Machine Learning Strategy for Fault Classification Using Only Nominal Data

Gianluca Nicchiotti[1], Luc Fromaigeat[2], and Lianel Etienne[3]

[1,2]*Meggitt Sensing Systems, CH-1701 Fribourg, Switzerland*
gianluca.nicchiotti@ch.meggitt.com
luc.fromaigeat@ch.meggitt.com

[3]*Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland*
lianelpierre.etienne@epfl.ch

ABSTRACT

Machine learning methods are increasingly used for rotating machinery monitoring. Usually at set up, only data associated to an engine in a good state, the so called nominal data, are available for the machine learning phase. Nevertheless a classifier requires faulty data to be trained at identifying the causes of the anomalies and this fact has generally limited the usage of data driven approaches to fault detection tasks. The paper suggests a strategy to use machine learning methods even for fault classification purposes and diagnostics. Within the proposed framework three different machine learning methods, Gaussian Mixture Model (GMM), Support Vector Machines (SVM) and Auto Associative Neural Networks (AANN) have been implemented, tested and compared. The idea is to take into account some 'a priori' knowledge about the faults to be classified, to drive the behavior of the machine learning methodology (SVM or AANN or GMM) to be more or less reactive to the different faults. The indicators (features) more sensitive to each kind of fault are firstly selected on the basis of expert knowledge. For each different fault, a set of indicators is defined and computed from nominal data only. Each set is then used to produce training data for one specific fault. Such data sets are then used to train one instance of each method for each different fault. The underlying logic is that fault tuned input data is able to produce fault tuned instances of the methods. For example the instance trained with the indicators associated to a fault 'A' reacts more powerfully in presence of the fault 'A' than the others. Once an anomaly is detected, the comparison among the reactions of the different 'fault tuned' instances allows classifying the fault, not just to detect it. The results show best detection performances for SVM whilst AANN outperforms the other two methods for classification.

1. INTRODUCTION

Data driven approaches are methodologies which are progressively more employed for anomaly and fault detection for machine condition monitoring purposes. However high integrity systems could not always use the traditional learning\classification method for a number of reasons: abnormalities are very rare or there are no data that describes the fault conditions. One of the main limitations lies indeed in the fact that, at engine setup, only valid (nominal) data are available for training. Novelty detection offered a solution to this problem by modeling normal data and using a distance measure and a threshold for determining abnormality. However in this framework, learning algorithms can be taught only the nominal behavior of the system and they cannot be used for fault classification. (Samanta, Al-Balushi, & Al-Araimi, 2003), (Jack & Nandi, 2002), (Booth & McDonald, 1998), (Sanz, Perera, & Huerta, 2007), (Guttormsson, Marks, El-Sharkawi, & Kerszenbaum, 1999), (Rojas & Nandi, 2006) (Prego, et al., 2013) (Alguindigue & Uhrig, 1991), (Fulufhelo, Tshilidzi, & Unathi, 2005), (Rubio & Jáuregui, 2011) developed methods to detect anomalous behaviors (anomaly detection) using Neural Networks (NN), Support Vector Machines (SVM) and Gaussian Mixture Models (GMM).

In the following we will propose an idea to extend machine learning capabilities from fault detection to fault classification with the constraint that only nominal data are available for training. The logic is to use a priori knowledge about the effects of each fault to be classified in order to produce training data which are somehow fault tuned. These training data are generated by computing, on nominal

data, features which are known to be the most responsive to each kind of fault which has to be classified.

For instance, in the case of neural networks, for each fault one NN is trained with its own fault tuned training data set. When a new sample to be analyzed arrives it will be processed by each one of the networks. The results of each network are then treated with a simple logic to determine if there is a fault and which fault it is, respectively fault detection and fault classification The paper is organized as follows: An overview of data driven approaches for anomaly detection will be presented in section 2. The focus will be on Auto Associative Neural Networks (AANN), Unsupervised Support Vector Machines and Gaussian Mixture Models.

The proposed strategy to extend the data driven capabilities from detection to classification will be described in section 3. Section 4 will illustrate the test rig. The results obtained with the three data driven methodologies will be then compared and discussed.

