Application of AI failure identification techniques in condition monitoring using wavelet analysis

Cecilia Gattino[1], Elia Ottonello[2],Giovanni Berselli[3], and Jacek Stecki[4]

[1,2,3] *Università degli studi di Genova, Genoa, 16146, Italy*
cecilia.gattino@gmail.com
eliaottonello@gmail.com giovanni.berselli@unige.it

[4] *PHM Technology, Melbourne, Fitzroy North VIC, 3068, Australia*
jstecki@phmtechnology.com

ABSTRACT

In the context of Industry 4.0, Condition Based Maintenance (CBM) for complex systems is essential in order to identify failures and mitigate them. After the identification of a sensor set that guarantees the system monitoring, three main problems must be addressed for effective CBM: i) collection of the right data; ii) choice of the optimal technique to identify the specific data-set; iii) correct classification of the results. The solutions currently used are typically data driven and, therefore, the results are variable, as it is sometimes challenging to identify a pattern for all specific failures. This paper presents a solution that combines a data driven approach with an in-depth knowledge of the mechanical system's behaviour. The choice of the right sensor set is calculated with the aid of the software MADe (Maintenance Aware Design environment), whereas the optimal data-set identification technique is pursued with a second tool called Syndrome Diagnostics. After an overview of such methodology, this work also presents RSGWPT (Redundant Second Generation Wavelet Packaged Transform) analysis to show different possible outcomes depending on the available sensor data and to tailor a detection technique to a given data set. Supervised and unsupervised learning techniques are tested to obtain either an anomaly detection or a failure identification depending on the chosen sensor set. By using the described method, it is possible to identify potential failures in the system with sufficient notice to implement the optimal maintenance actions.

Keywords: Failure identification, Condition Based Maintenance, RSGWPT, Unsupervised learning, Supervised learning.

1. INTRODUCTION

The evolution of advanced sensor technology, computer science, Big Data, Artificial Intelligence (AI), Internet of Things (IOT) has laid the foundations for a new industrial revolution, namely Industry 4.0 (Yan, Meng, Lu, & Li, 2017). One of the Industry 4.0 goals is to truly enable Smart Factories, also taking into account maintenance issues starting from the early design stages. An intelligent/smart factory operates using interconnected advanced sensors. Thus, Big Data processing technology has became essential to build an integrated environment in which the production process can be controlled/managed in a more efficient way (Kang et al., 2016). Big data has prevailed in recent years with its potential to ascertain valued insights for enhanced decision-making, and it has become a hot spot in both academic research and practical applications (Yi, Liu, Liu, & Jin, 2014). In this scenario, reliability and safety are regarded among the most crucial factors of the intelligent system. Industry Big Data analytics will have great benefits, such as improving system performance, achieving near zero downtime, and ensuring predictive maintenance (Yan et al., 2017). Focusing on the latter aspect, failure identification or anomaly detection are essentials to perform Condition Based Maintenance (CBM) on industrial systems. Typically, many machine learning techniques are associated with a specific data-set in order to test the functioning of the different algorithms based on specific cases. However, there are several possible mathematical and engineering errors that often compromise the functioning of these algorithms and, consequently, the failure identification. In order to possibly improve the current practice, this paper proposes a method which aims at combining a system behavioural model with machine learning techniques, trying, as far as possible, to improve failure identification. The methods leverages on the use of two software tools named MADe (Maintenance Aware Design environment) and Syndrome Diagnostics (currently in development stage), whose main features are recalled hereafter along with some failure identification techniques implemented within the above mentioned tools. The rest of the paper is organized as follows: Sec. 2 describes the proposed methodology; Sec. 3 introduces the creation of a system model used to select the ideal sensor and to simulate the propagation of failures through the system; Sec. 4 describes the two signal processing techniques used in this paper, namely Redundant Second Generation Wavelet Transform (RSGWPT) and fault frequency calculation; Sec. 5 describes the clustering techniques; Sec. 6 describes the link between the causal model and the numerical results, whereas Sec. 7 explains the failure identification; Sec. 8 present a case study: finally, Sec. 9 draws our conclusions.

2. THE PROPOSED METHODOLOGY

The software *MADe - Maintenance aware design environment* (Hess, Stecki, & Clark, 2008) is a versatile tool that can be used for different purposes. It enables better decisions about the selection of critical equipment and aids in reducing risks via analysis capabilities that consider technical, operational and economic requirements of system's operators and maintainers. As said, building upon *MADe* existing capabilities, the aim of this paper is to improve failure identification techniques by combing MADe with a new tool named *Syndrome diagnostic*, in order to enable accurate, precise and reliable CBM. The main aspect that is discussed in this report is the failure identification process inside a system, a component or a part. The procedure is illustrated in Fig. 1. The three groups represent the main steps to follow in order to obtain the correct information from the system, leading to a complete failure identification, namely:

- Failure coverage aware data collection;
- Signal processing;
- Machine learning classification.

During the first phase, the mechanical system is modelled using MADe. Then, components failures are simulated in order to study their propagation trough the system. Using this information combined with the PHM analysis in MADe (that will be explained in section 3.2), it is possible to evaluate the ideal sensor set that is essential to collect the right data. In the second phase, explained in section 3.3, two different signal processing techniques are used depending on the data set. The second one (bottom block in Figure 1) is a classical fault frequency calculation that leads to an immediate failure identification. The first one (upper block) is based on a wavelet transformation (explained in section 4.2.7) followed by a statistical features extraction. Those features are then used in the third group to classify the signal using different machine learning techniques. When a single signal has been classified, it is fed to the tool named Syndrome Diagnostic that has the ability to correlate a pattern of signal variation to a specific failure (see section 6).

3. FAILURE COVERAGE AWARE DATA COLLECTION

The first section of this procedure is the Failure Coverage aware Data Collection, which is, in turn, divided into: i) MADe model creation and failure modeling; ii) PHM Analysis; iii) sensor set selection and Data acquisition.

3.1. MADe Model of system and system's failures

It is important to emphasize the usefulness of a causal model within a fault identification procedure, namely the fundamental reason leading to the use of MADe. The main objective, in this phase, is to create a Functional Block Diagram (FBD) of a system, in order to investigate flow perturbations and observing the system responses. MADe has a hierarchical structure; the three main types of blocks that can be incorporated in the system model are subsystems, components and parts (Hess et al., 2008). Each model comprises an input and an output block to display the flows coming in and out of the system. Inside each subsystem, there are components and/or parts and inside of each components could be parts or other components (although not mandatory). Components are connected together using flows and efforts connections related to the functions that have been performed by them. For each component, it is essential to define a function and also the inflows and outflows that are passing through.There are three types of flows, namely Energy, Materials, Signal. Figure 2 shows the main steps to create a MADe system model, which can be built following two different methods: FCM (Fuzzy Cognitive Mapping) model or Bond modelling. As for FCM, it can be used to model the functional behaviour of a complex system. FCM is suitable to model signal and process-based systems and it can be used to generate simulation response graphs and fault propagation tables to *qualitatively* analyse (and possibly mitigate) the system's criticalities. To better understand the functioning of MADe models, it is essential to show an example to analyse the relevant aspects for this failure identification process. Figure 3 shows the model of a vehicle system, that is used during MADe training, where several components are interconnected. Such model allows to simulate a system failure, although it does not provide quantitative data about the systems dynamic behaviour. As for the Bond graph technique, it is a domain-independent graphical representation of the system dynamic (D. Karnopp, 1968) and it is required to *quantitatively* assess the system response (thus, the knowledge of those coefficients employed in the model become essential), hence building up a so-called dynamic equivalent. Following from the system models, Failure diagrams can be built, which are a useful way to define all possible system failures. Failure diagrams look at the causes and progression of failures, focusing on their physical aspect. Figure 4 shows an example of an air filter failure diagram: green diamonds represent the failure concept mechanisms, blue triangles represent the causes, red circles represent the effects that are connected to the component outflow.

