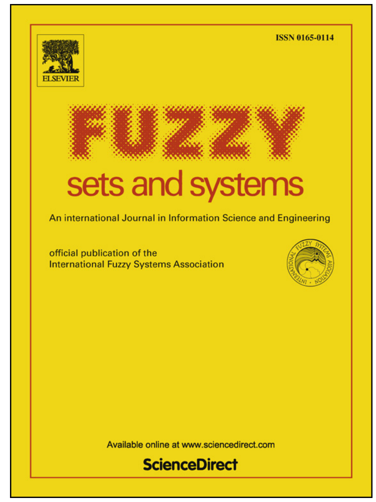


A fuzzy transition based approach for fault severity prediction in helical gearboxes

Mariela Cerrada, Chuan Li, René-Vinicio Sánchez, Fannia Pacheco, Diego Cabrera, Jose Valente de Oliveira

PII: S0165-0114(16)30458-4
DOI: http://dx.doi.org/10.1016/j.fss.2016.12.017
Reference: FSS 7150

To appear in: *Fuzzy Sets and Systems*

Received date: 25 April 2016
Revised date: 4 December 2016
Accepted date: 24 December 2016

Please cite this article in press as: M. Cerrada et al., A fuzzy transition based approach for fault severity prediction in helical gearboxes, *Fuzzy Sets Syst.* (2016), http://dx.doi.org/10.1016/j.fss.2016.12.017

Highlights

-
- A monitoring system based on fuzzy transitions for fault severity prediction is presented.
- Features extracted from vibration signals of rotating devices are used as input information.
- A static fuzzy model is used for computing the weights of fuzzy transitions (WFT).
- WFT depends on the knowledge of temporal behavior of samples associated to a fault degradation pattern.
- A dynamic equation using WFT allows predicting the next degradation state of the rotating device.

A fuzzy transition based approach for fault severity prediction in helical gearboxes

Mariela Cerrada[a,b,*], Chuan Li[c], René-Vinicio Sánchez[b], Fannia Pacheco[b], Diego Cabrera[b], Jose Valente de Oliveira[d]

[a] *CEMISID. Universidad de Los Andes, Mérida-Venezuela*
[b] *GIDTEC-Mechanical Engineering Department. Universidad Politécnica Salesiana, Cuenca-Ecuador*
[c] *National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, Chongqing 400067-China*
[d] *Universidade do Algrave, Faro-Portugal*

Abstract

Rotating machinery is an important device supporting manufacturing processes, and a wide research works are devoted to detecting and diagnosing faults in such machinery. Recently, prognosis and health management in rotating machinery have received high attention as a research area, and some advances in this field are focused on fault severity assessment and its prediction. This paper applies a fuzzy transition based model for predicting fault severity conditions in helical gears. The approach combines Mamdani models and hierarchical clustering to estimate the membership degrees to fault severity levels of samples extracted from historical vibration signals. These membership degrees are used to estimate the weighted fuzzy transitions for modelling the evolution along the fault severity states over time, according to certain degradation path. The obtained fuzzy model is able of predicting the one step-ahead membership degrees to the severity levels of the failure mode under study, by using the current and the previous membership degrees to the severity levels of two available successive input samples. This fuzzy predictive model was validated by using real data obtained from a test bed with different damages of tooth breaking in the helical

*Corresponding author
Email addresses: cerradam@gmail.com (Mariela Cerrada), chuanli@21cn.com (Chuan Li), rsanchezl@ups.edu.ec (René-Vinicio Sánchez), fannikaro@gmail.com (Fannia Pacheco), dcabrera@ups.edu.ec (Diego Cabrera), jvo@ualg.pt (Jose Valente de Oliveira)

gears. Results show adequate predictions for two scenarios of fault degradation paths.

Keywords: Fuzzy transition probability, Fuzzy prediction, Fault severity prediction, Fault severity classification, Fault detection and diagnosis

1. Introduction

Monitoring condition is an important activity to reach adequate reliability levels in industrial processes. Particularly, rotating machinery is a main device that supports an important class of manufacturing processes, where bearings and gears are important components of such rotating devices. Several works can be found that are devoted to detecting and diagnosing faults in rotating machinery by applying signal processing techniques, usually to vibration signals [1–6]. Intelligent systems have widely contributed in fault detection and diagnosis of rotating machinery; mainly, the obtained intelligent models are designed as classifiers which receive features extracted from vibration signal as the input vector. Expert systems are classical approaches to build intelligent models. In [7], an expert system is proposed as a tool for diagnosing the cause of abnormal vibration in rotating machinery and a decision tree is used for the acquisition of structured knowledge. A similar work is presented in [8], a set of rules is obtained from fault charts which are constructed with information extracted from vibration signals. In [9] is provided an extensive vision about machine learning based intelligent fault diagnosis, particularly the author focuses the attention on Neural Networks (NN), Deep Learning (DL), Statistical Learning (SL), Fuzzy Logic (FL), and hybrid approaches using neuro-fuzzy (NF) models.