2. MACHINE LEARNING METHODS

According to Johannes (2001) we can identify three major classes among the machine learning techniques: the density methods like GMM, the boundary methods as SVM and the reconstruction methods AANN. In recent times, all these possible data driven approaches have been employed for machine condition monitoring purposes (Chandola, Banerjee, & Kumar, 2009). In (Samanta, Al-Balushi, & Al-Araimi, 2003), (Jack & Nandi, 2002) and (Booth & McDonald, 1998), ANN and SVM have been employed to diagnose bearing faults and faults in power transformers. In (Alguindigue & Uhrig, 1991) ANN, (Fulufhelo, Tshilidzi, & Unathi, 2005) GMM, (Rojas & Nandi, 2006) SVM and (Prego, et al., 2013) ANN multiclass fault diagnostics was achieved using fault seeded data during training phase. All these approaches require the availability of fault data during the learning phase.

However, when new machinery is set up, fault data are not available and such diagnostic approaches are not viable. For this reason, in an operational context, data driven methods are more frequently used for fault detection tasks. Detection methods used in (Rubio & Jáuregui, 2011), (Guttormsson, Marks, El-Sharkawi, & Kerszenbaum, 1999) and (Sanz, Perera, & Huerta, 2007) are also called one-class classification. In one-class classification, it is assumed that only information of one target class is available and can be used for training. These methods using Auto-Associative Neural Networks (Sanz, Perera, & Huerta, 2007) or boundaries (Guttormsson, Marks, El-Sharkawi, & Kerszenbaum, 1999) only need nominal/healthy data for training.

One-class classification methods have been used so far only for fault detection tasks. In the following we will propose a strategy for their use for fault classification. The endeavor of this approach is to include expert ('a priori') knowledge in these data driven ('a posteriori') methods.

In the proposed classification strategy we have integrated one data driven method from each of the three previously identified classes: GMM as a density method, SVM as a boundary method and AANN as a reconstruction method.

The following subsections give a theoretical overview of the three data driven methods which have been integrated in our classification framework.

2.1. Gaussian Mixture Models

The normal distribution is a widely used model for the distribution of continuous variables (Bishop, 2006).

For a *D*-dimensional vector x of variables, the multivariate Normal distribution can be written in the form

$$N(\mathbf{x}|\boldsymbol{\mu},\Sigma) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}}{(2\pi)^{D/2}|\Sigma|^{1/2}} \qquad (1)$$

Where $\boldsymbol{\mu}$ is a *D*-dimensional mean vector, Σ is a $D \times D$ covariance matrix, and $|\Sigma|$ denotes the determinant of Σ.

The Normal distribution makes very strong assumptions about the model of the data. It should be unimodal and convex (Johannes, 2001). To obtain a more flexible density method, the normal distribution can be extended to a mixture of Gaussians (MoG). It is a linear combination of normal distributions:

$$p_{MoG}(\mathbf{x}) = \frac{1}{N_{MoG}}\sum_{j}\alpha_j\, N(\mathbf{x}|\boldsymbol{\mu_j},\Sigma_j) \qquad (2)$$

where α_j are the mixing coefficients and $p_{MoG}(\mathbf{x})$ is the Gaussian mixture distribution. If the number of mixture N_{MoG} is defined beforehand, the means $\boldsymbol{\mu_j}$ and covariances Σ_j of the individual Gaussian components can efficiently be estimated by the Expectation-Maximization (EM) algorithm (Bishop, 2006) and (Bilmes, 1998).

Once the model is estimated, it is possible to associate to each new measure a probability of belonging to the distribution.

2.2. One Class SVM

The SVM algorithm is usually constructed as a two-class algorithm which needs negative and positive examples for training (Ng, 2015). Schölkopf in (Schölkopf, Platt, Shawe-Taylor, & Smola, 2001) proposed a modification to allow its

use for only positive examples. This is called the one-class learning or unsupervised SVM. Basically, the algorithm returns a function f that takes the value +1 in a "small" region capturing most of the data points and -1 elsewhere. The strategy is to map the data into a feature space using an appropriate kernel function, and then try to separate the mapped vectors from the origin with maximum margin.
Let $x_1, x_1, \dots, x_l$ be training examples belonging to one class X, where $l \in \mathbb{N}$ is the number of observations and X is a compact subset of R^N. Let $\Phi : X \longrightarrow H$ be a kernel function which maps the training examples from X to the feature space H.Then, to separate the mapped vectors from the origin, the following quadratic problem has to be solved:

$$\min \frac{1}{2}\|w\|^2 + \frac{1}{\nu l}\sum_{i=1}^{l}\xi_i - \rho \qquad (3)$$

subject to $\left(w \cdot \Phi(x_i)\right) \geq \rho - \xi_i \quad i = 1,2,\dots,l \quad \xi_i \geq 0$, where ρ is the offset and w the weight vector. The weight vector and the offset determine the hyperplane in the feature space associated with the kernel. In Equation (3), $\nu \in (0,1]$ sets an upper bound on the fraction of outliers (training examples regarded out-of-class) and a lower bound on the number of training examples used as support vectors.