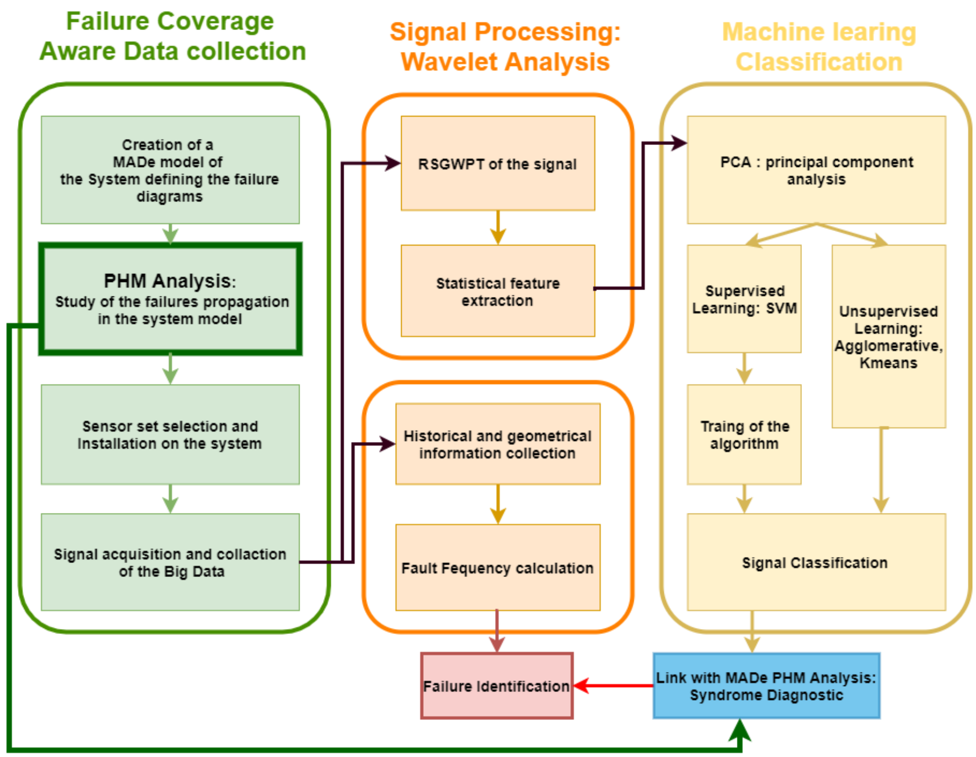

Figure 1. Conceptual steps of the proposed methodology

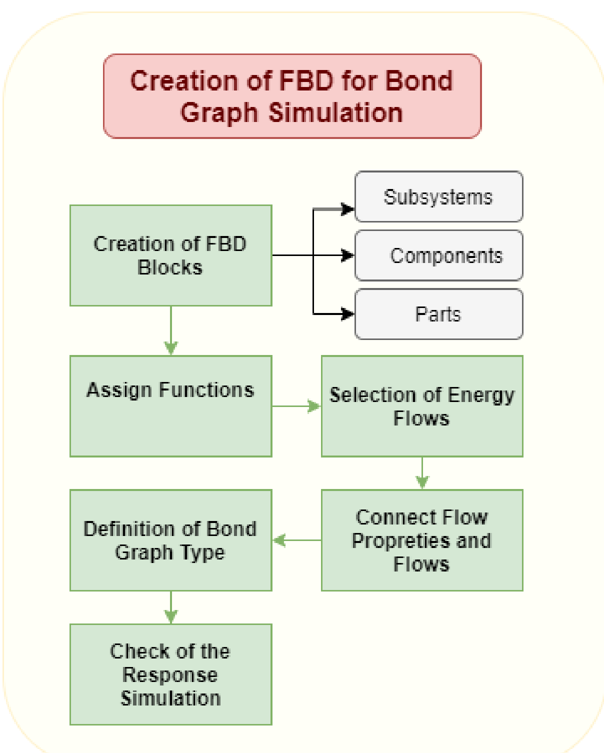

Figure 2. MADe modelling phases

3.2. PHM analysis

Once the system has been modelled, it is possible to perform the Prognostic Health Management (PHM) Analysis to generate the so-called propagation table. The main objective is to perform the following:

- Analysing a Functional Block Diagram (FBD) system to determine sensor test points;
- Modify existing sensor arrangements based on user knowledge or trade-offs;
- Optimising sensor coverage using sensor sets;
- Enter/customise sensor information into a Sensor Library.

The PHM analysis provides a sensor set selection technique adjusted on the system needs (Rudov-Clark, Ryan, Stecki, Stecki, & Hess, 2010). For example, by the choice of the system coverage, the user can adjust the percentage of coverage that is needed and calculate the number and the type of sensors required. The so-called Propagation Table (Shanna, Ryan, Stecki, & Stecki, 2009) is used for the analysis, and it is derived directly from the system model. It is one of two sets of data interrogated for sensor set analysis, the other being the observed failure responses that each sensor in the set may detect. The table consists of:

- Initiating causes of failures listed in the left-most columns (Item, Flow Property, Failure);
- Item Responses for remaining columns across the table;

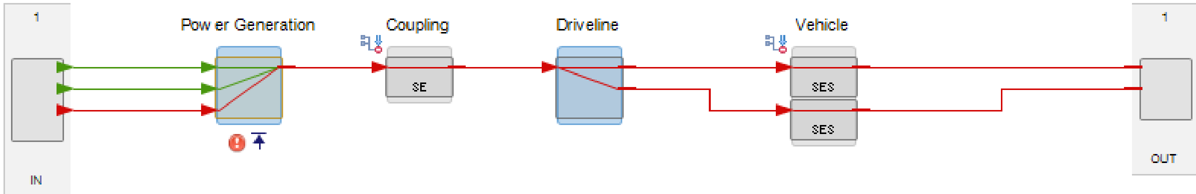

Figure 3. System model of a vehicle system

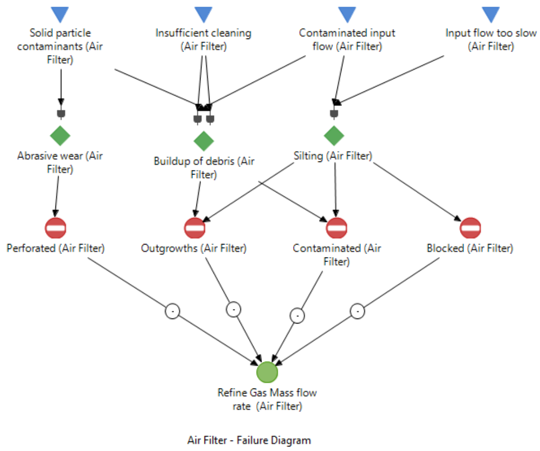

Figure 4. Failure diagram

- The healthy system state (first row), which represents the system at a nominal state, included in the propagation table, to ensure that a sensor set can observe a failure occurrence.

