J. A. Carino et al., "Fault Detection and Identification Methodology Under an Incremental Learning Framework Applied to Industrial Machinery," in *IEEE Access*, vol. 6, pp. 49755-49766, 2018.

DOI: 10.1109/ACCESS.2018.2868430

Fault detection and identification methodology under an incremental learning framework applied to industrial machinery

Jesus A. Carino, *Student Member*, *IEEE*, Miguel Delgado-Prieto, *Member IEEE*, Jose Antonio Iglesias, *Member IEEE*, Araceli Sanchis, Member *IEEE*, Daniel Zurita, *Student Member*, *IEEE*, Marta Millan, Juan Antonio Ortega Redondo, *Member*, *IEEE*, Rene Romero-Troncoso, *Senior Member*, *IEEE*

Abstract— **This paper presents an Industrial Machinery Condition Monitoring Methodology based on Ensemble Novelty Detection and Evolving Classification. The methodology aims to solve the main challenges presented when classical electromechanical-system monitoring approaches are applied in an industrial framework. Some of these challenges are: 1) the presence of unknown events caused by fault modes or new operating conditions, 2) the lack of information as starting knowledge which correspond only to the nominal healthy condition and 3) the incorporation of new events into the available starting knowledge without requiring an historical database of measurements of the monitored machine. The proposed methodology has four stages. First, double feature calculation and reduction over the available physical magnitudes in order to optimize the novelty detection and fault classification performances. Second, a novelty detection stage using an ensemble of One-Class Support Vector Machines to identify new events not previously considered. Third, a diagnosis stage by means of an evolving classifier, eClass0 or eClass1, to recognize the considered patterns. And fourth, re-training stage to include new scenarios to the novelty detection and fault classification models. The effectiveness of the methodology has been verified by experimental results obtained from an automotive End-of-Line test machine.**

Index Terms— Condition Monitoring; Fault Detection; Machine Learning; Novelty Detection.

I. INTRODUCTION

Currently, one of the most important challenges towards the implementation of reliable Condition-Based Monitoring (CBM) schemes in the industrial sector, refers to the management of unexpected events. That is, classical CBM is supported by a set of malfunction conditions that, a priori characterized, can be recognized later during the diagnosis process. However, the presence of faults not previously considered, or even deviations of performance over the nominal behaviour, represent common conditions that lead to diagnosis errors. Indeed, classical CBM schemes are being redefined by the separation of the fault detection and identification (FDI) tasks in two different stages. Novelty detection is the first stage, and becomes critical since the objective is to detect whether the measurement under analysis corresponds to a *known* or *unknown* condition. In this stage, no specific information about the fault is provided. Thus, the diagnosis of the fault is the objective of the second stage, and it presents its own challenges and obstacles to identify the fault, that can be handled independently from the previous one.

Certainly, such CBM scheme including novelty detection is being considered as one of the new trends in fault diagnosis research for electrical machines [1]. However, most of the related works available up to now correspond to static approaches, where the healthy and a set of fault conditions are previously characterized following a classical diagnosis approach, and uncorrelated events are detected and set apart [2]. Nevertheless, in most of industrial applications, just the nominal operating condition is available (the healthy condition), which, from one side, makes unfeasible a previous characterization of fault conditions and, from the other side, requires the proposal of adaptive CBM schemes capable of update its available knowledge and, then, its diagnosis capabilities.

Indeed, an important limitation of classical approaches is that the possibility of incursion of new classes to the base knowledge is not considered. That is, traditional data-driven CBM methodologies face the knowledge increase by means of a batch scheme, where a complete retrain of the whole diagnostic model structure is carried out with the data combining the initial and new knowledge. However, storing all the historical measurements is not a desired solution and, moreover, the complexity of the retraining process is increased as the data is accumulated, which represents an unsustainable approach. As alternative, adaptive strategies for novelty detection are being proposed, first, ensemble-based, and second, incremental approaches [3]–[5]. The objective of both is to provide a more flexible option capable to work in on-line mode. That is, the advantages of these methods focus in lessen the computation efforts of the models, decrease the number of configuration parameters, and provide the capability to update the models without necessity of the base knowledge used for the initial training. Dealing with fault diagnosis, a third strategy is being also considered, the evolving approach. Indeed, considering the need of data labelling for diagnosis purposes, such evolving strategy offers the possibility of model growing while optimizing the global computational complexity.

Such adaptive approaches for novelty detection and fault diagnosis, however, present important restrictions that must be overcome considering the specific application. The processing of the available signals, and estimation of relevant features to analyse the machine condition can only be performed with information of the healthy condition. Therefore, the characterization process of faults to emphasize specific patterns is not an affordable option. Some adaptive approaches have been successfully applied in industrial plant processes monitoring, network intrusion, aerospace vehicles, and others. Nevertheless, to the extent of the authors' knowledge, adaptive methodology considering both, novelty detection and fault diagnosis, has not been applied yet to industrial machinery, which requires specific consideration to allow its applicability.

In this work, the implementation of an adaptive CBM scheme for an electromechanical system with novelty detection and fault diagnosis capabilities, considering just the nominal condition as initial knowledge, is presented. For this objective, the available stator currents of an electric motor are considered for a time-frequency inspection in order to extend the applicability of the method to non-stationary operating conditions. Then, a set of statistical time-based features is estimated in order to characterize the stator current signature along the operating cycle. Next, an ensemble based novelty detection model is proposed to describe the available knowledge and identify novelty conditions. Finally, an adaptive diagnosis structure is implemented by the evolving classifiers *eClass0* and *eClass1*, whose performance is compared with a classical neural network based structure.

Three main contributions are proposed in this work. First, a new methodology for incremental learning of the novelty detection and diagnosis models starting from data corresponding only to the nominal healthy condition. Second, the verification of the evolving classifier and ensemble based novelty detection as reliable and performing approach for CBM in front of classical schemes. And, third, the validation of the proposed methodology in front of multiple conditions, including *known* and *unknown* events, measured from an industrial electromechanical system provided by an automotive sector company.

Note that it is the first time that this methodology is applied to an industrial electromechanical system, and it represents a significant step towards the understanding of the adaptive condition based monitoring in such application field [6], [7].

This paper is organized as follows. A background regarding the state of the art of adaptive novelty detection and fault diagnosis methods is presented in Section II. In Section III, the proposed adaptive CBM method is described. The experimental system and the corresponding operating cycles are described in Section IV. The results obtained are presented and discussed in Section VI. Finally, conclusions are summarized in Section VII.

II. BACKGROUND

In regard with the implementation of an adaptive novelty detection stage, two strategies are considered mainly in the literature: incremental models and ensemble of one-class classifiers.