Since nonzero slack variables ξ_i penalize the objective function, we can expect that if w and ρ solve this problem, then the decision function

$$f(x) = sign\left((w \cdot \Phi(\mathrm{x})) - \rho\right) \qquad (4)$$

will be positive for most examples x_i contained in the training set.

2.3. AANN

Auto-Associative Neural Networks (AANN), also known as Replicator Neural Networks or Autoencoders, are a family of ANN which are trained to reproduce their input at the output (Kramer, 1992).

An ANN can be viewed as a system of interconnected processing elements called "neurons" exchanging messages between each other. Each neuron possesses numeric parameters (weights), which are set by means of a training process. In AANN, during training, the network learns to duplicate the input at the output, whose sizes are therefore the same as it can be observed in Figure 1. At first sight, this replication task could seem trivial; however, the network structure has a "bottleneck" as the hidden layer has fewer nodes than the input and output layers. This means that within the hidden layer(s) a compression process of the input data takes place. This forces the network to learn the significant features of the input data.

Once trained with healthy data, the AANN will able to replicate unseen nominal data with good accuracy. However faulty data are expected to possess information content which is structured differently from the healthy ones; as a consequence the compression step in the hidden layer cannot be performed efficiently and the reconstruction result will be inaccurate. Once a new sample is processed by the AANN, the measure of the difference between output and input, the Reconstruction Error (RE) of an input vector X, is computed as

$$RE(X) = \left\| X\text{-}Out_X \right\| \qquad (5)$$

where Out_X is the output of the AANN and || symbol stands for any p- norm.

The RE measures how much the new sample belongs to the same class of data used for training i.e. the healthy class. Once computed the RE, a fault or anomaly detection logic can be easily implemented for instance by thresholding.

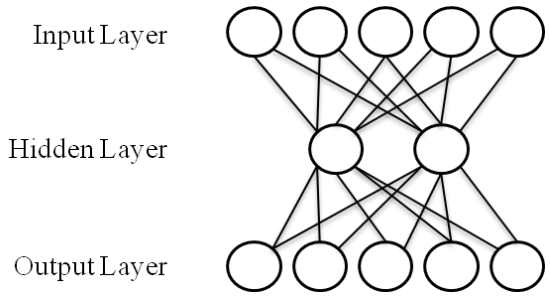

Figure 1. A simple Auto-Associative Neural Network

3. DIAGNOSTICS STRATEGY

As highlighted in the previous section, data driven methods have been proven to be successful in accomplishing anomaly detection task when trained with examples belonging to the healthy class. Our strategy proposes to combine several anomaly detection subtasks to perform a multi class fault classification.

In the following, feature is meant an individual measurable property of the phenomenon being observed. In this study, the features are the different characteristics of the signals that can be extracted by the monitoring system.

Let $F = \{f_1, f_2, \dots, f_\mathrm{N}\}$ be the set of the features generated by the monitoring system. For example, f_1 could be the amplitude of the vibration signal.

The idea is to exploit the fact that certain features are more responsive to certain faults than others. Let L be the number of faults which have to be discriminated and F_i a subset of F

containing the features most reactive to fault *i.* For example, $F_i = \{f_j, f_k\}$ means that features *j* and *k* have been selected to train the machine learning module for fault *i.* One machine learning module M_i is then trained for each fault *i.* *M* can be either AANN, SVM or GMM method and Figure 2 represents the above described definitions.

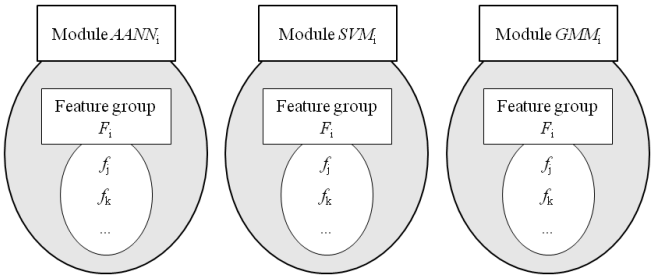

Figure 2. Definition of the different modules for a fault *i*

It is important to clarify that at this stage of the research each machine learning method is processed independently from the others.