The result of the PHM analysis is a list of possible sensor sets that are as close as possible to the characteristics chosen by the user.

3.3. Sensor set selection

The choice of the sensor set that is more suitable for the analysed system is guided by the results proposed by MADe PHM analysis, that provides a list of combinations of sensor sets taking into account the parameters chosen by the user. The most important parameters that needs to be considered are: i) coverage; ii) cost; iii) ambiguities groups. The first two usually direct the choice but it is important to consider the ambiguity groups, a group of failures that cannot be distinguished with the sensor set under examination. This aspect is essential for the uniqueness of the pattern linked to a specific failure, so it is crucial that all the critical failures are not contained in an ambiguity group if they need to be detected separately. After those considerations, a sensor set needs to be selected and installed on the system in order to start the data acquisition.

4. SIGNAL PROCESSING

The second section of the proposed methodology is related to data processing, which is carried out using two techniques, depending on the given Data set, namely: i) fault frequency calculation; ii) RSGWPT. The latter approach consists in a wavelet transform of the signal to extract features from the wavelet coefficients. Those features will then be used to perform a machine learning classification of the given points. The first approach is more classical and it is based on historical and geometrical information to calculate the fault frequency.

4.1. Fault frequency calculation

In vibration monitoring, Fault Frequency Calculation is essential when direct failure identification of a critical component is performed. Unfortunately, every component has a different way to calculate the fault frequencies associated to it. An example of one of the most common component considered in vibration monitoring for mechanical systems is listed in Tab. 1, summarizing different faults frequencies of a rolling bearing (Saruhan H., 2014). Symbols are defined as follows: z is the number of balls, n_e and n_i are the angular velocities of outer and inner rings respectively, α is the ball/roller contact angle, λ is the diameter ratio.

4.2. RSGWPT

4.2.1. Fourier analysis vs wavelet analysis

Traditional vibration signal analysis has generally relied upon the spectrum analysis via the Fourier Transform (FT). Fourier analysis transforms a signal from a time-based domain to a frequency-based one, thus generating the spectrum that includes all of the signal's constituent frequencies (fundamental and its harmonics) (Al-Badour, Sunar, & Cheded, 2011). Fu-

Table 1. Bearing fault Frequency

Inner ring fault	$f_i = \frac{z\|n_e - n_i\|}{120}(1 + \lambda cos\alpha)$
Outer ring fault	$f_e = \frac{z\|n_e - n_i\|}{120}(1 - \lambda cos\alpha)$
Balls fault	$f_v = \frac{z\|n_e - n_i\|}{120}\frac{1-(\lambda cos\alpha)^2}{\lambda}$
Cage fault	$f_p = \frac{n_e + n_i}{120} + \frac{n_e - n_i}{120}\lambda cos\alpha$

Item	Flow Property	Failure	Differential Mechanical - rotational Angular velocity	Driveshaft Mechanical - rotational Torque	Half Shaft Front Mechanical - rotational Torque	Half Shaft Rear Mechanical - rotational Torque	Planetary Gearbox Front Mechanical - rotational Angular velocity
Driveline		Healthy System (SS,TR)	⇔	⇔	⇔	⇔	⇔
Differential	● Mechanical - rotational - Angular velocity	⇧ High (SS)	⇔	⇔	⇧ High	⇧ High	⇔
Differential	● Mechanical - rotational - Angular velocity	⇩ Low (SS)	⇔	⇔	⇩ Low	⇩ Low	⇔
Driveshaft	● Mechanical - rotational - Torque	⇧ High (SS)	⇔	⇧ High	⇧ High	⇧ High	⇔
Driveshaft	● Mechanical - rotational - Torque	⇩ Low (SS)	⇔	⇩ Low	⇩ Low	⇩ Low	⇔
Half Shaft Front	● Mechanical - rotational - Torque	⇧ High (SS)	⇧ High	⇧ High	⇧ High	⇩ Low	⇩ Low
Half Shaft Front	● Mechanical - rotational - Torque	⇩ Low (SS)	⇩ Low	⇩ Low	⇩ Low	⇧ High	⇧ High
Half Shaft Rear	● Mechanical - rotational - Torque	⇧ High (SS)	⇧ High	⇧ High	⇩ Low	⇧ High	⇧ High
Half Shaft Rear	● Mechanical - rotational - Torque	⇩ Low (SS)	⇩ Low	⇩ Low	⇧ High	⇩ Low	⇩ Low

Figure 5. Example of the so-called Propagation table, generated by MADe

elled by its huge success in processing stationary signals, unfortunately FT technique does not provide promising results on non stationary signals. In particular, the FT main problem is that it lacks time localization. A solution can be a time-frequency representation like Short-time Fourier Transform (STFT), although its disadvantage is that it provides constant resolution for all frequencies since it uses the same window for the analysis of the entire signal (Loutas & Kostopoulos, 2012). The Wavelet Transform (WT) is actually a time-scale method as it transforms a function from the time domain to the time-scale domain (scale is indirectly associated with frequency). The WT is also a reversible transform which makes the reconstruction or evaluation of certain signal components possible. WT became very popular in condition monitoring of complex signals (transient and/or non-stationary) for two specific characteristics: de-noising and feature extraction. Feature extraction provides the input to an expert system towards autonomic health degradation monitoring and data-driven prognostics.

4.2.2. Mother wavelet

To perform the classical Wavelet Transforms, namely Continuous Wavelet Transform, Discrete Wavelet Transform or Wavelet Packaged Transform, it is essential to select the correct mother wavelet. Different types of wavelets have different time–frequency structures and thus it is always an issue of how to choose the best wavelet function for extracting fault features from a given signal. An "inappropriate" wavelet will reduce the accuracy of the fault detection. The choice is often challenging and it can affect the results. Some decision techniques are provided in the reference (Wai Keng, Leong, Hee, & Abdelrhman, 2013).

4.2.3. Continuous Wavelet Transform (CWT)

A wavelet is a wave-like oscillation that instead of oscillating forever like harmonic waves, drops rather quickly to zero. The continuous wavelet transform breaks up a continuous signal into shifted and scaled versions of the mother wavelet ψ (Loutas & Kostopoulos, 2012).