Incremental novelty detection models are commonly used in data streaming applications to cope with classical problems as the so called concept-drift, by including forgetting and adaptive capabilities to their structures [8]. For instance, Krawczyk *et al*., in [9], proposed an incremental one-class support vector machine based on a weighting matrix to adapt the knowledge' boundary to variations in the incoming data. This weighted approach lead to an improvement in classification of different data streams, especially with the presence of an incipient data drift. Similarly, Al-Behadili *et al.*, in [10], proposed an incremental parzen window kernel density estimator to address also the data drift problem. This approach obtained better results than the standard Support Vector Data Description (SVDD), nevertheless, to keep it computational efficient the user needs to define an initial number of clusters, in this case applying *k*-means algorithm, for each class. Indeed, most of the studied incremental approaches are developed to adapt models to current conditions of the monitored system, which means that past knowledge is considered obsolete and discarded [9]. Thus, incremental models are mainly applied within big data analytics, where a great deal of continuous data is available. The performance of such approach over electromechanical systems may be limited, considering the low inertia of multiple wear based faults and the necessity of multi-fault patterns recognition.

The ensemble of one-class classifiers is the other main alternative based on training one novelty detection model for each available new data set, combining later the outputs to determinate if the measurement under analysis corresponds to *known* condition or differs in some aspect from the available knowledge. In this sense, the work presented by Lazzaretti *et al.*, in [2], presents an ensemble of one-class classifiers to perform an automatic classification of voltage waveforms in electrical distribution networks. In such work, there is no clear division between the novelty detection and the fault diagnosis stages, therefore, both tasks are performed by an ensemble of Support Vector Data Descriptions (SVDD). Like most of the works dealing with novelty detection, the incursion of novel information is not faced; nevertheless, the method allows the addition of new SVDD models if data regarding a new type of fault is available.

In general terms, the use of an ensemble of one-class classifiers provides more design flexibility in comparison of the incremental based models. That is, dealing with an ensemble-based approach, a new model can be created when a new data set is detected; therefore, there is no loss of previous knowledge because it is retained within the set of models. In this sense, the discard of knowledge is user-dependent, by selecting the specific model to remove. Moreover, any novelty detection technique can be used to be part of an ensemble-based scheme. In this sense, dealing with electromechanical condition monitoring, where relative small sets of training data are usually available, a suitable option are the domain-based approaches One-Class Support Vector Machine (OC-SVM) or Support Vector Data Description (SVDD). For example, in [11], a method using SVDD is used to deal with an unbalanced and small sampled dataset for rotor severity classification. Dealing with ensemble approaches, some disadvantages are present as well, for instance, the necessity of an offline training stage for each new model. This fact requires that a representative set of data must be identified and temporally stored to train manually the corresponding new model.

Regarding fault diagnosis, the same both discussed strategies are also applicable with their respective modifications. Indeed, there is considerable literature on incremental learning and ensemble-based classifiers, and most of the characteristics discussed in the novelty detection side applies also for fault diagnosis [12]–[16]. Indeed, incremental and ensemble approaches have their variants for multi-class classification problem, mainly, the incremental Support Vector Machine (SVM) [17], and the ensemble of SVM classifiers [18], respectively. Other proposed models for incremental or ensemble fault diagnosis include an incremental version of probabilistic neural network (PNN) [19], a combination of discriminant analysis (DA) and principal component analysis (PCA) [20], decision trees based techniques [21], AdaBoost [22], Bagging [23] or Learn++ [13], among others. The differences and characteristics of the aforementioned incremental or ensemble models are focused on the classification accuracy improvement by modifying the training procedure, basically, adding robustness to outliers and improving the rules regarding the number of classifiers used in the ensemble-based scheme. It is important to note that, in general, such methods work under a supervised or semi-supervised environment, where the labeling process of a new data set as well as the model tuning is carried out manually and off-line.

However, as it has been aforementioned, dealing with fault diagnosis purposes, the evolving strategy is being considered as a superior adaptive approach in multiple studies [5]. For instance, in [7], an evolving approach is used for fault detection and identification. For novelty detection, the Recursive Density Estimation (RDE) calculation is used to detect outliers, meanwhile for fault identification an unsupervised evolving classifier AutoClass is used. Indeed, the fault diagnosis stage requires the consideration of a more complex data boundary structure. Unlike novelty detection problem, where a binary scenario is considered, the fault diagnosis applied to electromechanical system requires the consideration of a multi-fault scenario.

In this sense, the conclusions of some studies suggest that the computational complexity of an ensemble-based approach for diagnosis can lead to unfordable structures after different adaptions to new data sets. Evolving strategies, however, allow the possibility of modify the structure of a unique model in function of the different boundaries to be considered. Indeed, this evolving strategy avoids the risk of a complex ensemble-based fault diagnosis structure, in which the relations among the multiple models must be defined manually depending on their labels. A family of fuzzy-rule based evolving classifiers have been used in recent works, as for example in [24], based on eClass algorithms, in [25], based on simpleClass, or in [7], based on AutoClass. All of them provide an evolving and online solution for fault diagnosis under low-computational cost requirements.

eClass0 and eClass1, are two well-known and used evolving classifiers. Both approaches are Fuzzy-Rule-Based (FRB) and work under an online unsupervised framework. A set of prototypes (focal points) are selected from the stream of data with a Gaussian membership function to generate the corresponding fuzzy rules. A set of measurements, like the potential and age of the prototypes, are determined to change the fuzzy rules in case new measurements are available for re-training. Their architecture is different regarding the actions performed when a new measurement is evaluated after the activation of the rules; while eClass0 follows the typical construct of an FRB classifier with class labels as direct output, the eClass1 regresses over the feature vector using first-order multiple-input-multiple-output evolving Takagi-Sugeno (MIMO-eTS) models (MISO is also possible for two-class problems9i) and the normalized outputs per rule can be interpreted as the possibility of the data sample belonging to a certain class. It is important to stress that both methods are capable of including new classes as new information is presented and automatically tune the dynamically adapting parameters to define the classification boundaries for each class. A specific discussion about advantages and disadvantages of these methods can be found in [5], [24].

Therefore, dealing with the adaptive CBM implementation applied to an electromechanical system, the ensemble-based approach for novelty detection and the evolving approach for fault diagnosis, represent the most suitable solutions.

Indeed, considering real applications of electromechanical systems, generally, a small number of measurements per class is available. Once a fault condition is detected, the corresponding maintenance operations are immediately carried out. Also, due to the nature of the application domain, if an anomaly is detected, the user inspection is required to label the data; therefore, unsupervised and autonomous approaches are not convenient.

III. Experimental Setup

The experimental platform under study performs a friction test over the manufactured parts (steering system). It is important to mention that the EOL (End of Line) test machine applies its own algorithm to determine the healthy state of the Steering System, nevertheless the aim of this work is to monitor the proper function of the EOL test machine.

The end-of-line test machine is shown in Fig. 1, where a 1.48kW synchronous servomotor with 4 pair of poles, 3000 rpm of rated speed and a rated torque of 4.7Nm is connected to a 60:1 reduction gearbox.