To illustrate the process of fault detection and classification, the procedure for the AANN case is described (see also Figure 3). Consider the test sample *X;* its raw data are initially processed to produce the fault tuned feature subsets $F_i{}^X$ which are then analyzed by the corresponding AANN modules (one network per fault), to create a set of measures $D_i{}^X$ (*i=1,...L).*

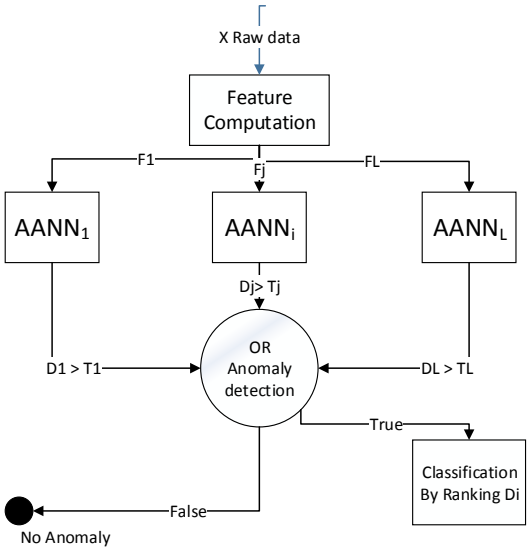

Figure 3. Block diagram for detection and classification process

If any of the values $D_i{}^X$ exceeds its corresponding threshold T_i an anomaly A^X is declared. If an anomaly is present (A^X is true), then the values $D_i{}^X$ are ranked in descending order. If

$$D_K = \max(D_i^X) \quad i = 1, \dots L \tag{6}$$

the algorithm concludes that K is the most likely fault which occurred.

The following subsections detail for each method how the thresholds and distances are computed and ranking is performed.

3.1. Detection and classification with AANN

Once the AANN module have been trained, the training samples are fed again into the networks to obtain mean $\mu_{REF}^{(i)}$ and standard deviation $\sigma_{REF}^{(i)}$ of the RE distribution.

The threshold of module $AANN_i$ of fault *i* is then defined as:

$$T_i = \mu_{REF}^{(i)} + n \cdot \sigma_{REF}^{(i)} \tag{7}$$

Here $n > 0$ is a parameter which enables to adjust the sensitivity of the detection model. Under the assumption that the distribution of the training RE values is Gaussian, with $n = 3$, 99.7% of the nominal RE values are smaller than *T.* For $AANN_i$, the distance of sample $\boldsymbol{x}$ is defined as:

$$D_i(\boldsymbol{x}) = \frac{RE_x - \mu_{REF}^{(i)}}{\sigma_{REF}^{(i)}} \tag{8}$$

where $\mu_{REF}^{(i)}$ and $\sigma_{REF}^{(i)}$ are the RE references of one $AANN_i$ associated to fault *i* and $RE_{\boldsymbol{x}}$ is the reconstruction error of the sample under analysis.

As described, if $D_i(\boldsymbol{x})$ exceeds the threshold for at least one of the modules $AANN_i$, an anomaly is detected.
In such case $D_i(\boldsymbol{x})$ are sorted in descending order:

$$D_j(\boldsymbol{x}) > \cdots > D_k(\boldsymbol{x}) \tag{9}$$

The fault with the greatest distance is selected as the most likely cause of the anomaly.

3.2. Detection and classification with GMM

Once the parameters of the Gaussian mixture have been calculated with the EM algorithm, for each training sample $\boldsymbol{X}$ the smallest Mahalanobis distance $D_{min}^{(i)}$ *(X)* from an element of the mixture is computed. The threshold T_i is then set as:

$$T_i = n \cdot \max\left(D_{min}^{(i)}(X)\right), X \in \{Training\ set\} \quad (10)$$

Where max indicates the maximum value and again $n > 0$ is a parameter which enables to adjust the sensitivity of the detection model.
When a new sample $\boldsymbol{x}$ is analyzed, the Mahalanobis distance is computed by each GMM_i module:

$$D_i(\boldsymbol{x}) = \sqrt{(\boldsymbol{x} - \boldsymbol{\mu}_i)^{\mathrm{T}} S_i^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_i)} \quad (11)$$

These distances are then compared to the thresholds T_i. If for at least one module $D_i(\boldsymbol{x}) > T_i$, an anomaly is detected. When the anomaly is detected, the classification is done by ranking the $D_i(\boldsymbol{x})$ for all i, similarly to the AANN classification:

$$D_j(\boldsymbol{x}) > \cdots > D_k(\boldsymbol{x}) \quad (12)$$

As for AANN, the fault with greatest distance represents the most likely cause of the anomaly.

