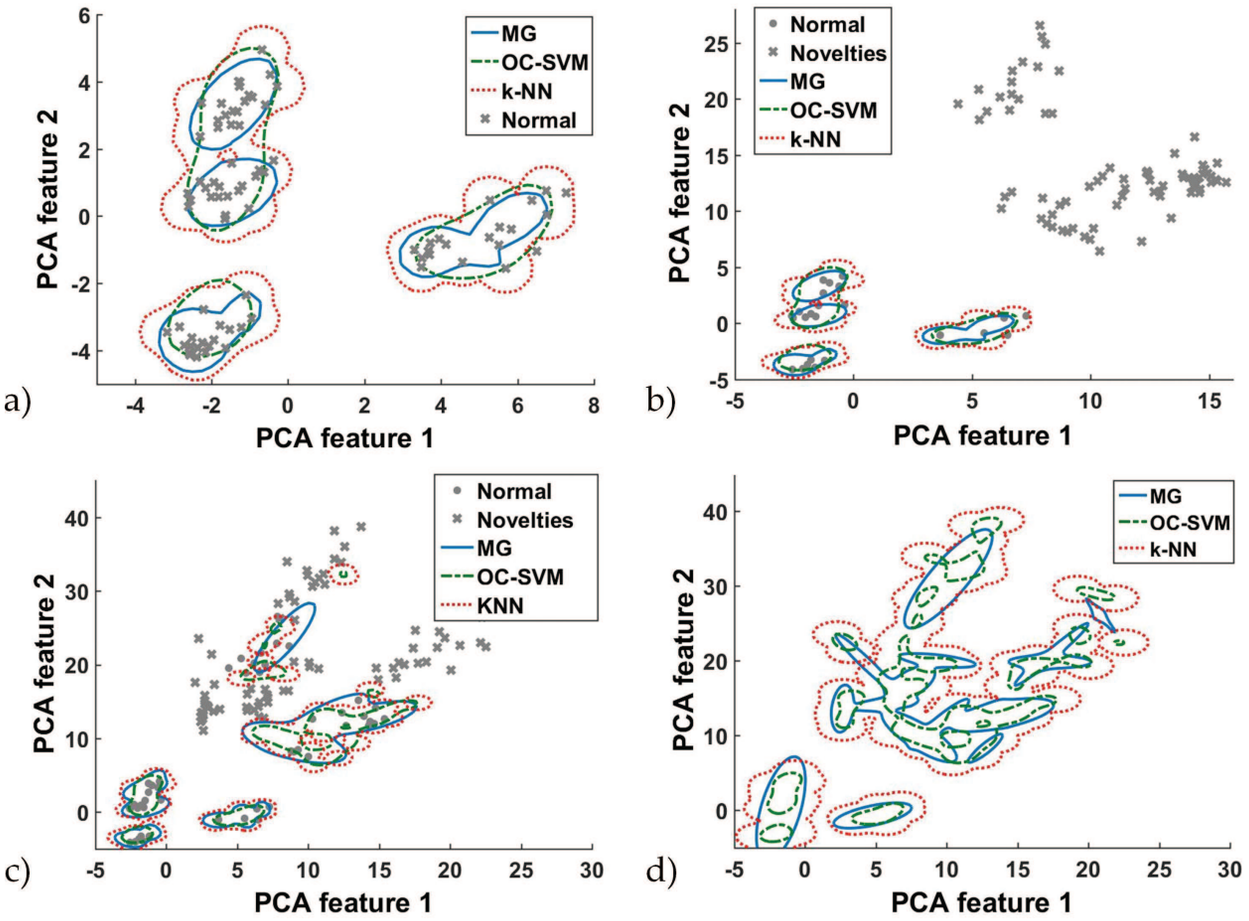

Figure 4. Performance of the three novelty detection models in a PCA-based 2-dimensional feature space. MG mixture of Gaussian, OC-SVM, one-class support vector machine and *k*-NN, *k*-nearest neighbor. (a) Novelty models' boundaries during the characterization of the *H* operating scenario. (b) Projection of new measurements corresponding to *H* and C_{LW} operating scenarios over the novelty models trained with the *H* operating scenario. (c) Novelty model boundaries' update by the incorporation of the *H* and C_{LW} operating scenarios' measures detected as novelties and the projection of new measurements corresponding to *H*, C_{LW} and C_{HW} operating scenarios over the novelty models. (d) Novelty model boundaries' update by the incorporation of the *H*, C_{LW} and C_{HW} operating scenarios' measures detected as novelties.