NN based classifiers are very popular models for fault classification, a classical application is presented in [10]. In [11] a Back-Propagation NN (BPNN) is combined with Rough Set Theory (RST) and Particle Swarm Optimization (PSO) for bearing fault diagnosis. In [12], a Radial Basis Function NN (RBFNN) is used as classifier, and the parameter optimization is accomplished by an improved Differential Evolution (DE) algorithm. More complex NN based architectures, such as Deep Neural Networks (DNN) are presented as a promising intelligent technique for mining massive data obtained from rotating machinery [13]. A diagnosis model using DNN is proposed in [14], particularly an unsupervised Boltzmann Machine is used for the statistical feature representation, and its outputs are the input to a classical one layer BPNN.

SL such as Support Vector Machines (SVM) combined with intelligent approaches, are also studied to address the fault classification in rotating machinery. In [15], an intelligent diagnosis model using SVM combined with Ant Colony Optimization (ACO) is proposed for fault classification in locomotive roller bearings. ACO is used for synchronous feature selection and parameter optimization. SVM with ensemble-based incremental learning approach is presented in [16]. Classical application of SVM and the modified algorithm called Reproducing Wavelet SVM, can be found in [17, 18].

Clustering is another SL technique that has been combined with intelligent approaches such as FL, to propose fault classification models for rotating machinery. In [19], a new clustering algorithm based on an improved Fuzzy C-means (FCM) algorithm is proposed for fault diagnosis. Two phases for feature selection and weighting are performed by using distance evaluation. In [20], FCM based clustering is also used to develop fault classification, feature weighting is accomplished through a Feed Forward NN (FFNN), and sample weighting is computed by using the distribution density function of a sample. In [21], a fuzzy clustering is proposed to segment the frequency spectrum of the informative frequency band (IFB) from the vibratory signal, into meaningful sub-bands. Three commonly-used selectors are combined using a fuzzy comprehensive evaluation method to guide the clustering in order to select the IFB with the minimum comprehensive cost. A novel method allowing for interactive clustering to bearing fault diagnosis is proposed in [22]. The method provides an intuitive way to control the cluster formation process to select a desirable level of granularity ranging from fault detection to classification of a variable number of faults.

Besides fuzzy clustering, FL based models are also extensively used for developing fault detection and classification models of rotating machinery. A classical application of the Mamdani-type Fuzzy Inference System can be found in [23, 24]. In [25], a fuzzy diagnosis method using sequential inference and possibility theory was proposed; in this case, a fuzzy model discriminating two different health conditions is developed in successive phases with the relevant possibility of symptom parameters. A similar work using sequential fuzzy diagnosis method is presented is [26], where ACO is applied previously to optimize clustering process for extracting the representative symptom parameters. In [27], a fault diagnosis matrix is built from two fuzzy relationship matrix representing the historical based fault diagnosis matrix and the expert knowledge based matrix.

NF approaches have been also used to construct intelligent hybrid fault diagnosis models, being Adaptive Neuro-Fuzzy Inference System (ANFIS) the most popular ones [28, 29]. In [30], the performance of the ANFIS model as bearing fault classifier is compared with the classical NN. In [31], an ANFIS is designed with features that are discretized and selected by using fuzzy clustering and RST, respectively. An interesting hybrid approach based on a Fuzzy Neural Network (FNN) is presented in [32], the ambiguous relationship between symptoms and fault types are captured through the possibility theory and the Dempster & Shafer theory. This knowledge is used to train a FNN which is implemented by a partial linearised NN, under a sequential diagnosis method. In [33], several ANFIS models are trained through two learning stages, for determining different bearing status; rules extraction and interpretation is performed to obtain an interpretable fuzzy model. In [34], a fuzzy knowledge base is obtained and refined through an empirical procedure and a NN architecture, respectively.