3.3. Detection and classification with one class SVM

In each SVM_i fault tuned module, the thresholds and the distances are computed by the method itself. The distance that the algorithm outputs is the argument of the "sign()" function of equation (4) :

$$D_i(\boldsymbol{x}) = w \cdot \Phi(\boldsymbol{x}) - \rho \quad (13)$$

with w, $\Phi()$ and ρ as defined in section 2.2 and i is the fault id .

$D_i(\mathrm{x})$ is positive for samples that are inside the boundary, zero at the boundary and negative for samples that are outside the boundary. Positive distances are considered as normal and an anomaly i is detected for negative distances $D_i(\boldsymbol{x})$ hence $T_i = 0\ \forall i$ .When at least one SVM module detects an anomaly, classification is performed and the distances are sorted, in this case, in ascending order:

$$D_j(\boldsymbol{x}) < \cdots < D_k(\boldsymbol{x}) \quad (14)$$

In this case the fault with most negative distance is classified as the cause of the anomaly.

4. EXPERIMENTAL SETUP

During experiments, data have been generated by a Rotor Kit STI (see Figure 4) and acquired by the VM600 platform and VibroSight™ software. Seventeen sensors can be located at different places on the Rotor Kit (proximity sensors, accelerometers and velocity sensors). The right-hand side motor enables the rotation of two disks at different regimes.

There are different configuration possibilities on the Rotor Kit which make possible data fault generation. Small weights of different size can be added on the disks to simulate unbalance conditions and misalignment faults can be generated at the shaft junction.

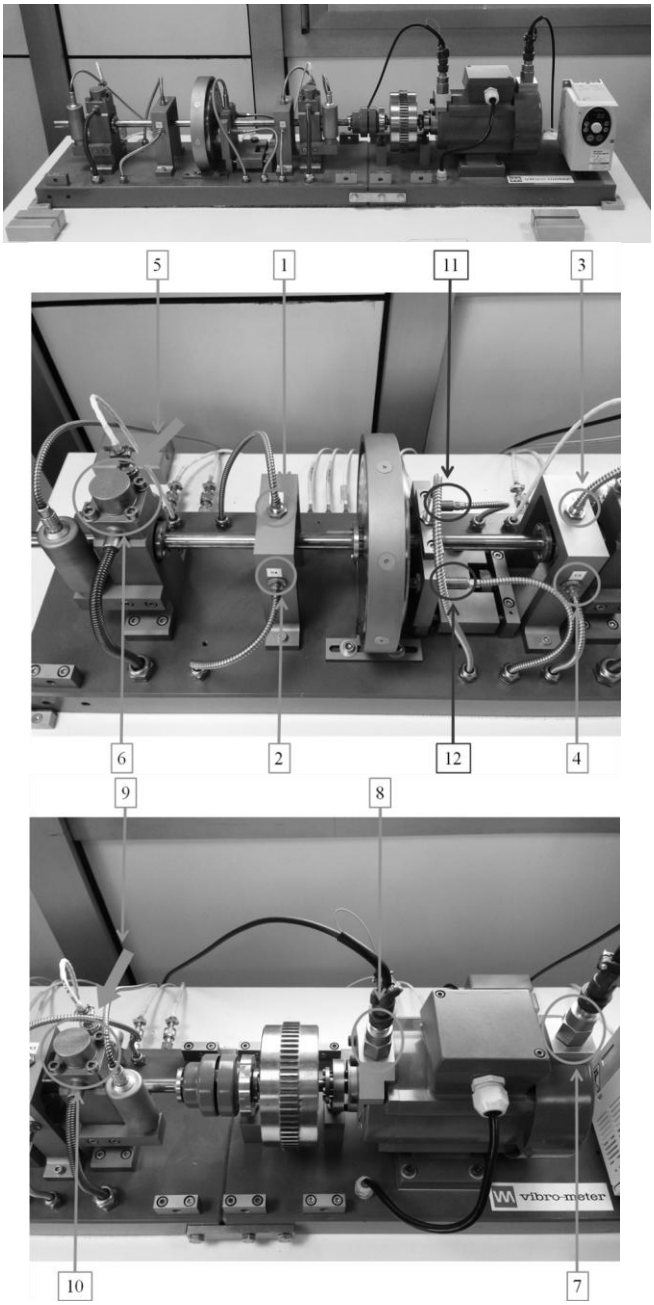

Figure 4. Experimental test rig (top); Sensor emplacements (bottom)