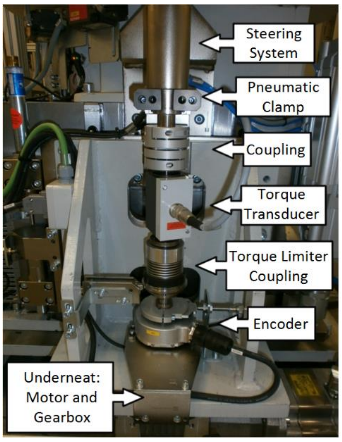

Fig. 1. Experimental setup that performs the test is composed by a servomotor, a gearbox, an encoder, a torque transducer and a pneumatic clamp to hold the intermediate shaft of the steering system.

An encoder of 9000 points of resolution is placed after the gearbox and is coupled to a 10Nm torque transducer by a torque limiter coupling. The other end of the torque transducer is coupled to the steering system under test. A scheme of the parts composing the EOL test machine is shown in Fig. 2. The acquisition system, in order to monitor the machine, register the torque signal of the transducer and the rotatory shaft position from the encoder. Data acquisition is done at 1 kHz of sampling frequency by a NI cDAQ-9188 composed by the modules NI 9411 and NI 9215.

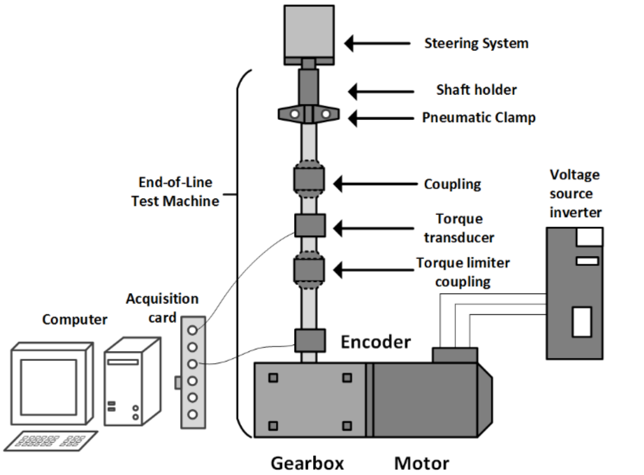

Fig. 2. Schematic of the EOL test machine under monitoring and the acquisition system.

A. *Description of the friction test*

The friction test consist on quantifying the DC value of the torque to rotate the steering system. The EOL machine forces the steering system column to follow a predefined speed profile. Such test consist of a complete clockwise turn (CW) and a complete counter clockwise turn (CCW). The speed profile performed by the test machine is shown in Fig. 3.

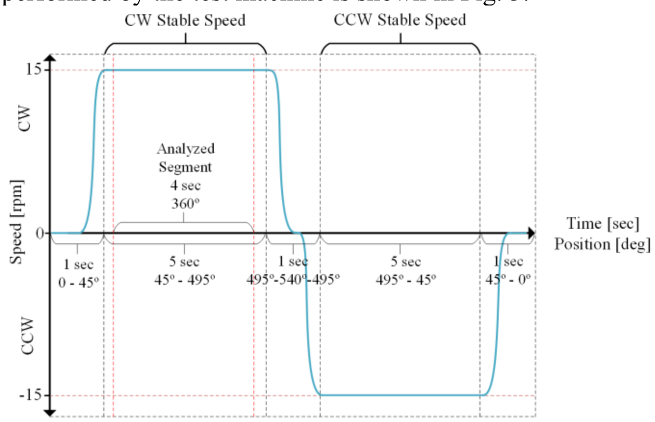

Fig. 3. Speed profile applied by the EOL test machine.

The test starts smoothly in a clockwise direction for the first 45º until a speed set point is reached. The corresponding acceleration time is related to the drive capability. During the subsequent 455º, the speed is fixed at a set point, in this case 15 rpm. After that, the same procedure is employed to return to the original start point in the opposite direction. The speed profile is shown in Fig. 3.

This speed profile provokes a torque in the shaft that it is measured by the torque transducer. An example of the torque measurement of a complete test under machine healthy conditions and the analyzed segment of three different conditions are shown in Fig. 4.

As can be seen in, a fault in the EOL test machine alters the torque signal form and, therefore, the statistical properties of the signal tend to vary (RMS value, crest and shape factor, etc.). On this work, the signal processing stage in the proposed methodology has the objective to highlight these alterations to perform the assessment of the machine without altering the part test itself; therefore, the methodology can be applied online.

A segment of 4 seconds is extracted of each torque signal measurement corresponding to the CW stable speed period. That is, the torque signal corresponding to the stationary speed set point during 360º rotation.

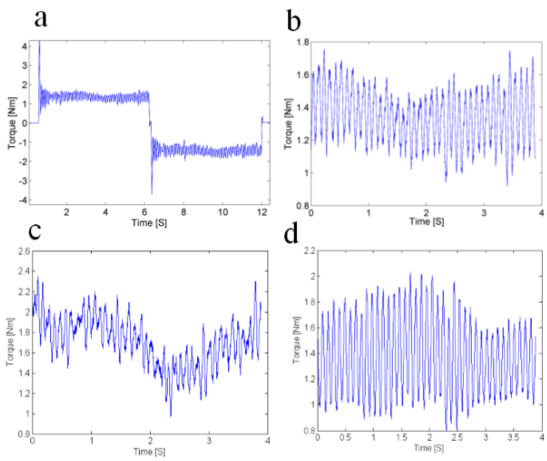

Fig. 4. Torque signal measured during the EOL test a) Complete torque measurement under machine healthy conditions b) Analyzed segment of a healthy machine measurement c)Analyzed segment of a machine misalignment fault d) Analyzed segment of a machine coupling wear fault

B. *Machine faults under consideration*

Some alterations have been induced in the machine to provoke two common fault conditions typically encountered on these EOL test machines; also, three severity levels are also considered for each fault. Therefore, two fault conditions with three severity degrees are present: three degrees of misalignment, MIS_5, MIS_6 and MIS_7, and three degrees of coupling wear, CW_1, CW_2, and CW_3.

The misalignment fault of the shaft has been provoked by a displacement of the base of the fixture holding the steering system. This induces a misalignment of the steering system respect to the shaft holder. Three degrees of severities are considered regarding the distance of horizontal displacement: 5mm (MIS_5), 6mm (MIS_6) and 7mm (MIS_7).

The coupling wear fault is emulated by employing three different intermediate elastomers in the torque limiter coupling, each one with different dynamic torsional stiffness (DTS). The values of the DTS of the used elastomers are all lower than the standard used in the healthy condition in order to emulate a degradation of the material.

The DTS values of the three elastomers corresponds to a low degradation degree, 2580Nm/rad (CW_1), intermediate degradation degree, 2540 Nm/rad (CW_2), and high degradation degree, 876 Nm/rad (CW_3).