	MG	OC-SVM	k-NN
Accuracy	0.912 (±0.033)	0.952 (±0.025)	0.950 (±0.034)
Recall	1.000 (±0.000)	1.000 (±0.000)	1.000 (±0.000)
Precision	0.898 (±0.035)	0.942 (±0.029)	0.949 (±0.038)
F1 score	0.946 (±0.019)	0.970 (±0.015)	0.974 (±0.02)

Table 3. Associated scores to the PCA and novelty detection performance dealing with the class *H* as normal and C_{LW} as novelty.

Also, in **Table 4**, the scores in regard to the PCA feature reduction and the novelty detection models' performance, dealing with the projection of new measurements corresponding to H, C_{LW} and C_{HW} operating scenarios over the novelty models trained with the H and C_{LW} operating scenarios, can be seen.

	MG	**OC-SVM**	**k-NN**
Accuracy	0.930 (±0.025)	0.863 (±0.014)	0.966 (±0.014)
Recall	0.997 (±0.005)	0.995 (±0.006)	0.985 (±0.005)
Precision	0.902 (±0.033)	0.822 (±0.013)	0.962 (±0.022)
F1 score	0.947 (±0.018)	0.901 (±0.01)	0.973 (±0.010)

Table 4. Associated scores to the PCA and novelty detection performance dealing with the H and C_{LW} as known operating scenarios and C_{HW} as novelty.

Similarly, next, the LS variant of the novelty detection methodology is shown in **Figure 5**.

The proposed scores during the assessment of this variant of novelty detection models in front of a new set of measurements are also shown next. Thus, in **Table 5**, the scores in regard to the LS feature reduction and novelty detection models' performance, dealing with the projection of new measurements corresponding to H and C_{LW} operating scenarios over the novelty models trained with the H operating scenario can be seen.

	MG	**OC-SVM**	**k-NN**
Accuracy	0.956 (±0.021)	0.971 (±0.026)	0.967 (±0.028)
Recall	1.000 (±0.000)	1.000 (±0.000)	1.000 (±0.000)
Precision	0.940 (±0.023)	0.965 (±0.031)	0.960 (±0.033)
F1 score	0.969 (±0.012)	0.982 (±0.016)	0.980 (±0.017)

Table 5. Associated scores to the LS and novelty detection performance dealing with the class H as normal and C_{LW} as novelty.

Also, in **Table 6**, the scores in regard to the LS feature reduction and novelty detection models' performance, dealing with the projection of new measurements corresponding to H, C_{LW} and C_{HW} operating scenarios over the novelty models trained with the H and C_{LW} operating scenarios can be seen.

	MG	**OC-SVM**	**k-NN**
Accuracy	0.910 (±0.016)	0.841 (±0.032)	0.890 (±0.014)
Recall	0.955 (±0.017)	0.925 (±0.021)	0.865 (±0.044)
Precision	0.922 (±0.027)	0.840 (±0.044)	0.940 (±0.030)
F1 score	0.938 (±0.012)	0.879 (±0.021)	0.899 (±0.014)

Table 6. Associated scores to the LS and novelty detection performance dealing with the H and C_{LW} as known operating scenarios and C_{HW} as novelty.