With the different setup options described, five faulty configurations have been be generated: unbalance on disk 1

(UnbalD1), unbalance on disk 2 (UnbalD2) misalignment (Misalignment), misalignment and unbalance on disk 1 and misalignment and unbalance on disk 2. Each fault has a major impact only on the signals recorded by a subset of sensors. As a consequence a subset of sensors has been associated to the detection of each fault as shown in Table 1.

Table 1. Sensors associated to the different faults ;ID column identifies the position in figure 4

Sensor Type	ID	Fault
Horizontal proximity sensor	1	UnbalD1
Vertical proximity sensor	2	UnbalD1
Horizontal proximity sensor	3	UnbalD1
Vertical proximity sensor	4	UnbalD1
Horizontal accelerometer	5	UnbalD1
Vertical accelerometer	6	UnbalD1
Vertical accelerometer	7	UnbalD2
Vertical accelerometer	8	UnbalD2
Horizontal accelerometer	9	UnbalD2
Vertical accelerometer	10	UnbalD2
Axial proximity sensor	11	Misalignment
Axial proximity sensor	12	Misalignment

5. RESULTS

Tests have been conducted at two different speed regimes: R1 (regime one) @3300 RPM and R2 (regime two) @2100 RPM. For each regime 10000 training samples have been recorded over several days to capture different room temperatures, and bearing warm-up conditions. Small weights have been added to one or both disks to increase the dispersion of the nominal data.

For each regime six test sets have been acquired in a different configuration of the rotor kit (see Table 2). They represent nominal data and five combinations of 3 types of faults: UnbalanceD1, UnbalanceD2 and Misalignment. For Misalignment fault, four features have been selected, twelve for UnbalD1 and eight for UnbalD2.

Each AANN module has been designed with a single hidden layer, and the number of mixtures in GMM is provided in the table 3. Several trials with different set of parameters (number of neurons in the hidden layer, number of mixtures) have been performed in order to identify the optimal configuration for each algorithm for which results will be reported in the following. The features used as input are computed from harmonic analysis of the vibration.

Table 2. Data sets for testing phase

Configuration name	Description	Set size (R1)	Set size (R2)
Nominal	No weight and no misalignment	400	300
UnbalanceD1	Weight on disk 1	400	400
UnbalanceD2	Weight on disk 2	400	400
Misalignment	Misalignment	100	100
Misal&unbalD1	Misalignment and weight on disk 1	200	200
Misal&unbalD2	Misalignment and weight on disk 2	200	200

As metrics for detection, the percentages of false alarms (FA) and missed detections (MD) have been estimated. The accuracy is determined as the percentage of tests where the fault detection process is performed correctly divided by the size of the test set. Table 4 summarizes the detection results for the 3 methods.

Table 3. GMM modules description

GMM modules	Number of mixtures
Misalignment	2
UnbalD1	5
UnbalD2	2

For detection, the method with the best accuracy is SVM. It appears that SVM, as a binary classifier, is the most efficient in defining the nominal region in the feature space when only two classes (healthy and unhealthy) are present. As a consequence, this minimizes the number of classification errors. However, what represents strength in the case of fault detection, will become a weakness within the proposed classification framework.

Table 4. Detection results

Method	FA%	MD%	Accuracy
AANN	11.4	11.0	88.7%
SVM	0.4	0	99.9%
GMM	8.1	0	98.3%

To determine the classification performance, the confusion matrices have been computed for the three methods. When two faults were present in a sample, classification has been considered correct if one of the two faults have been ranked first. Using this assumption, accuracy has been determined as the percentage of the number of samples correctly classified divided by the size of the test set. Results are presented for each machine learning method in tables 5,6 and 7.