4.2.4. Discrete Wavelet Transform (DWT)

The DWT is a discrete form of the CWT. It adopts the dyadic scale and translation to reduce computation time (Loutas & Kostopoulos, 2012). The curves of scales function can be modified by the scale parameter which is the inverse ratio to frequency. The DWT analysis of a signal is calculated by passing through a series of filters. The filters include both high-pass filters and low-pass filters. The high-pass filters analyse the high frequency bands (A_J), whereas the low-pass filters analyse the low frequency band (DJ). A_J and D_J represent the approximation and the detail signals of the J^{th} level. The decomposition is repeated to increase the resolution of the frequency domain and the transient signals can be used to detect the fault in the DWT domain. The number of decomposition levels N is related to the sampling frequency of the signal being analysed (fs). In order to get an approximation signal containing frequencies below frequency f, the number of decomposition levels that have to be considered is given by (M. Steinbuch, 2005):

$$N = int\left(\frac{log(\frac{fs}{f})}{log(2)}\right) \tag{1}$$

4.2.5. Wavelet Packaged Transform (WPT)

The WPT simultaneously decomposes approximations and details whilst the DWT only breaks up the approximations. In the first resolution, $j = 1$, the signal is decomposed into two packets: A and D. Camp A represents the lower frequency component of the signal, whilst camp D represents the higher frequency component of the signal. Then, at the second resolution, $j = 2$, each packet is decomposed again into two sub-camps forming AA, AD, DA, DD and so on (Loutas & Kostopoulos, 2012). The wavelet packets contain the information of the signal in different time windows

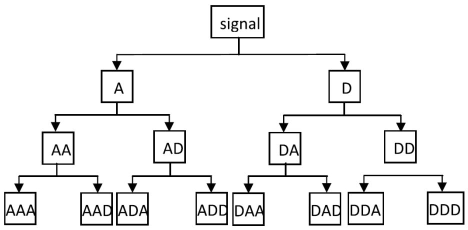

Figure 6. Wavelet packaged transform decomposition (Loutas & Kostopoulos, 2012)

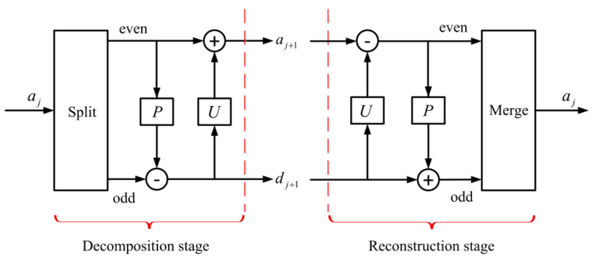

Figure 7. Lifting scheme (Liu et al., 2017)

at different resolutions. Each camp corresponds to a specific frequency band. Both WPT and DWT operate within the framework of multi-resolution analysis (MRA). Unlike DWT though, WPT has the same frequency bandwidth in every level. Figure 6 depicts the WPT decomposition tree with A and D corresponding to approximation and detail respectively. The WPT can thus be seen as a generalization of the WT and the wavelet packet function is also a time–scale function (Richard L. Lemaster, 2012).

4.2.6. Second Generation Wavelet Transform (SGWT)

SGWT is a new wavelet construction method using the lifting scheme. The main feature of the SGWT is that it provides an entirely spatial domain interpretation of the transform as opposed to the traditional frequency domain based constructions (Tong et al., 2017). Compared with the classical wavelet transform, the lifting scheme possesses several advantages including the possibility of adaptive design, in-place calculations, irregular samples and integers-to-integers wavelet transforms. The lifting scheme provides high flexibility, which can be designed according to the properties of the given signal and ensures that the resulting transform is always invertible. The multi-resolution analysis property is preserved, and the implementation is faster than the first generation wavelet transforms.

Figure 7 shows the decomposition of the SGWT that consists of three main steps: split, predict, and update. In the split step, an approximate signal at level l is split into even and odd samples. The even samples are passed to the predict function and they are subtracted from the odd samples then the result is sent to the update function and is added to the even part. The construction of the SGWPT is presented in Figure 8 (Tong et al., 2017).

4.2.7. Redundant Second Generation Wavelet Packaged Transform (RSGWPT)

In the redundant lifting scheme, the splitting step is discarded. Assuming P^l and U^l represent the prediction and update operators of the redundant lifting scheme at level l, the coefficients of P_l and U_l are obtained by padding prediction coefficients p_n and update coefficients u_n of initial operator P and U with zeroes (Hongkai, Zhengjia, Chendong, & Peng, 2006).

$$p_i^l = p_0^0, \underbrace{0, \ldots 0}_{2^l-1}, p_1^0, \underbrace{0, \ldots 0}_{2^l-1}, p_2^0, \ldots, p_{N-2}^0, \underbrace{0, \ldots 0}_{2^l-1}, p_{N-1}^0 \quad (2)$$

$$u_j^l = u_0^0, \underbrace{0, \ldots 0}_{2^l-1}, u_1^0, \underbrace{0, \ldots 0}_{2^l-1}, u_2^0, \ldots, u_{N-2}^0, \underbrace{0, \ldots 0}_{2^l-1}, u_{N-1}^0 \quad (3)$$

The redundant lifting scheme possesses time invariant property and keeps the signal information. The redundant decomposition results of an approximation signal s_l at level l with the lifting scheme are expressed by the following equations:

$$d_{l+1} = s_l - P^l s_l \quad (4)$$

$$s_{l+1} = s_l - U^l d_{l+1} \quad (5)$$

where d_{l+1} and s_{l+1} are detail signal and approximation signal at level $l+1$. The reconstruction procedure of the redundant lifting scheme is expressed as:

$$s_l = \frac{1}{2}\left(s_{l+1} - U^l d_{l+1} + d_{l+1} + P^l\left(s_{l+1} - U^l d_{l+1}\right)\right) \quad (6)$$

The forward and inverse transforms of the redundant lifting scheme are shown in Figure 9. RSGWPT is easy to be constructed starting from the redundant lifting scheme and SGWPT. The prediction step and update step of RSGWPT at level l are performed by using P_l and U_l, which are expressed as follows:

$$\begin{cases} s_{l,l} = s_{(l-1),1} - P^l\left(s_{(l-1),1}\right) \\ s_{l,2} = s_{(l-1),1} + U^l\left(s_{l,1}\right) \\ \cdots \\ s_{l,(2^l-1)} = s_{(l-1),2^l-1} - P^l\left(s_{(l-1),2^l-1}\right) \\ s_{l,2^l} = s_{(l-1),2^l-1} + U^l\left(s_{l,(2^l-1)}\right) \end{cases} \quad (7)$$

The reconstruction stage of RSGWPT can be obtained from its decomposition stage and expressed by following equations (Liu et al., 2017):

$$\begin{cases} s_{(l-1),2^{l-1}} = \frac{1}{2}\left(s_{l,2} - U^l\left(s_{l,1}\right)\right) + s_{l,(2^l-1)} \\ \quad + P^l\left(s_{l,2^l} - U^l\left(s_{l,(2^l-1)}\right)\right) \\ \cdots \\ s_{(l-1),1} = \frac{1}{2}\left(s_{l,2} - U^l\left(s_{l,1}\right) + s_{l,1}\right. \\ \quad \left. + P^l\left(s_{l,2} - U^l\left(s_{l,1}\right)\right)\right) \end{cases} \quad (8)$$