Previous works show that bearing fault diagnosis using intelligent systems is the most studied device of rotating machinery. Our paper is focused on gearbox, as a component of rotating machinery that has an important role in mechanical power machines, due to its ability to transmit motion in most of the manufacturing systems. Recently, the authors have reported some results in the fault detection and diagnosis of gears by using vibration signals and machine learning techniques combining several paradigms of the artificial intelligence, such as Random Forest (RF) with Genetic Algorithms (GA) [35, 36], Artificial Neural Networks with GA [37, 38], SVM and RST [39], and RF with Deep Learning [40]. Additionally, other works combining artificial intelligence techniques for fault diagnosis in gears can be found in the recent literature. ANFIS adjusted through a Kohonen self-organizing feature map and GA is presented in [41]. In [42], multiple classifiers, such as FFNN, RBFNN and K-Nearest Neighbour (KNN) with different sets of input features are combined through a weighted averaging technique using GA to weight optimization. A weighted KNN classification algorithm is used to identify the gear crack levels in [43]. A similar work using KNN is discussed in [44]. Classical applications of SVM to fault diagnosis in gears can be found in [45, 46]; a modified SVM classifier called Relevance Vector Machine (RVM) optimized with ACO is studied in [47] to propose a better gear fault classification, two pairwise-coupled RVM are also used in [48] to develop a Probabilistic Committee Machine optimized through PSO. Rule based reasoning is presented in [46] by using a fuzzy reasoning strategy applied to a fuzzy relation matrix between fault causes and fault symptoms. Decision Tree (DT) and Random Tree are evaluated as classifiers for gear fault diagnosis in [49, 50], DT and GA with RST are used for feature selection. Moreover, acoustic emission analysis by artificial intelligence techniques has been also reported as a signal providing useful information for fault diagnosis in gears [50–52].

The next step after detecting and diagnosing faults aims at evaluating and predicting the fault severity of future condition, in order to take maintenance actions for minimizing the effect of the detected faults in the rest of the process. Recently, prognosis and health management in rotating machinery have received high attention as a research area, and some advances in this field are discussed in [53–56]. Some results in prognosis on gearboxes are presented in [54], by using different techniques: (i) signal processing such as empirical model decomposition wavelet transform, Hilbert-Huang transform, adaptive amplitude and phase demodulation, (ii) machine learning such as neural networks, fuzzy logic, neuro-fuzzy architectures and genetic algorithms, (iii) Statistical and probabilistic models such as Kalman filter, Support vector machines, auto-regressive models and particle filter.

One important task in prognosis is the assessment of fault severity or fault degradation in order to estimate the remaining useful life. For this purpose, probabilistic models are interesting frameworks for estimating futures stages of the damages. Recent results in this field using particle filtering are presented in [56], when signals are under non-Gaussian additive noise; particularly, the work refers to the usage of particle filtering on vibration feature data from a fatigue-driven fault in a critical aircraft component [57]. Stochastic filtering is applied in [58] for estimating a stochastic degradation process and uncertain condition monitoring measurements; the approach is validated with a simplified fatigue crack growth process, which can be used to model certain rotating machinery faults. A probabilistic model is proposed in [59] to estimate the conditional distribution of the system state with respect to the specific information available about this system. In [60], recursive Bayesian updating scheme is used to assess the current state of fatigue damages like cracks. Bayesian inference-based probability is developed in [61] as a quantification indicator of machine health degradation by analysing vibration signals. A high-order Markov process is used to predict the evolution of the machine health in the form of a probability density function.

On the other hand, Fuzzy Logic (FL) has been used for fault severity assessment with focus on bearing damages, some recent works are developed in [62–64]. In [63] a fuzzy-logic inspired process for estimating the severity of bearing faults is constructed. Gaussian distribution associated to the spectral content of vibration signals across frequency bins are identified, and they are used to define characteristic membership functions for each severity level. The approach is tested on different severity levels in the bearing inner race. An approach for early detection of faults in fan bearings and severity assessment is proposed in [64] by using a Wavelet Filter (WF). A fuzzy rule is introduced in order to match the WF that maximizes the amplitudes of the Sum of the Amplitudes of Bearing Characteristic Frequencies (SABCF), which is an indicator of bearing faults. Basically, SABCF is defined in a fuzzy way by using Gaussian membership functions that is described in terms of the theoretical BCF, such that the fuzzy membership function indicates that the closer the frequency to the theoretical BCF, the higher the confidence that the frequency is related to a bearing fault. The work in [62] proposes a monotonic degradation assessment index of rolling bearings using Fuzzy Support Vector Data Description (FSVDD) and running time. FSVDD constructs the fuzzy-monitoring coefficient which is sensitive to the initial defect and stably increases as faults develop.