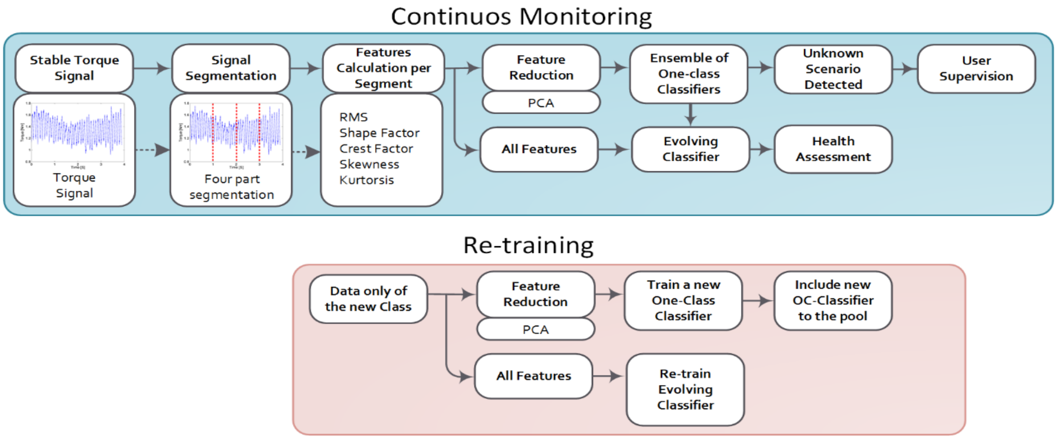

Fig. 5. Proposed methodology for the EOL test machine. The monitoring method is composed by a signal processing stage where statistical features are calculated and analyzed by a novelty detection and a multi-fault classification models to assess the condition of the machine.

Additionally, a sliding malfunction is caused by varying the tightening torque of the screws of the coupling between the torque transducer and the pneumatic clamp. The screws are loosened 0.5 Nm from the nominal tightening measured by a torque wrench. The measurements corresponding to this fault are going to be considered as an emerging novelty condition (*Nc*).

IV. METHODOLOGY

The majority of condition-based monitoring schemes operate on the premise that *a priori* information of healthy and faulty scenarios of the monitored machine is available. However, in real-world industrial applications, information regarding faulty conditions is not always accessible; therefore, additional problems arise and classical approaches need to be adapted. The main considerations to overcome are:

- The capability to work under the premise that only information of the healthy condition is initially available.
- The adaptation of the novelty detection and fault identification models to incorporate new scenarios detected without the need of a database that stores measurements of the previous scenarios already incorporated.

Such requirements are addressed on this work by means of the proposed condition-monitoring methodology depicted in Fig. 5. The general structure of the methodology is intended to serve as a guide to extend such condition monitoring scheme to other electrical machine based systems presenting the same aforementioned circumstances; yet, the modules to process the data and the feature calculation have been proposed to obtain relevant information dealing an automotive EOL test machine.

The proposed method is composed by two stages: a continuous monitoring stage to assess the condition of the machine and a re-training stage to include new information to the novelty detection and fault detection models.

During the continuous monitoring stage, a torque signal analysis is performed, then the novelty detection and fault identification stages assess if the measurement analyzed of the machine correspond to a: healthy condition, faulty condition or novel condition.

If a novel condition is detected a user analyze the machine to find the cause of the anomaly. If the user confirms that the novel condition corresponds to a new fault, the re-training stage is triggered to incorporate the new fault to the novelty detection model and the fault identification model.

A. Continuos Monitoring Stage

First, a torque signal analysis is carried out during the stationary speed set point corresponding to a 360º turn of the steering system. It is expected that malfunctions and anomalies could be reflected in the torque signal during segments of the revolution of the steering system, therefore, the segmentation represents a viable strategy to gain resolution during the characterization. Thus, the four seconds torque signal is segmented in four parts of 1 second. The number of segments chosen represents a tradeoff between resolution and total number of features. A larger number of segmentations increase the resolution but also increase the number of features, and could lead to overfitted models, meanwhile choosing a lower number of segments could not provide enough resolution.

A set of five statistical time-domain features are calculated from each segment of the torque signal. The proposed features are listed in Table I. These features have been successfully employed in different studies for electromechanical systems fault detection [26]. Therefore, a total of 20 features are calculated from each torque signal measurement.

TABLE I.
STATISTICAL TIME FEATURES

Root Mean Square (RMS)	$RMS = \sqrt{\frac{1}{n} \cdot \sum_{k=1}^{n} (x_k)^2}$	(10)
Shape Factor	$SF = \frac{RMS}{\frac{1}{n} \cdot \sum_{k=1}^{n} \lvert x_k \rvert}$	(11)
Crest Factor (CF)	$CF = \frac{\max(x)}{RMS}$	(12)
Skewness	$S_k = \frac{\sum_{k=1}^{n} (x_k - \bar{x})^3}{n\sigma^3}$	(13)
Kurtosis	$k = \frac{\sum_{k=1}^{n} (x_k - \bar{x})^4}{n\sigma^4}$	(14)

In order to exploit the potentiality of a separate novelty detection and fault identification stages, two different dimensionality reduction approaches are applied over the features sets.

For the novelty detection module, a PCA is used to extract a reduced set of features that maximize the variance of the dataset. Indeed, from the novelty detection point of view, the data variance represents one of the most convenient characteristics to be considered. Thus, most of the data variance is enhanced and preserved by a reduced set of features called principal components.

The fault identification task is classically approached by previous feature reduction techniques in order to maximize the distances among available labeled classes. Unlike PCA, that preserves as much data variance as possible in a reduced set of features, the classification task requires supervised approaches. Nevertheless, one of the challenges considered in this work is the initial availability of only one class, the healthy condition, therefore such supervised approaches are not viable. Thus, in order to deal with such scenario, all the twenty estimated features are considered.

After the corresponding feature reduction, the novelty detection and fault identification models are initially trained using healthy measurements of the friction tests. As the monitoring of the machine progresses and new faults are identified, the models are eventually re-trained with new classes of detected faults.

To analyze a single torque signal measurement, the corresponding reduced set of features is first examined by the novelty detection model. Then, the measurement can be cataloged as *novel* or *known*. If the measurement is catalogued as novel, the machine is considered to be working under unknown conditions, therefore an alarm is activated to the user for supervision. This can be triggered by different scenarios, including outliers, the presence of a new fault or by a new operation condition of the machine. If the user determinates that the unknown condition correspond to a new fault in the machine or a new operation point the re-training stage is activated, otherwise, the alarm is considered to be a false alarm and the measurements are discarded.

If the measurement is catalogued as *known*, it means that the machine is working under a previously known scenario, which can be healthy or faulty. To discern between the known scenarios, the measurement is analyzed by an evolving classifier. The output of the model is a label that identifies the analyzed measurement as one of the considered classes.

In this application, instead of analyzing each EOL test individually, a batch-type analysis is performed, where 20 measurements (which corresponds to 20 EOL tests) are stored and then evaluated simultaneously by the methodology. The number of analyzed measurements corresponds, in this case, to the number of EOL tests performed for each shaft of the steering system. The number of tests is empirically selected to provide a robust decision of the friction test; if only one test is performed then the rate of false alarms could be increased drastically due to outliers. The condition of the machine is determined according to the majority of the labels.