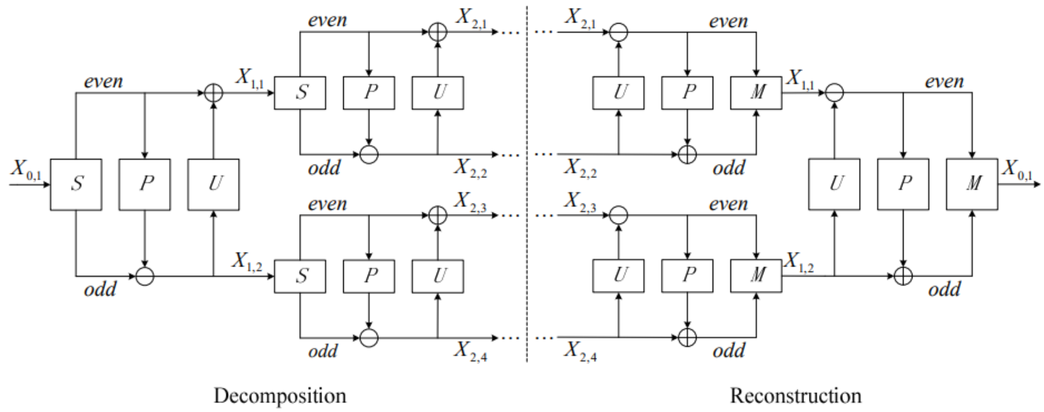

Figure 8. SGWPT Lifting scheme (Tong et al., 2017)

The RSGWPT not only possesses time invariance but can also match the characteristics of vibration response signals, so the features extracted from the resultant wavelet packet coefficients of RSGWPT have a greater ability to reveal the state changes of the system.

4.3. Statistical feature extraction

Feature extraction is performed using statistical variables extracted from the wavelet coefficients of the signal. The features that have been chosen are classical Statistical time-domain features such as mean, root mean square (RMS), standard deviation and variance. These were usually used in past studies to identify the differences between one vibration signal and another. More advanced statistical-based features such as skewness and kurtosis can be applied to the signal which is not purely stationary. These features examine the probability density function (PDF) of the signal. It is a well-known fact that if the condition of the component changes, the PDF also changes, thus the skewness and kurtosis might also be affected. In particular, skewness is used to measure whether the signal is negatively or positively skewed, whereas kurtosis measures the peak value of the PDF and indicates if the signal is impulse in nature. For a signal with a normal distribution i.e., bearing without faults in a nominal state, the skewness has value of zero (Caesarendra & Tjahjowidodo, 2017). Percentile also is been tested and provides a good representation of the vibration signals. The features that have been used are: RMS, Variance, Skewness, Kurtosis, Shape factor, Crest factor, Entropy, Percentile 50, 75, 5.

5. MACHINE LEARNING CLASSIFICATION

5.1. Principal component analysis (PCA)

Considering the number of features and the amount of data that is analysed, a dimension reduction technique is essential to complete the data analysis. Principal Component Analysis (PCA) is a technique for reducing the dimensionality of large data-sets thereby increasing interpretability but, at the same time, minimizing information loss. It does so by creating new uncorrelated variables that successively maximize variance (Maltoni, 2019).

5.2. Supervised and unsupervised learning

Supervised and unsupervised learning are used under different conditions. Supervised learning uses built data labels keeping track of the system failures during the data collection, or calculated using different signal processing techniques. Unsupervised learning can be used to classify unlabelled datasets, although not always with excellent results (Sathya & Abraham, 2013).

5.2.1. Supervised Learning: SVM

Support Vector Machines (SVM) have been developed in the framework of statistical learning theory (Vapnik, 1998; Cortes, 1995), and have been successfully applied to a number of applications ranging from time series prediction (Fernández, 2020), to face recognition (Tefas, Kotropoulos, & Pit, 2000), to data processing for medical diagnosis (Veropoulos, Cristianini, & Campbell, 1999; Evgeniou & Pontil, 2001). SVM is a supervised learning technique based on the methods of separating hyper-planes. The function used in this paper is the Radial Basis Function (RBF) kernel, a kernel that is in the form of a radial basis function (more specifically, a Gaussian function). The RBF Kernel SVM is used in many applications because there are many classification problems that are not linearly separable or regressable in the space of the inputs. These problems might be classifiable in a higher dimensionality feature space, given a suitable mapping. To correctly perform a classification, it is essential to optimize two main hyper parameters to avoid problems such as over-fitting. Those two parameters are γ and C (Pedregosa et al., 2011). To avoid over-fitting, it is also important the cross-validation of the results.

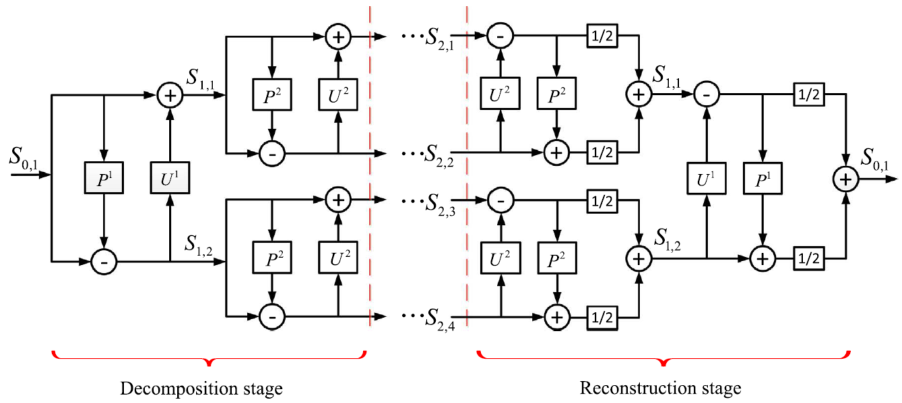

Figure 9. Redundant Lifting scheme (Liu et al., 2017)

5.2.2. Unsupervised Learning: Agglomerative

Agglomerative clustering schemes start from the partition of the data set into singleton nodes and merge step by step the current pair of mutually closest nodes into a new node until there is one final node left, which comprises the entire data set (Müllner, 2011). The algorithms are generally bottom-up and the end goal is trying to aggregate individual elements. At each step (level) the agglomerative algorithm aggregates the most similar elements (pattern to pattern, pattern to cluster, or cluster to cluster), i.e. the less distant ones from a threshold which depends on the level (Maltoni, 2019). The algorithms to compute the distance between clusters are average, centroid, complete, median, single, ward, weighted (Müllner, 2011).

5.2.3. Unsupervised Learning: K-Means

K-means minimizes distances from centroids. It requires as an input the number of clusters (S) and an initial solution. It produces good results as long as a reasonable initial solution and an adequate number of classes are provided. It identifies hyper-spherical clusters if the Euclidean distance is used as a measure of distance between the patterns or hyper-ellipsoidal clusters in the case of Mahalanobis distance (Maltoni, 2019). The limitations of this technique are the possibility to create only spherical clusters that sometimes are not the ideal shape for all the data-sets. In addition, it exists the risk of convergence towards local minima (Likas, Vlassis, & J. Verbeek, 2003).

5.3. Signal classification

Usually, by using only a numerical approach with clustering algorithms, it is very complicated to identify faults in complex mechanical systems. In fact, the quantity of signals and the relationships between them are often manageable only with a solid knowledge of system failure mechanisms. Anomalies are easier to be found and they are usually classifiable as follows:

- Several types of system, components, or parts failures;
- A parameters change linked with the data acquisition;
- False positives (non-existent failures incorrectly identified) or true negatives (non identified failures).