http://dx.doi.org/10.5772/67531

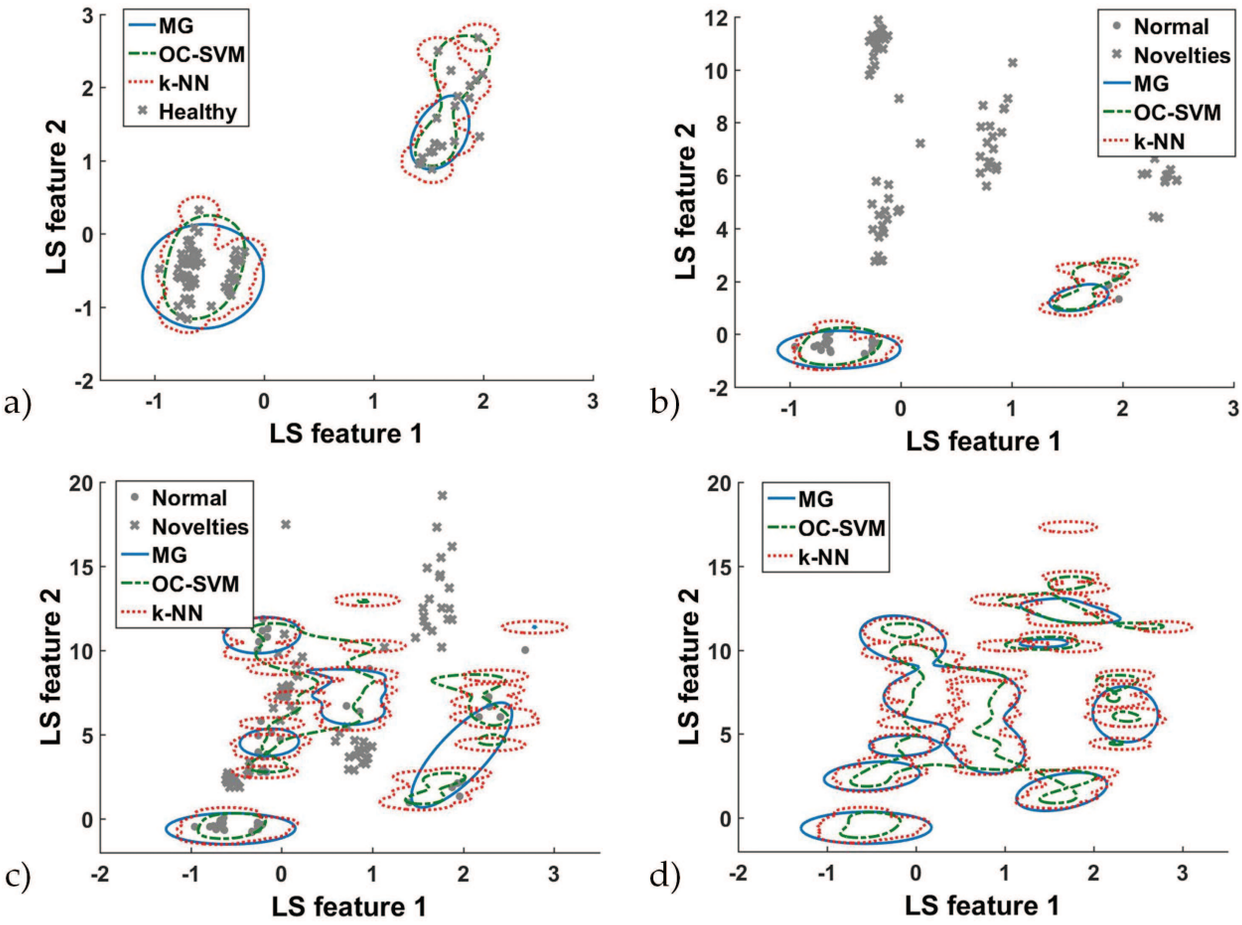

Figure 5. Performance of the three novelty detection models in a LS based 2-dimensional feature space. MG mixture of Gaussian, OC-SVM, one class support vector machine, and *k*-NN, *k*-nearest neighbor. (a) Novelty models boundaries during the characterization of the *H* operating scenario. (b) Projection of new measurements corresponding to *H* and C_{LW} operating scenarios over the novelty model trained with the *H* operating scenario. (c) Novelty model boundaries update by the incorporation of the *H* and C_{LW} operating scenarios measures detected as novelties, and projection of new measurements corresponding to *H*, C_{LW} and C_{HW} operating scenarios over the novelty models. (d) Novelty model boundaries update by the incorporation of the *H*, C_{LW} and C_{HW} operating scenarios measures detected as novelties.

In regard to the feature reduction effect over the novelty detection performance, it has been taken into account that both methods, PCA and LS, represent linear approaches to the reduction of the initial 20-dimensional feature set. This premise is not a limitation to the analysis of the novelty detection models considered; however, the feature space could be further improved in order to maximize the obtained results for the specific application.

Independently of the novelty detection model, the first test stage, that is, the assessment of new measurements corresponding to *H* and C_{LW} operating scenarios over the novelty models trained with the *H* operating scenario, shows a clear superiority of the LS approach. The accuracy obtained with the LS approach reaches till 97% and is, in all cases, better than using PCA that reaches a maximum of 92%. However, the second test stage, that is, the assessment of new measurements corresponding to *H*, C_{LW} and C_{HW} operating scenarios over the novelty models trained with the *H* and C_{LW} operating scenarios, shows a clear superiority of the PCA approach. The accuracy obtained with the PCA approach reaches till 96% and is, in most of the cases, better than using LS that reaches a maximum of 91%.