The fault case in which most miss-classification occurred is the unbalanceD2. This is due to the fact that disk 2 is heavier and has a smaller radius than disk 1. As the same weight has been used to generate the unbalance on both disks, the specific unbalance is less on disk 2 than on disk 1: then UnbalanceD2 fault has less impact on the vibration signal. The misclassification is larger for SVM and GMM whilst AANN performs much better. Being a reconstruction method AANN has a more 'holistic' approach in the analysis of the data, which results in the capability to take into account the relative sizes of the components of the feature vector and not simply their individual positions in the feature space.

Tests also show that whenever two faults are present, the anomaly is always detected and one of the two faults is always the first in the ranking for all machine learning modules. The overall best accuracy for classification, presented in Table 8, is obtained with the AANN method. SVM, in this case, significantly underperforms the other two methods.

SVM is a "boundary method" which gives more weight to samples close to the border during learning. It is by "nature" a binary classifier and this clearly makes it more suitable for a two class problem. However our classification strategy compares the results of several two-class modules. In this case a class membership indication appears more appropriate than the binary output provided by SVM. Reconstruction methods as AANN and density methods like GMM make a more " democratic" use of the learning set allowing to use information about the whole class data distribution not just about its boundary. GMM and AANN produce a measure of belonging rather than a binary decision. This turns out into a better and more meaningful integration of the results from the different modules.

The logic conclusion of this analysis is that the best results are produced by the synergy between the different machine learning methods. In a schema which uses SVM for anomaly detection and then classifies the anomalies with AANN, 94.5% of correct classifications is achieved on our test set. This represents a remarkable result considering we only used healthy data for training.

Table 5. AANN classification confusion matrix

		Classification			
		Nominal	UnbalD1	UnbalD2	Misalig
Fault	Nominal	620	0	0	80
	UnbalanceD1	0	701	0	99
	UnbalanceD2	287	0	411	102
	Misalignment	0	0	0	200
	Mis.&unbalD1	0	0	0	400
	Mis.&unbalD2	0	0	0	400

Table 6. SVM classification confusion matrix

		Classification			
		Nominal	UnbalD1	UnbalD2	Misalig
Fault	Nominal	697	3	0	0
	UnbalanceD1	0	474	100	226
	UnbalanceD2	0	485	216	99
	Misalignment	0	0	0	200
	Mis.&unbalD1	0	246	0	154
	Mis.&unbalD2	0	0	103	297

Table 7. GMM classification confusion matrix

		Classification			
		Nominal	UnbalD1	UnbalD2	Misalig
Fault	Nominal	643	27	0	30
	UnbalanceD1	0	797	0	3
	UnbalanceD2	0	577	223	0
	Misalignment	0	0	0	200
	Mis.&unbalD1	0	263	0	137
	Mis.&unbalD2	0	0	0	400

Table 8. Summary of classification results

Method	Accuracy%
AANN	82.8
SVM	72.3
GMM	80.7

6. CONCLUSIONS

A strategy for data driven fault classification when only healthy data is used for training has been proposed. Results obtained with three different machine learning methods have been presented. The experimental analysis conducted at two different regimes indicated that whilst SVM performances are close to 100% accuracy for anomaly detection, AANN method outperforms for fault classification. A synergic exploitation of the different machine learning techniques allows obtaining overall 94.5 % of correct classifications over a test set containing four data classes (one healthy and three types of fault). Using this strategy provides also a fault ranking estimation, and this is particularly useful for this application on rotating machinery, such as gas and steam turbines, where, almost always, an unbalance fault shows up as well, even if another fault is the root cause of the anomaly. Examining the classification details is then a true benefit of the method.

The proposed method currently produces one classification result for each new sample. The focus of our research is nowadays on the integration of the classification results over time hence considering more samples to produce the classification decision. This should allow enforcing the robustness of the method and further improving the classification results. In parallel, the same method can be extended to other fault types. There should be no limit in the number of classes, as soon as separate sets of features are available for each fault.

NOMENCLATURE

L	Number of faults
T_i	Threshold for the fault i
D_i	Distance for the module i
$\boldsymbol{x}$	Sample under test
X	Sample of the training set

REFERENCES

Alguindigue, I., & Uhrig, R. E. (1991). *Vibration monitoring with artificial neural networks.* Tennessee: OECD publishing.

Bilmes, J. A. (1998). *A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models.* Berkeley: International computer science institute.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* New York: Springer Science+Business Media.