B. Novelty detection and fault identification model selection

An ensemble of one-class classifiers is used to perform the novelty detection task. An increasing amount of studies point out that domain based novelty detection models present promising results. Thus in this work the OC-SVM with Gaussian kernel is used as a one-class classifier for the ensemble method. The preparation of each OC-SVM includes the selection of the parameters for configuration and the training of the model.

The ensemble method consist on training one novelty model for each class known and combine their outputs to detect if an analyzed measurements are *known* or *novel*. Each OC-SVM is trained with information of one known scenario (can be healthy and faulty sets). This means that the model defines a novelty threshold that encloses all the *known* scenarios, if a new measurement evaluated has a novelty score lower than the threshold defined, then, it is considered *novel*, on the contrary, if the novelty score is above the threshold, it is considered *known*. Each measurement is evaluated by all the OC-SVM trained, and if at least one of the model labels the measurement as *known* then the final label for that measurement is *known*, consequentially if none of the models label a measurement as *known*, then the final label is *novel.*

Regarding the fault identification stage, the evolving classifiers eClass0 and eClass1 are used. These classifiers are able to adapt dynamically to the new data with no need of any specific threshold to be specified. The FRB structure of both classifiers changes according to the data streams. In addition, in the case of eClass1, the parameters of the regression models are also constantly updated. The prototypes (existing data samples) to create the fuzzy rules are selected via the calculation of the potential, which is a Cauchy function of the sum of the distances between a certain data sample and all other data samples in the feature space. It is very important to remark that since the formulation of the potential is calculated in a recursive manner, instead of using the complete dataset, the current measurement uyses only $(n + 1)$ memorized quantities, where n is the number of features [24]. This aspect is essential in online applications.

To evaluate a new measurement, a firing level (degree of confidence) of the fuzzy rules is calculated and the output of the rules determinate the class of the evaluated measurements. For eClass0 the output label is directly associated to the activated rule, meanwhile, for the eClass1 an *eTS* model regress the feature vector to determinate the confidence value.

C. Re-training Stage

The re-training stage is triggered when a novel scenario is detected and the user determinates that it correspond to a new fault or a new operating condition of the machine. The models used in the novelty detection stage and the fault identification stage have different re-training procedures to include a new class to their base knowledge.

Regarding the re-training of the novelty detection stage, a new OC-SVM model is trained including only information of the new class, this allows the monitoring system to include new information without needing access of the measurements initially used for training. Since the proposed approach for novelty detection considers a low number of features, three after the PCA, the model can perform an adequate novelty boundary for the new class with a limited number of samples and avoid the curse of dimensionality. The training of the model consists on selecting the configuration parameters and tuning them according the distribution of the data to select an appropriate novelty threshold.

Regarding the re-training of the evolving classifiers, both eClass0 and eClass1 have the capability of including automatically new measurements to their base knowledge to increase the robustness of classification of the known classes or to include new classes. In case a set of measurements with a new class is used for re-training, for both of the classifiers a new prototype is selected and a new rule is created for the new class; additionally, for eClass1, the parameters of the *eTS* models are updated. This is performed automatically as long as the true label of the measurements of the new class is provided.

V. Experimental Results

In order to validate the feasibility of the proposed methodology, four different models of steering systems have been tested. The four models possess the same structure described previously, but with different brands of components. It is important to note that all the models of steering systems used were in healthy state, in order to focus the analysis on the state of the test machine.

The expected torque response is slightly different for each steering system model. The four steering system models have a different reference pattern, therefore it is expected that the performance of the novelty detection and fault identification models are affected by the variability of the torque response of the models. Nevertheless, it is desired to assess the capability of a condition monitored approach to generalize between different models of steering systems and correctly identify the machine condition.

Eight classes regarding the condition of the machine are considered on this work:

- Healthy condition: *Hc.*
- Six faulty conditions: MIS_5, MIS_6, MIS_7, CW_1, CW_2, CW_3.
- Novelty condition: *Nc.*

For each one of the 4 models, 20 friction tests are performed, that leads to a total of 80 measurements for each class. Then, the complete dataset consist of a total of 640 measurements.

A. Performance analysis

Seven different scenarios for test are used to evaluate the capability of the methodology to detect and classify novel scenarios and the response of the models to the incorporation of new classes to the initially available information. The distribution of the classes for each scenario is presented in Table II.

The 8 classes are grouped in three sets: *training set*, *known set* and *novelty set*. Each of the scenarios correspond to a progressing stage of the continuous monitoring approach, from an initial knowledge of only the Healty condition, (*Hc*) to a scenario where information of 7 classes is available.

These scenarios are intended to test the capabilities of the proposed methodology in the industrial framework where initially the healthy condition is initially available, and progressively new classes are detected and incorporated, in this case, one class to the training stage in each iteration.

The dimensionality reduction for novelty detection, the PCA, is performed in all seven scenarios using only measurements of the healthy condition (*Hc*). Since some of the contributions in this work are focused on providing an alternative of classical approaches that need storing a historical database of measurements of the monitored machine, a selection of more appropriate features whenever a new class is incorporated is out of scope in this work.

The proposed methodology is based on a sequential monitoring scheme, which implies that each sample is analyzed by the novelty detection model and then, if is labeled as *known*, analyzed by the fault identification model. If the tests are performed using this sequential monitoring scheme, the results obtained can mislead the analysis of the performance of the models due to an integration of error. For example, if all the measurements are labeled incorrectly by the novelty detection stage, the performance of the fault identification stage is not analyzed. Therefore, to avoid misleading interpretation of results, the tests are performed separately to the models. Thus, to analyze the performance of each stage, two sets of performance metrics are considered: one for the novelty detection stage and the other for the fault identification stage.

Regarding the results obtained from the novelty detection model, the following metrics are calculated:

- Test Set Performance: This metric refers to the number of correctly classified measurements of the *novelty set* and the *known set* divided by the total of test examples. This metric is used to obtain a novelty model global performance. Nevertheless, it is important to notice that it does not contemplate the accuracy of discriminating between the different classes composing the *known class* in the fault identification stage.
- Known Set Performance: This metric refers to the number of correctly classified measurements of the *known set* divided by the total of measurements belonging to the same set. This metric can be seen as the true negative rate (TNR).
- Novelty Set Performance: This metric refers to the number of correctly classified measurements of the *novelty set* divided by the total of measurements belonging to the same set. This metric can be seen as the true positive rate (TPR).

Regarding the results obtained from the fault identification stage, the following metrics are calculated:

- Training set Performance: This metric represents the capacity of the fault identification model to classify the samples used in the training. A low training performance indicates that the model is not able to discriminate among classes, which can be caused by an overlapping of the data in the feature space.
- Test Set Performance: This metric represents the measurements analyzed of the *known set* by the fault identification model that are correctly classified divided by the total number of measurements of the *known set.* It is important to notice that the *novelty set* is not contemplated in the fault identification stage because it is assumed that these samples were previously discarded by the novelty detection model. Including such class in the comparison between fault identification models induce an unnecessary constant error in the tests.