This effect is reasonable in dealing with the available data because the LS approach allows a better representation in terms of novelty detection. That is, considering that all the available data corresponds to the unique class *normal* or *known*, the performance of the novelty detection model will be enhanced if less data dispersion is presented. In this sense, the LS feature space shows a more compact projection of the data, at least during the first test stage, a fact that facilitates the definition of the novelty detection boundaries and the posterior accuracy. However, dealing with the second test stage, the maximization of the variance by means of the PCA avoids false negatives. In fact, the dispersion of data is desired when complexity of data is considered, since new operating scenarios could be assessed as known measurements. This performance of the feature reduction techniques over the novelty detection performance is a critical aspect during the condition monitoring configuration, since a trade-off between feature space complexity and data dispersion must be reached. Nevertheless, in this case study, the proposed novelty detection methodology including both the feature reduction techniques exhibits high ratios of performances.

In regard to the novelty detection models, independently of the feature reduction technique, the first test stage, that is, the assessment of new measurements corresponding to H and C_{LW} operating scenarios over the novelty models trained with the H operating scenario, shows a clear superiority of the OC-SVM and k-NN approaches in terms of accuracy, precision and F1-score, considering that the recall is maximum in all three cases. However, the second test stage, that is, the assessment of new measurements corresponding to H, C_{LW} and C_{HW} operating scenarios over the novelty models trained with the H and C_{LW} operating scenarios, shows a superiority, of the k-NN approach, mainly in terms of accuracy and precision, although the MG shows also good behavior in terms of recall.

In fact, as it has been mentioned, the probabilistic novelty detection approach, represented by the MG technique, assumes a data dispersion that, dealing with unknown operating scenarios, cannot be the optimum. This fact is smoothened when the data density increases, since more information is available in order to infer a proper PDF. In case of OC-SVM and k-NN, both techniques show wide novelty detection boundaries, which allow a better characterization of the data distribution by means of good generalizations. However, it must be taken into account that, qualitatively, a more complex partition of the feature space is reached by the k-NN, and although, as it has been explained, this can be controlled by the value of k, such tuning is not trivial and, then, OC-SVM represents a more simple solution.

4. Conclusions

A condition monitoring scheme for novelty detection is applied to an industrial end-of-line test machinery of electrical-assisted power steering columns, where the *healthy* data is the initial available information. The fault conditions considered consist of two severities of one commonly presented fault in the mechanical parts of the electrical drive of the test machine, coupling wear. The fault condition is presented in two stages, in order to analyze the detection and learning capabilities of the considered approaches. These fault severities represent a challenge for the data analysis due to the similitudes between the torque signals characterizing each fault.

http://dx.doi.org/10.5772/67531

Six variants of the methodology are proposed and analyzed. Thus, two feature reduction approaches by means of PCA and LS are considered in order to emphasize the information contained in the 20-dimensional vectors of statistical time-based features in which each torque measurement is characterized. Later, three novelty detection modelling approaches are introduced and implemented, that is, the probabilistic method by means of the mixture of Gaussians, domain-based methods by means of one-class support vector machine and distance-based methods, by means of *k*-nearest neighbors. A comparison and analysis between the novelty models and the feature reduction procedures is performed to analyze the proper selection of novelty models for these scenarios. The results have shown that the combination of PCA as feature reduction and *k*-NN as the novelty detection model reaches, in general, the best-considered scores, mainly the accuracy, 96 and 90%, and precision, 96 and 94%. However, the OC-SVM alternative must also be considered due to its simpler configuration requirements and good performances.

Acknowledgements

The authors gratefully acknowledge the financial support from the MINECO of Spain, under the Project CICYT TRA2013- 46757-R, the Generalitat de Catalunya under the Project GRC MCIA, Grant n° SGR 2014-101, and the CONACyT of Mexico under the scholarship 313604 grant. Also, the authors would like to thank the support and the access to the friction test machine database provided by MAPRO Sistemas de ensayo S.A., especially to Álvaro Istúriz and Alberto Saéz.