Booth, C., & McDonald, J. R. (1998). The use of artificial neural networks for condition monitoring of electrical power transformers. *Neurocomputing*, 97-109.

Fulufhelo, V. N., Tshilidzi, M., & Unathi, M. (2005). Early classifications of bearing faults using hidden Markov models, Gaussian mixture models, Mel-frequency cepstral coefficients and fractals. *International Journal of Innovative Computing, Information and Control*, 1281-1299.

Guttormsson, S., Marks, R., El-Sharkawi, M., & Kerszenbaum, I. (1999). Elliptical novelty grouping for on-line short-turn detection of excited running rotors. *Energy Conversion, IEEE Transactions on*, 16-22.

Jack, L. B., & Nandi, A. K. (2002). Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms. *Mechanical Systems and Signal Processing*, 373-390.

Johannes, M. D. (2001). *One-class classification.* Delft: Advanced School for Computing and Imaging.

Kramer, M. A. (1992). Autoassociative neural networks. *Computers & chemical engineering*, 16(4), 313-328.

Ng, A. (2015, September 21). *Machine Learning Course Materials.* Retrieved January 15, 2016, from http://cs229.stanford.edu/materials.html

Prego, T. d., de Lima, A. A., Netto, S. L., da Silva, E. A., Gutierrez, R. H., Monteiro, U. A., . . . Vaz, L. (2013). On Fault Classification in Rotating Machines using Fourier Domain Features and Neural Networks. *Circuits and Systems (LASCAS), 2013 IEEE Fourth Latin American Symposium on*, 1-4.

Rojas, A., & Nandi, A. K. (2006). Practical scheme for fast detection and classification of rolling-element bearing faults using support vector machines. *Mechanical Systems and Signal Processing 20*, 1523-1536.

Rubio, E., & Jáuregui, J. C. (2011). Time-Frequency Analysis for Rotor-Rubbing Diagnosis. In F. Ebrahimi, *Advances in Vibration Analysis Research* (pp. ISBN: 978-953-307-209-8, InTech, DOI:10.5772/15186).

Samanta, B., Al-Balushi, K. R., & Al-Araimi, S. A. (2003). Artificial neural networks and support vector machines with genetix algorithm for bearing fault detection. *Engineering Applications of Artificial Intelligence*, 657-665.

Sanz, J., Perera, R., & Huerta, C. (2007). Fault diagnosis of rotating machinery based on auto-associative neural networks and wavelet transforms. *Journal of Sound and Vibration 302*, 981-999.

Schölkopf, B., Platt, J. C., Shawe-Taylor, J., & Smola, A. J. (2001). Estimating the Support of a High-Dimensional Distribution. *Neural Computation, 13*(7), 1443-1471.

BIOGRAPHIES

Gianluca Nicchiotti received a MSc in Physics from Università di Genova, Italy in 1987 and a MSc in Applied Science from Cranfield University UK in 2013. Since 2005, he has been working in Meggitt Sensing System as SW engineer. Prior to Meggitt, he managed an image processing research team for offline cursive handwritten recognition at Elba Research Center. He has also developed algorithms for sport video special effects at Dartfish. His career started at Elsag Bailey R&D in underwater acoustic cameras field. He is author of about 40 papers and 4 patents in signal and image processing domain.

Lianel Etienne received a M.Sc degree in Communication System from Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, in 2016 as well as a Minor in Spatial Technology at the EPFL. He spent the last 6 months in Meggitt Sensing System, Fribourg, developing and writing its master project in automation of rotating machinery diagnostic process. In 2014, he had the opportunity to do a six-month placement at HUBER+SUHNER in the Radio-Frequency Research and Development division where he designed an active fiber optic frontend with application to WLAN. In 2010 he did his military service in the Grenadier Battalions 30 of the Swiss Special Forces Command (SFC) in which he is still incorporated.

Luc Fromaigeat Received a Doctorate in Geophysics from the INPL (Nancy, Lorraine, France) in 1982 on a new method for borehole exploration. After working 8 years in Alstom as Noise and Vibration specialist for measurement and diagnostics, he is now at Meggitt Sensing Systems for projects in Vibration Monitoring systems and site diagnostics worldwide. Focus is on signal acquisition, processing, electronics, and software, machinery protection and diagnostics, with particular focus on turbomachines applications. He is author of 5 publications and one patent in the field of vibration or acoustics. He was member of the International Hydropower Association