The biggest challenge is to identify the specific failure mechanisms that are occurring in the system. One strategy can be the direct identification using the fault frequencies calculation of critical component for vibration signals. It is essential to keep track of all the parameters adjustments such as gain changes, to avoid false alarms and also to have a correct correlation with the collected data and the results of the analysis usually performed later from another person.

6. LINK WITH MADE PHM: SYNDROME DIAGNOSTIC

The link between MADe PHM Analysis, explained in section 3.2, and failure identification using machine learing techniques has lead to the creation of Syndrome Diagnostic, a tool under development at PHM Technology. In this section, the idea behind it is explained. The propagation table (see e.g. Figure 5) contains a trinary flow variation pattern (Nominal, High, Low) that associates all failures to the sensed flows in the system. Unfortunately, the symptoms, like vibrations, are not present in this table as it only shows the modelled flows. Hence, starting from such propagation table (explained in section 3.2), Syndrome Diagnostics will have the ability to build a second table connecting the symptoms variation to the flows behaviour. This table contains a specific pattern composed by the signals of the sensors installed on the system, which is directly connected to a specific failure. The Syndrome Diagnostics capability is not only limited at failures detection but also aims at isolating failure and at providing a understandable visualization to the user. The outputs of the algorithm contained in Syndrome Diagnostic can be divided in three steps: i) the detection raises an alarm to the user when something is happening; ii) the isolation tends to identify the

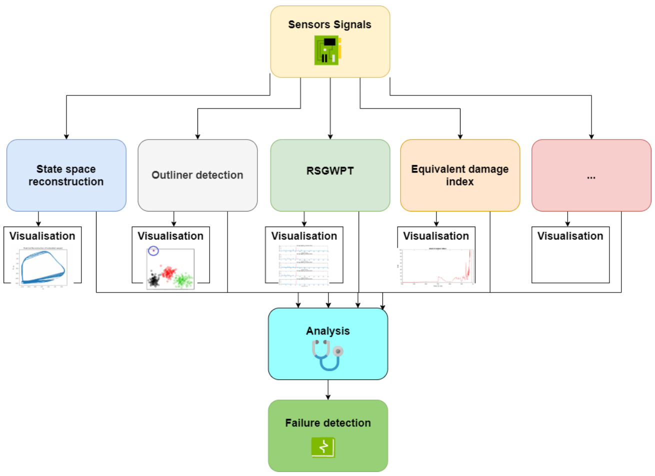

Figure 10. Failure identification techniques

said failure and to help the user understand what happened and where; iii) the signals are visualized and displayed in an easy way to understand. The new tool shows the reliability of a specific failure using the probability of detection associated with each sensor. The results are compared to the propagation table in order to quantify the probability/confidence of each possible state of the entire system, i.e. of each possible failure. A higher probability indicates more confidence in the result.

7. Failure identification

At last, the failure detection process is run in parallel with several analyses shown in Figure 10. The first one, State Space Reconstruction (SSR) (Casdagli, Eubank, Farmer, & Gibson, 1991), is a method focusing on signal showing recurrent patterns (dynamic systems) such as periodic signals. The idea is to draw a signal against itself delayed by a well chosen time. If the system is healthy, the drawn picture should be close to a geometric figure, whereas if a failure is occurring, the drawing is completely messy. This is extremely visually detectable and easy to understand. The second approach aims at detecting outliner from healthy situations. The idea is easy to understand and to visualize: signals are being clustered by common properties. If a data point (corresponding to a time) does not belong to any of the healthy groups (each group can be, for example, of a different regime), it is likely that a failure is occurring. The third one is the RSGWPT with machine learning classification (the one described in this paper), whilst the fourth one is the calculation as a cumulative damaging index that can be employed for monitoring fatigue sensitive components. PHM Technology is currently working on the development of other algorithms, which will be implemented in the failure identification procedures.

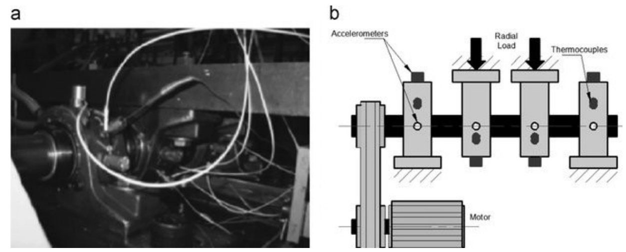

Figure 11. Data collection set up (Qiu et al., 2006)

8. Case study: Bearing IMS Data set

The mentioned features have been tested on data generated by the NSF I/UCR Center for Intelligent Maintenance Systems (Qiu, Lee, Lin, & Yu, 2006) with support from Rexnord Corp. in Milwaukee, WI. Four bearings are installed on a shaft. The rotation speed is kept constant at 2000 RPM by an AC motor coupled to the shaft via rub belts. A radial load of 6000 lbs is applied onto the shaft and bearing by a spring mechanism. All bearings are lubricated. Rexnord ZA-2115 double row bearings are installed on the shaft as shown in Figure 11. PCB 353B33 High Sensitivity Quartz ICP accelerometers are installed on the bearing housing (two accelerometers for each bearing [x- and y-axes] for data set 1, one accelerometer for each bearing for data sets 2 and 3). Sensor placement is also shown in Figure 11. All failures occurred after exceeding designed life time of the bearing, which is more than 100 million revolutions. Three Data sets are included in the data packet, were each data set describes a test-to-failure experiment. Each data set consists of individual files that are 1-second vibration signal snapshots recorded at specific intervals. Each file consists of 20,480 points with the sampling rate set at 20 kHz.

8.1. Signal processing: Fault frequency calculation and RSGWPT

The Dataset contains a bearing with a defect on the external ring. Therefore, by knowing the characteristic fault frequencies of the specific bearings, the signal spectrum can be used to identify the fault peaks related to the defect for each timestep. They are a precious tool for data labeling of the signal. Figure 12 shows the comparison between a healthy instant and an instant with a fault. As shown in Figure 12 (Outer race fault), the peak is clearly detectable at the calculated fault frequency. Using a trigger, all the time-steps have been labelled. Then, the RSGWPT of the signal has been performed for every time-step and the wavelet coefficients have been

Table 2. Fault frequencies

Vibration Frequency Fundamental Train	0.0072
Vibration Frequency Inner Ring Defect	0.1617
Vibration Frequency Outer Ring Defect	0.1217
Vibration Frequency Roller Spin	0.0559

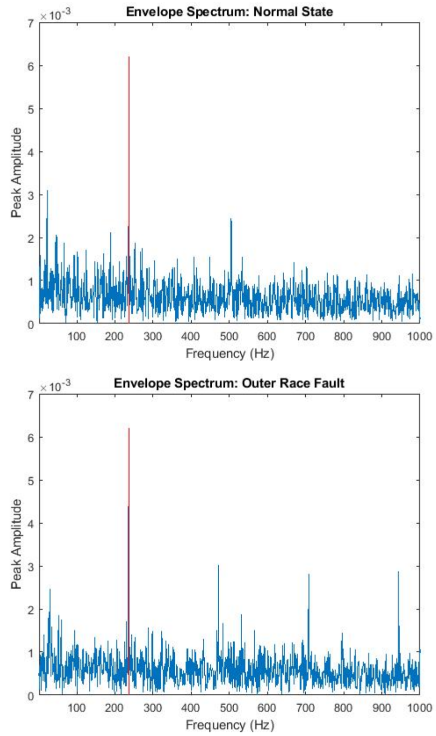

Figure 12. Comparison of bearing conditions at time-step 5 vs time-step 600

stored. The features mentioned on the previous chapters have been extracted and PCA has been performed to reduce the dimensions of the Coefficients of the wavelet*Feature Matrix. Subsequent computations are shown in the next paragraph.