TABLE II.
CONTENTS OF THE TRAINING AND TESTING SETS FOR EACH SCENARIO

Scenario Name	Training Set	Testing Set	
		Known Set	Novelty Set
T_1	*Hc*	*Hc*	*CW1, CW2, CW3, MIS5, MIS6, MIS7, Nc*
T_2	*Hc, CW1*	*Hc, CW1*	*CW2, CW3, MIS5, MIS6, MIS7, Nc*
T_3	*Hc, CW1, CW2*	*Hc, CW1, CW2*	*CW3, MIS5, MIS6, MIS7, Nc*
T_4	*Hc, CW1, CW2, CW3*	*Hc, CW1, CW2, CW3*	*MIS5, MIS6, MIS7, Nc*
T_5	*Hc, CW1, CW2, CW3, MIS5*	*Hc, CW1, CW2, CW3, MIS5*	*MIS6, MIS7, Nc*
T_6	*Hc, CW1, CW2, CW3, MIS5, MIS6*	*Hc, CW1, CW2, CW3, MIS5, MIS6*	*MIS7, Nc*
T_7	*Hc, CW1, CW2, CW3, MIS5, MIS6, MIS7*	*Hc, CW1, CW2, CW3, MIS5, MIS6, MIS7*	*Nc*

TABLE III.
PERFORMANCE OF THE PROPOSED NOVELTY DETECTION SCHEME

	Novelty Detection PCA - 3 Features						
	T1	T2	T3	T4	T5	T6	T7
Test Set Performance	0.959(±0.004)	0.955(±0.008)	0.894(±0.002)	0.890(±0.014)	0.876(±0.019)	0.836(±0.027)	0.810(±0.031)
Known Set Performance	0.969(±0.005)	0.962(±0.002)	0.917(±0.007)	0.916(±0.009)	0.928(±0.017)	0.903(±0.028)	0.840(±0.068)
Outlier Set Performance	0.842(±0.046)	0.842(±0.064)	0.767(±0.053)	0.804(±0.044)	0.772(±0.029)	0.763(±0.033)	0.765(±0.023)

TABLE IV.
PERFORMANCE OF THE NOVELTY DETECTION CONSIDERING 20 FEATURES CALCULATED

	Novelty Detection 20 Features						
	T1	T2	T3	T4	T5	T6	T7
Test Set Performance	0.971(±0.004)	0.949(±0.014)	0.908(±0.014)	0.834(±0.011)	0.813(± 0.01)	0.748(±0.006)	0.656(±0.006)
Known Set Performance	0.981(±0.003)	0.957(± 0.01)	0.920(±0.015)	0.815(±0.012)	0.793(±0.012)	0.691(±0.015)	0.395(±0.029)
Outlier Set Performance	0.850(±0.054)	0.863(±0.047)	0.861(±0.034)	0.896(±0.024)	0.855(±0.029)	0.811(±0.013)	0.780(±0.008)

TABLE V.
RESULTS OF THE T_7 SCENARIO FOR THE FAULT IDENTIFICATION STAGE. THE EVOLVING CLASSIFIERS ARE COMPARED TO A CLASSICAL APPROACH IN BOTH FEATURE SELECTION APPROACHES.

	Fault Identification					
	20 Features			PCA		
	eClass0	eClass1	NN	eClass0	eClass1	NN
Training Set Performance	0.872(±0.024)	0.736(±0.024)	0.942(±0.015)	0.586(±0.052)	0.364(±0.038)	0.636(±0.022)
Test Set Performance	0.839(±0.016)	0.726(±0.032)	0.874(±0.014)	0.547(±0.044)	0.369(±0.035)	0.601(±0.023)

B. Model estimation and parameter selection

A 70% of the available measurements per class are used for the *training set*. From the *training set*, a five-fold cross-validation is used in order to adjust each of the OC-SVMs parameters of the ensemble method. The kernel used is the *Gaussian* and the value of the width of the kernel is limited among the following set of discrete values: {1, 2, 3, 5, 10, 15}. Regarding the neural network, a configuration of one hidden layer with 10 neurons is used. The neurons are configured with a sigmoid activation function and the training procedure corresponds to a classical back propagation algorithm using all the training samples.

Once the models are trained and adjusted in each scenario, the test is performed using the remaining 30% of the measurements of each class of the *known set* and the *novelty set*. This process was repeated five times with five different training-test set distributions, randomly selected and fixed.

C. Results and discussions

In order to highlight the contribution and motivation of this work, the outline of the results will be presented as follows: first, the seven scenarios are tested by the novelty detection model proposed, then, the seven scenarios are tested again but using the 20 calculated features instead of the proposed PCA, after the novelty detection is tested, the proposed method for fault identification is tested and compared with a classical approach. The performance metrics are analyzed on each case to highlight the advantages and disadvantage of each model and each dimensionality reduction approach.

Different configurations regarding the dimensionality of the features are used to have an insight of the advantages of discarding irrelevant features. The number of selected features is reduced from an initial 20-dimensional space to a reduced 3-

dimensional space, taking into consideration that the reduced set of features fulfill the respective restrictions from each dimensionality reduction approach.

The ensemble of novelty models are first tested with the seven scenarios proposed, the number of OC-SVM models trained correspond to the number of classes included on the training set of each scenario. It is important to mention that the scenarios tested are performed in consecutive order, therefore, the OC-SVM model trained with the Healthy class, *Hc*, is used on all the scenarios and not modified in a re-training. A re-training is performed when a new class is included, for example, from the scenario T_1 to the scenario T_2 to include a new class in a new OC-SVM, in this case the fault CW1.

For a better understanding of the test, the novelty score of the testing set of the scenario T1 is presented in Fig. 6. The horizontal line represents the novelty threshold, the samples on the upper side of the novelty threshold are labeled as *known* and the samples on the lower side are labeled as *novel*. Varying the novelty threshold, for example lowering the threshold value, can lead to a better detection of the *known class*, *Hc*, increasing the known set performance, but that variation also leads to labeling samples of the novelty class, *Nc*, as known, which lead to a lower outlier set performance. The vertical lines represent the division of classes among the measurements used of the testing set.

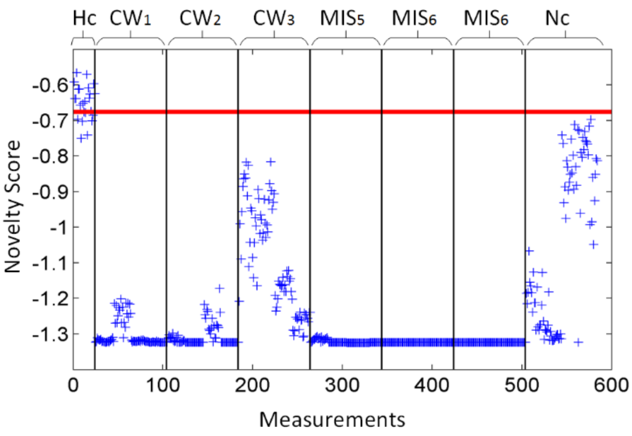

Fig. 6. Resulting Novelty Score of the T1. The red line is the novelty threshold, *Th*, which is set to -0.67. The black lines represent the division among the different classes during the test.