Author details

Miguel Delgado Prieto[1]*, Jesús A. Cariño Corrales[1], Daniel Zurita Millán[1], Marta Millán Gonzalvez[2], Juan A. Ortega Redondo[1] and René de J. Romero Troncoso[3]

*Address all correspondence to: miguel.delgado@mcia.upc.edu

1 MCIA Research Center, Technical University of Catalonia, Terrassa, Spain

2 MAPRO Test and Assembly Systems S.A., Sant Fruitós de Bages, Spain

3 University of Guanajuato, Salamanca, Mexico

References

[1] Yin S., Ding S. X., Xie X., Luo. H. A review on basic data-driven approaches for industrial process monitoring. IEEE Transactions on Industrial Electronics. 201;**61**(11):6418–6428. doi:10.1109/TIE.2014.2301773

[2] Davies, A., editor. Handbook of condition monitoring: Techniques and methodology. 1st ed. Ireland: Springer Science & Business Media Dordrecht; 1998. 564 p. doi:10.1007/978-94-011-4924-2

[3] Ahmad R., Kamaruddin S. An overview of time-based and condition-based maintenance in industrial application. Computers & Industrial Engineering. 2012;**63**(1):135–149. doi:10.1016/j.cie.2012.02.002

[4] Boashash B., editor. Time-frequency signal analysis and processing: A comprehensive reference. 2nd ed. UK: Academic Press; 2016. 1020 p

[5] Möller D.P.F. Digital manufacturing/industry 4.0. In: Möller D.P.F., editor. Guide to computing fundamentals in cyber-physical systems. 1st ed. Switzerland: Springer; 2016. pp. 307–375. doi:10.1007/978-3-319-25178-3_7

[6] Teti R., Ferretti S., Caputo D., Penza M., D'Addona D. M. Monitoring systems for zero defect manufacturing. In: Teti R., editor. Eighth CIRP conference on intelligent computation in manufacturing engineering, 18 July, Naples. Procedia CIRP; 2013. pp. 258–263. doi:10.1016/j.procir.2013.09.045

[7] Pimentel M. A. F., Clifton D. A., Clifton L., Tarassenko L. A review of novelty detection. Signal Processing. 2014;**99**(6):1165–1684. doi:10.1016/j.sigpro.2013.12.026

[8] Jumutc V., Suykens J. A. K. Multi-class supervised novelty detection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2014;**36**(12):2510–2523. doi:10.1109/TPAMI.2014.2327984

[9] Jyothsna V., Rama Prasad V. V. A review of anomaly based intrusion detection systems. International Journal of Computer Applications. 2011;**28**(7):26–35

[10] Chandola V., Banerjee A., Kumar V. Anomaly detection: A survey. ACM Computing Surveys. 2009;**41**(3):15–58. doi:10.1145/1541880.1541882

[11] Markou M., Singh S. Novelty detection: A review—Part 1: Statistical approaches. Signal Processing. 2003;**83**(12):2481–2497

[12] Markou M., Singh S. Novelty detection: A review—Part 2: Neural Network Based Approaches. Signal Processing. 2003;**83**(12):2499–2521. doi:10.1016/j.sigpro.2003.07.019

[13] Duda R. O., Hart P. E., Stork D. G., editors. Pattern classification. 2nd ed. Chichester: Wiley; 2000. 680 p

[14] Reynolds D. Gaussian mixture models. In: Li S. Z., Jain A. K., editors. Encyclopedia of biometrics. 2nd ed. United States: Springer US; 2015. pp. 827–832. doi:10.1007/978-1-4899-7488-4_196

[15] Parra L., Deco, G., Miesbach, S. Statistical independence and novelty detection with information preserving nonlinear maps. Neural Computation. 1996;**8**(2):260–269. doi:10.1162/neco.1996.8.2.260

[16] Elmoataz A., Lezoray O., Nouboud F., Mammass D., Meunier J., editors. Image and signal processing. 1st ed. New York: Springer; 2010. 602 p

[17] Bishop C. M., editor. Pattern recognition and machine learning. 1st ed. New York: Springer-Verlag; 2006. 740 p

[18] Tax D. M. J., Duin R. P. W. Support vector domain description. Pattern Recognition Letters. 1999;**20**(11–13):1191–1199. doi:10.1016/S0167-8655(99)00087-2

[19] Delgado-Prieto M., Cirrincione G., Espinosa A. G., Ortega J. A., Henao H. Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks. IEEE Transactions on Industrial Electronics. 2013;**60**(8):3398–3407. doi:10.1109/TIE.2012.2219838

[20] He X., Cai D., Niyogi P. Laplacian score for feature selection. In: Weiss Y., Schölkopf B., Platt J. C., editors. Neural information processing systems. Vancouver: MIT Press; 2005. pp. 1–8

IntechOpen