8.2. Machine learning classification

As previously explained, three techniques have been tested for classification purposes, namely SVM, Agglomerative and K-means. SVM Kernel supervised learning algorithm has been used to classify the time-steps in 2 clusters, dividing 50% of the data in the training set and 50 % in the testing set. The algorithm has been cross validated and the hyper-parameters have been optimized. The results are visible in figure 13 (a). After those steps, the Agglomerative Unsupervised learning clustering Method has been performed to classify the time-steps in 3 clusters representing the Normal State, an Incipient Failure or a Failure. The results of the clustering algorithm are visible in Figure 13 (b). Also, the K-means Unsupervised learning clustering method has been performed to classify the time-steps in 3 clusters as in the previous case. The results of the clustering algorithm are visible in Figure 13 (c).

8.3. Failure identification

To evaluate the efficiency of the failure identification technique, the labels created using the fault frequency calcula-

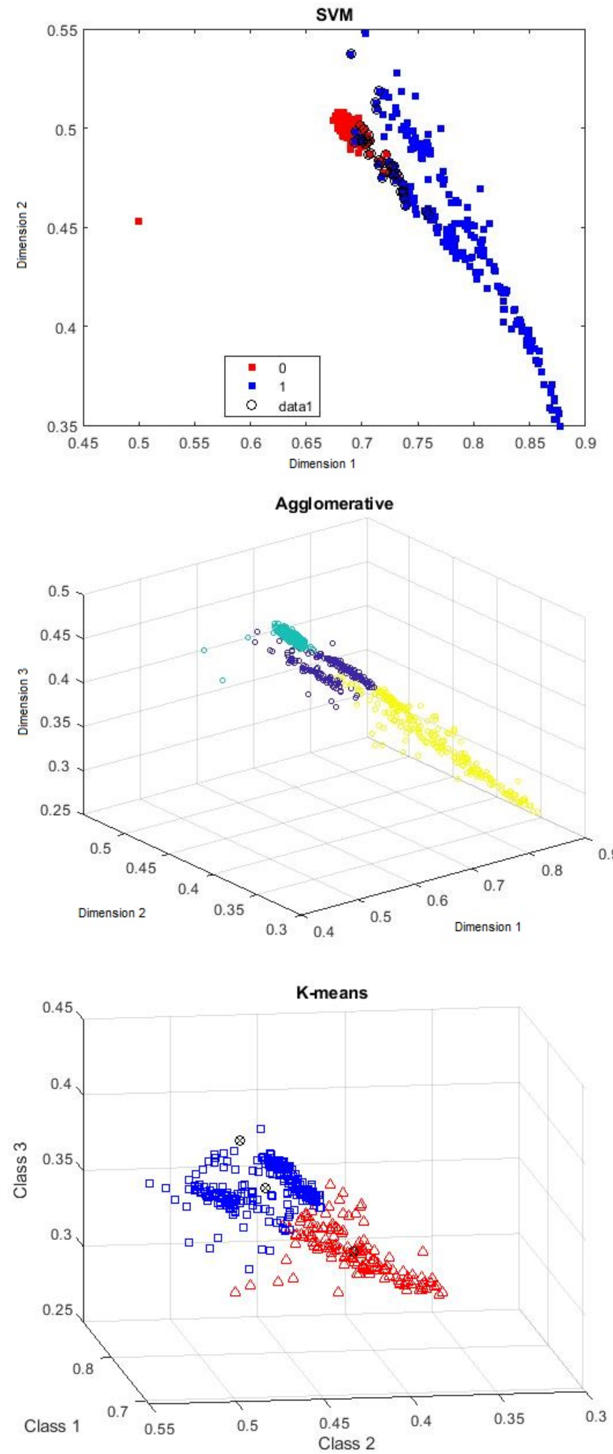

Figure 13. (a) SVM supervised learning classification (b) Agglomerative unsupervised learning classification (c) K-Means unsupervised learning classification

Table 3. Accuracy and F1score

Accuracy and F1score		
	Accurancy	F1 Score
SVM	0.9756	0.9930
Agglomerative	0.9726	0.9897
K-means	0.9685	0.9907

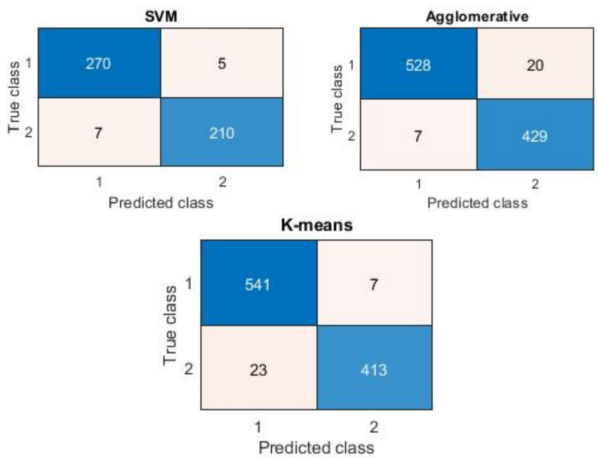

Figure 14. SVM (left) Agglomerative (right) an k-means (down) classification

tions have been used as ground truth to calculate the confusion Matrix that is visible in Figure 14. The tree techniques are able to identify the fault. Table 3 shows the results: the SVM performs better but agglomerative and K-means are still giving accurate results. Note that there is a difference between Agglomerative and K-means. In fact, Agglomerative appears to have a more conservative approach, reporting a greater number of false negatives. Therefore, it is less convenient in situations where there is a lot of noise and false alarms would be harmful. K-means is less conservative and records a higher number of false positives. Therefore, it is more advantageous in situations where the progression of failure is slow.

9. CONCLUSION

By using the methodology presented in this paper, it is possible to identify potential failures in a system with sufficient notice, in order to implement the optimal maintenance actions. On one hand, the aspect of combining a causal model associated with the use of artificial intelligence is certainly a promising approach. In parallel, the advantage of combining different data processing methods is also very useful, since some failures are not detectable from every signal or every techniques. Therefore, the coexistence of different techniques provides a winning weapon in those cases in which particularly complex identification issues are present. Among all possible ways to enable failure identification, this paper has focused on a traditional method (i.e. calculation of characteristic fault frequencies) and on the RSGWPT analysis, which allows the identification of faults in nonlinear transient signals. Future work will test a large variety of data sets to optimize the calculation times associated with each algorithm, in order to determine the best performing. In addition, a comparative cross evaluation of the provided failure identification accuracy (to be optimized on critical components) and the computational times (possibly allowing real-time monitoring applications) will be carried out.