In this scenario only one class is included in the training set, therefore only one novelty score per measurement is obtained in the novelty detection stage, which implies that the labeling of measurements depends only of one model. To perform the test of the scenario T_2, where two classes are included in the training set, a new OC-SVM is trained for the *CW1* class and two novelty scores are obtained. The novelty scores of both OC-SVM is shown in Fig. 7.

As can be seen, all measurements from the test set are evaluated by the ensemble of models, therefore, all samples are labeled as *novel* or *known* by both OC-SVMs. As mentioned before, if at least one of the models label a measurement as *known*, then the final label of that measurement is *known*, consequentially if none of the models label a measurement as *known*, then the final label is *novel*. Since each of the OC-SVMs is trained to identify similar measurements of different classes, both output labels are equally important, therefore the only case when a sample is considered novel is when it does not belong to any distribution learned by the method. In this case, in a continuous monitoring approach, if a measurement is labeled as known the fault identification stage analyze such measurement to identify if the measurement belongs to a healthy case, *Hc*, or the fault case, *CW1*.

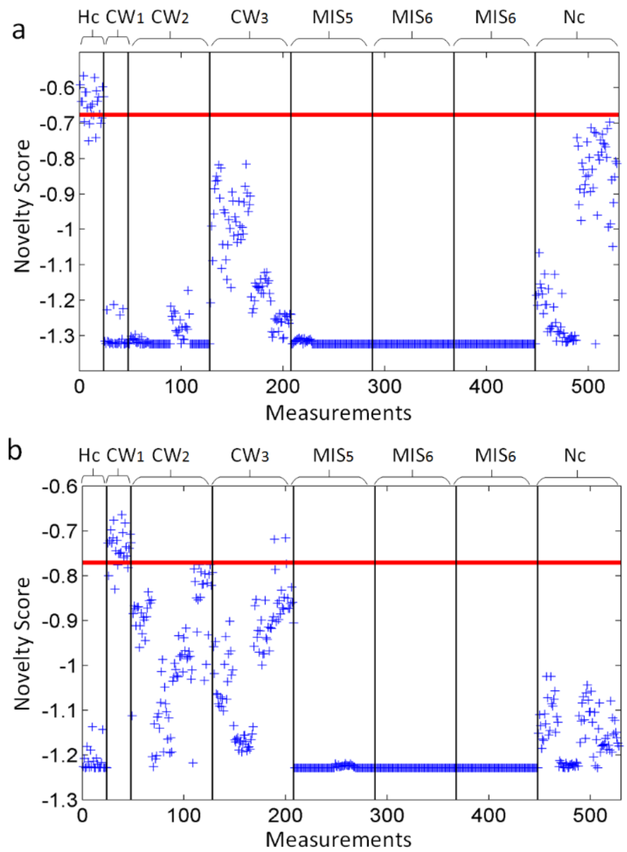

Fig. 7. Resulting Novelty Score of the T_2. The red line is the novelty threshold, *Th*, which is set to -0.67 for the first OC-SVM and -0.77 for second. The black lines represent the division among the different classes during the test. a) The novelty results of the OC-SVM trained using measurements of the *Hc* class. b) The novelty results of the OC-SVM trained using measurements of the CW_1 class.

Some authors propose to use the output of the ensemble of the OC-SVMs to perform the fault identification of the measurement; nevertheless, while in this scenario, T_2, such approach seems like a viable option, problems appear when classes with similar distributions are included. A solution for this problem is addressed in the methodology proposed in this work and discussed further ahead in this section.

The results of the proposed novelty detection approach in the seven scenarios are shown in Table III.

A high test set performance, around 96%, is obtained on the T_1 and T_2 scenarios, nevertheless, the performance gradually decrease to a final 81% as new classes are incorporated to the base knowledge. This decrease of performance is expected, in each scenario new information is incorporated to the model, therefore more variability and cases are considered normal to the model and this limits the capacity of the model to detect anomalies.

To verify if the dimensionality reduction stage improves the performance of the model, the seven scenarios are also tested using the 20 calculated features, the results are shown in Table IV.

As can be seen that, regarding the T_1 to T_3 scenarios, a performance around 96% is obtained in both approaches, nevertheless starting from the T_4 scenario, the performance of the model using the 20 features start decreasing at a higher rate in comparison to the PCA proposed approach. The decrease of performance is caused by the misclassification of the known set, labeling them as *novel*, which implies that reducing the features allows a more adequate tuning of the novelty detection boundary to achieve a more robust detection of the known classes.

To test the fault identification stage a separate analysis is performed where only the scenario T_7 is used. Testing the fault identification method in a scenario where the class discrimination is evident would lead to an excellent performance of every model and does not contribute in highlighting the limitations and advantages of each model and features used for comparison. Therefore, the scenario which presents a more adequate challenge regarding number of classes for discrimination is used. The result of the fault identification stage using evolving classifiers and a supervised classical approach used in many condition monitoring systems, the multi-layer Neural Network (NN), is shown in Table V. Regarding the Neural Network, a configuration of one hidden layer with 10 neurons is used. The neurons are configured with a sigmoid activation function and the training procedure corresponds to a classical back propagation algorithm.

A comparison between using the 20 calculated features and a dimensionality reduction by PCA is also performed. Better results are obtained with all three models using the 20 features than the PCA approach. Meanwhile the PCA represents a better option to detect outliers in the novelty detection stage, for the fault identification stage the use of 20 features represent a better option to discern between classes. The misclassification is caused by an overlapping of classes in the feature space, this can be highlighted by the training set performance and the consistent low performance in all three models. If the novelty detection task and the fault identification task are performed in the same stage, the selection of a different feature selection for each stage would lower the performance of the methodology.

Regarding the test set performance of the three models using the 20 calculated features as input, the Neural Network approach present a slightly better performance, around 3%, than the evolving classifier eClass0 and a 15% increased performance than the eClass1, nevertheless, the advantage of the evolving methods in terms of incursion of different classes and low computational cost of training and testing, represent a better option for the fault identification stage for online applications.

VI. Conclusions

This paper proposes a methodology for continuous learning of condition monitoring applied to an End-of-Line test machine of the automotive sector by analyzing the torque signal. The methodology assess the condition of the machine under monitoring without altering the undergoing operation.

Taking into consideration that the methodology presented work under the assumption that only the healthy condition is initially available, the main contributions presented are focused on: (i) the analysis and selection of a separate feature calculation and reduction method to increase the performance of the novelty detection and fault classification stages and (ii) the proposition of an ensemble of novelty detection models and an evolving classifier to perform incremental learning when a database storing the measurements of the machine is not available.

The methodology proposed is compared in each stage with different feature reduction approaches to perform a proper selection of features for each model taking in consideration the challenges of an industrial framework. A set of several tests with the corresponding performance metrics are proposed for the evaluation of both stages, novelty detection and fault identification. By monitoring separately the accuracy of the models, it is possible to identify the advantages and limitations of the methodology and compare different models for each stage.