ACKNOWLEDGMENT

We thank Paul Lebaigue, Navid Zaman, Christopher Stecki, Daniel Chan, Joshua Hi and all the PHM Technology team.

REFERENCES

Al-Badour, F., Sunar, M., & Cheded, L. (2011). Vibration analysis of rotating machinery using time–frequency analysis and wavelet techniques. , *25*(6), 2083–2101.

Caesarendra, W., & Tjahjowidodo, T. (2017). A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing. *Machines*, *5*(4), 21.

Casdagli, M., Eubank, S., Farmer, J., & Gibson, J. (1991). State space reconstruction in the presence of noise. *Physica D: Nonlinear Phenomena*, *51*(1), 52 - 98.

Cortes, V., Corinna. (1995, 09). Support vector networks. *Machine Learning*, *20*, 273-297.

D. Karnopp, R. C. R. (1968). Analysis and simulation of multiport systems: the bond graph approach to physical system dynamics. M.I.T. Press, University of California.

Evgeniou, T., & Pontil, M. (2001). Support vector machines: Theory and applications. In G. Paliouras, V. Karkaletsis, & C. D. Spyropoulos (Eds.), *Machine learning and its applications* (Vol. 2049, pp. 249–257). Springer Berlin Heidelberg. (Series Title: Lecture Notes in Computer Science)

Fernández, R. (2020, 07). Predicting time series with a local support vector regression machine.

Hess, A., Stecki, J., & Clark, S. (2008). The maintenance aware design environment: Development of an aerospace phm software tool. In *Proceedings of PHM08.*

Hongkai, J., Zhengjia, H., Chendong, D., & Peng, C. (2006). Gearbox fault diagnosis using adaptive redundant lifting scheme. *Mechanical Systems and Signal Processing*, *20*(8), 1992–2006.

Kang, H. S., Lee, J. Y., Choi, S., Kim, H., Park, J. H., Son, J. Y., . . . Noh, S. D. (2016). Smart manufacturing: Past research, present findings, and future directions. *International Journal of Precision Engineering* and Manufacturing-Green Technology, *3*(1), 111–128.

Likas, A., Vlassis, N., & J. Verbeek, J. (2003). The global k-means clustering algorithm. *Pattern Recognition*, *36*(2), 451–461.

Liu, Z., Guo, W., Hu, J., & Ma, W. (2017). A hybrid intelligent multi-fault detection method for rotating machinery based on RSGWPT, KPCA and twin SVM. *ISA Transactions*, *66*, 249–261.

Loutas, T., & Kostopoulos, V. (2012). Utilising the wavelet transform in condition-based maintenance: A review with applications. *Department of Mechanical Engineering and Aeronautics University of Patras, Rio, Greece*.

Maltoni, D. (2019). Riduzione dimensionalità. In *Dispense di machine learning.* Retrieved from `http://bias.csr.unibo.it/maltoni/ml`

M. Steinbuch, M. v. d. M. (2005). Wavelet theory and applications. *Eindhoven University of Technology Department of Mechanical Engineering Control Systems Technology Group Eindhoven*.

Müllner, D. (2011). Modern hierarchical, agglomerative clustering algorithms. *arXiv:1109.2378 [cs, stat]*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Qiu, H., Lee, J., Lin, J., & Yu, G. (2006). Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *Journal of Sound and Vibration*, *289*(4), 1066–1090.

Richard L. Lemaster. (2012). *The use of the wavelet transform to extract additional information on surface quality from optical profilometers.* INTECH Open Access Publisher.

Rudov-Clark, S., Ryan, A., Stecki, C., Stecki, J., & Hess, A. (2010). Extending advanced failure effects analysis to support prognostics and health management. In *2010 prognostics and system health management conference* (p. 1-5).

Saruhan H., A. U. I., Sandemir S. (2014). Vibration analysis of rolling element bearings defects. *Journal of Applied Research and Technology*, *12*(3), 384 - 395.

Sathya, R., & Abraham, A. (2013, 02). Comparison of supervised and unsupervised learning algorithms for pattern classification. *International Journal of Advanced Research in Artificial Intelligence*, *2*.

Shanna, R.-C., Ryan, A. J., Stecki, C., & Stecki, J. S. (2009). Automated design and optimisation of sensor sets for condition-based monitoring..

Tefas, A., Kotropoulos, C., & Pit, I. (2000, 07). Enhancing the performance of elastic graph matching for face authentication by using support vector machines. *IEEE Transactions on Pattern Analysis and Machine Intelligence - PAMI*.

Tong, Q., Cao, J., Han, B., Zhang, X., Nie, Z., Wang, J., ... Zhang, W. (2017, 03). A fault diagnosis approach for rolling element bearings based on rsgwpt-lcd bilayer feature screening and extreme learning machine. *IEEE Access*, *PP*, 1-1.

Vapnik. (1998, 01). Statistical learning theory. *Wiley*, *16*.

Veropoulos, K., Cristianini, N., & Campbell, C. (1999, 06). The application of support vector machines to medical decision support: A case study. *Adv Course Artif Intell*.

Wai Keng, N., Leong, M., Hee, L., & Abdelrhman, A. (2013, 09). Wavelet analysis: Mother wavelet selection methods. *Applied Mechanics and Materials*, *393 (2013)*, 953-958.

Yan, J., Meng, Y., Lu, L., & Li, L. (2017). Industrial big data in an industry 4.0 environment: Challenges, schemes, and applications for predictive maintenance. *IEEE Access*, *5*, 23484–23491.

Yi, X., Liu, F., Liu, J., & Jin, H. (2014). Building a network highway for big data: architecture and challenges. *IEEE Network*, *28*(4), 5–13.

BIOGRAPHIES

Cecilia Gattino she received the Bachelor degree in Mechanical Engineering from the University of Genova, and the Master degree in Mechanical Engineering Design and Production from the University of Genova in 2017 and 2020, respectively. She was an International student in Institut National des Sciences Appliquées de Strasbourg (INSA) Strasbourg.

Elia Ottonello he received the Bachelor degree in Mechanical Engineering from the University of Genova, and the Master degree in Mechanical Engineering Design and Production from the University of Genova in 2018 and 2020, respectively.

Giovanni Berselli is Associate Professor in Design Methods and Tools for Industrial Engineering at the University of Genova. He has been Research Associate with the Escuela Superior de Ingenieros de la Universidad de Navarra and with Monash University. He has also been Visiting Professor at the University of Twente and at the Medical Devices and Simulation Lab - Harvard Medical School.

Jacek Stecki Chief Technology Officer CTO, PHM located in Melbourne, Australia. Former Associate Professor and Director of Centre for Machine Condition Monitoring at Monash University, Melbourne, Australia. International Consultant in the field of subsea engineering, simulation, prognostics and health management (PHM) and hybrid vehicles. Over 35 year experience in research and consulting on fluid power, condition monitoring and simulation. Over 150 technical and scientific papers on the subject of fluid power, simulation, condition monitoring and subsea engineering. Currently involved in PHM programs in Mining and Aerospace industries.