Regarding the novelty detection stage, the ensemble method of OC-SVMs successfully discerned among the known set and the outlier set of the seven performed tests. Using the PCA to select a reduced number of features in comparison of using the 20 features lead to a better performance, especially at the last test, where an increment of 44% of accuracy on the known set is achieved.

Regarding the fault identification stage, among the evolving classifiers, the eClass0 obtained better results than the eClass1, where an increment of 11% of accuracy on the test set is achieved using the 20 features and an increment of 17% on the test set is achieved using the PCA. Since each class is composed by four different models of steering systems, four different distributions are expected among the data; therefore the eClass0 is more suitable in this case to enclosure a multi-modal distribution of the data with the Gaussian enclosure of the prototypes than the regression performed by the eClass1. Regarding the Neural Network classifier comparison, the eClass0 obtained a slightly lower accuracy; nevertheless Neural Network the intensive offline training of the Neural Network is not suitable for online applications where a database of historical measurements is not available.

In general, the results obtained in this work suggest that this methodology may be also useful for any other industrial machines, with a corresponding signal processing stage to identify a suitable set of features of the monitored machine.

Acknowledgment

The authors wish to acknowledge financial support from the Generalitat de Catalunya (GRC MCIA, Grant nº SGR 2014-101). This work is also supported in part by the Spanish Ministry of Economy and Competitiveness under the TRA2016-80472-R Research Project. 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. This research was partially supported by CONACyT scholarship 313604 grant.

References

[1] H. Henao, 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 Ind. Electron. Mag.*, vol. 8, no. 2, pp. 31–42, 2014.

[2] 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.

[3] A. Lemos, W. Caminhas, and F. Gomide, “Adaptive fault detection and diagnosis using an evolving fuzzy classifier,” *Inf. Sci. (Ny).*, vol. 220, pp. 64–85, 2013.

[4] Y. Hu, P. Baraldi, F. Di Maio, and E. Zio, “A Systematic Semi-Supervised Self-adaptable Fault Diagnostics approach in an evolving environment,” *Mech. Syst. Signal Process.*, vol. 88, no. October, pp. 0–1, 2016.

[5] R.-E. Precup, P. Angelov, B. S. J. Costa, and M. Sayed-Mouchaweh, “An overview on fault diagnosis and nature-inspired optimal control of industrial process applications,” *Comput. Ind.*, vol. 74, pp. 75–94, 2015.

[6] C. G. Bezerra, B. S. J. Costa, L. A. Guedes, and P. P. Angelov, “An evolving approach to unsupervised and Real-Time fault detection in industrial processes,” *Expert Syst. Appl.*, vol. 63, pp. 134–144, 2016.

[7] 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.

[8] B. Krawczyk and M. Woźniak, “Reacting to different types of concept drift with adaptive and incremental one-class classifiers,” in *2015 IEEE 2nd International Conference on Cybernetics (CYBCONF)*, 2015, pp. 30–35.

[9] B. Krawczyk and M. Woźniak, “One-class classifiers with incremental learning and forgetting for data streams with concept drift,” *Soft Comput.*, vol. 19, no. 12, pp. 3387–3400, 2015.

[10] H. Al-Behadili, A. Grumpe, and C. Wöhler, “Incremental Learning and Novelty Detection of Gestures in a Multi-class System,” in *2015 3rd International Conference on Artificial Intelligence, Modelling and Simulation (AIMS)*, 2015, pp. 304–309.

[11] L. Duan, M. Xie, T. Bai, and J. Wang, “A new support vector data description method for machinery fault diagnosis with unbalanced datasets,” *Expert Syst. Appl.*, vol. 64, pp. 239–246, 2016.

[12] P. Baraldi, R. Razavi-Far, and E. Zio, “Classifier-ensemble incremental-learning procedure for nuclear transient identification at different operational conditions,” *Reliab. Eng. Syst. Saf.*, vol. 96, no. 4, pp. 480–488, Apr. 2011.

[13] M. D. Muhlbaier, A. Topalis, and R. Polikar, “Learn++.NC: Combining Ensemble of Classifiers With Dynamically Weighted Consult-and-Vote for Efficient Incremental Learning of New Classes,” *IEEE Trans. Neural Networks*, vol. 20, no. 1, pp. 152–168, 2009.

[14] M. Chen, P. Chang, and J. Wu, “Engineering Applications of Arti fi cial Intelligence A population-based incremental learning approach with arti fi cial immune system for network intrusion detection,” *Eng. Appl. Artif. Intell.*, vol. 51, pp. 1–11, Jan. 2016.

[15] D. Sannen, E. Lughofer, and H. V Brussel, “Increasing On-line Classification Performance Using Incremental Classifier Fusion,” in *2009 International Conference on Adaptive and Intelligent Systems*, 2009, pp. 101–107.

[16] M. Dave, S. Tapiawala, M. J. Er, and R. Venkatesan, “A novel progressive multi-label classifier for class-incremental data,” in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2016, pp. 3589–3593.

[17] M. Karasuyama and I. Takeuchi, “Multiple Incremental Decremental Learning of Support Vector Machines,” *IEEE Trans. Neural Networks*, vol. 21, no. 7, pp. 1048–1059, 2010.

[18] Z. Erdem, R. Polikar, F. Gurgen, and N. Yumusak, “Ensemble of SVMs for Incremental Learning,” in *Multiple Classifier Systems: 6th International Workshop, MCS 2005, Seaside, CA, USA, June 13-15, 2005. Proceedings*, N. C. Oza, R. Polikar, J. Kittler, and F. Roli, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 246–256.

[19] P. M. Ciarelli, E. Oliveira, and E. O. T. Salles, “An incremental neural network with a reduced architecture,” *Neural Networks*, vol. 35, pp. 70–81, Nov. 2012.

[20] D. Skočaj, M. Uray, A. Leonardis, and H. Bischof, “Why to Combine Reconstructive and Discriminative Information for Incremental Subspace Learning,” in *Computer Vision Winter Workshop 2006*, 2006, pp. 1–6.

[21] A. S. Al-Hegami, “Classical and Incremental Classification in Data Mining Process,” *Int. J. Comput. Sci. Netw. Secur.*, vol. 7, pp. 179–187, 2008.

[22] Y. Freund and R. E. Schapire, “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting,” *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997.

[23] E. Bauer and R. Kohavi, “An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants,” *Mach. Learn.*, vol. 36, no. 1, pp. 105–139, 1999.

[24] P. P. Angelov and X. Zhou, “Evolving Fuzzy-Rule-Based Classifiers From Data Streams,” *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 6, pp. 1462–1475, 2008.

[25] R. D. Baruah, P. Angelov, and J. Andreu, “Simpl_eClass: Simplified potential-free evolving fuzzy rule-based classifiers,” in *2011 IEEE International Conference on Systems, Man, and Cybernetics*, 2011, pp. 2249–2254.

[26] 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 Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3398–3407, 2013.