In case of gears, FL has been applied to diagnosis, as presented in some works discussed above (see [23–29, 32–34, 41]), but its use for modelling fault degradation can only be found in few works. In [65], FL was applied to the oil debris and vibration data in order to build a model that discriminates between different stages of pitting wear in spur gears. An evolving fuzzy predictor based on clustering is proposed in [66] to forecast the beta kurtosis index for cracked and pitting gear monitoring; fatigue degradation process is also analysed in that work. Most solutions are developed through NF approaches where ANFIS models are devoted to predicting the degradation state. In [67], an ANFIS architecture is used to forecast damage propagation trends in several kinds of gear faults, based on the evolution of a monitoring index; this index uses a wavelet reference function measuring the energy concentration over a specific bandwidth. In [68], ANFIS and high-order particle filtering are used to forecast the time evolution of the fault indicator. ANFIS is used to model the fault degradation and the high-order particle filter is used to carry out p-step-ahead predictions via a set of particles. In [69], a NF system (NFS) is employed as a prognostic model to forecast the evolution over time of the machine fault state; in order to improve the degree of belief in the forecasting estimations, an updating scheme using Bayesian estimation algorithms, solved with the particle filtering method, is integrated to the NFS by taking into account the probability density function of residuals between the real (on-line measurements) and predicted condition data by the NFS.

Probability theory and FL are complementary and fuzzy probability theory arises in a natural manner [70, 71]. The application of probability, fuzzy membership and fuzzy probability have been discussed in [72, 73] to propose Fuzzy Transition Probabilities (FTP) for state monitoring. FTP combines the transition probability of Markov process with fuzzy sets. Given a set of historical samples associated with the process evolution in time, the matrix of FTP is computed in a training phase. For a new sample in the current time t the fuzzy probability of evolving from the state k to the state j at time $(t+1)$ is calculated by using the computed FTP. The work in [74] is based on the theory in [72, 73] to propose the prediction of the membership degree of the current sample to a certain state at time $(t+1)$. The approach allows computing the Weights of Fuzzy Transition (WFT) from certain current state k to a next state j based on the knowledge of the membership degree of the training samples to the state k at time t. Once the WFT are computed in the training phase, the prediction of the membership degree of a given current sample to a next state can be done.

This work applies the theory developed in [74] to predict the fault severity level in helical gears. According to our knowledge, FL is not widely reported as a technique for prediction of fault degradation assessment in rotating machinery, and this is our main contribution. Classical probabilistic approaches such as those mentioned previously using particle filtering imply complex computation of probability distributions to estimate the time evolution of the system output. In case of fault diagnosis for rotating machinery, the system output is associated to a performance index extracted from signals that, in some cases, is not easy to define. Additionally, the model for this index is a dynamic model that must be identified. On the other hand, the mentioned fuzzy approaches are static ones oriented to estimate fault severity states, but not for predicting them. The proposed approach in this work avoids this complexity and define the fault severity prediction as a combination of static and dynamic approaches. The static approach is for diagnosing the fault severity state, as a classic problem fuzzy classification, and the dynamic part is for taking into account the time evolution between severity states using the concept of fuzzy transitions, from the knowledge of temporal patterns.

Particularly, a progressive gear tooth breaking is considered as failure mode. Mechanical devices are subject to progressive faults, then the approach in [74] is modified to consider the future states that are associated with more degraded states regarding the current one, as proposed in [72]. The training phase assumes three classical scenarios of changes in the degradation path, such as [75]: (i) stationary to lineal trend,(ii) stationary to polynomial trend, and (iii) stationary to lineal to polynomial trend. In order to calculate the membership function of the samples to the different severity states, a Mamdani-type fuzzy inference system is designed as classifier and Fuzzy C-Means Algorithm is used to estimate the fuzzy rules. Next, a set of Gaussian membership functions are estimated by the hierarchical clustering of the output membership functions that are proposed by the classifier, in order to estimate the membership degree of the current output crisp value of the classification to the class associated with the severity level. The application of this algorithm uses real data from an experimental test bed, that collects vibration signals under different gear damages that occur progressively into a temporal pattern. Results show that the prediction of future states corresponds to the expected state according to the essayed scenarios.

This work is organized as follows. Section 2 presents the main results of the work in [74]. Section 3 discusses the application of this algorithm for fault severity prediction. Section 4 describes the experimental setup for generating real data about fault severity states in helical gears. Section 5 shows the application of the algorithm on the case of study, under different scenarios of fault degradation. Finally, Section 6 gives the conclusions.

2. Situation prediction based on fuzzy clustering and fuzzy transitions

The work in [74] is inspired by [72] for predicting functional states in industrial complex process. This is accomplished by estimating the membership degree of a future sample to different functional states, from the knowledge of the membership degree of the current sample to all states. One of the advantages of this approach is that it provides information about the evolution of the process. The methodology combines a static measurement, i.e. the result of a fuzzy classifier based on clustering trained with historical data to model situations, and an estimation algorithm inspired on FTP for discrete event systems. As mentioned in section 1, given a sample measured from the process which is associated to the functional state k, the FTP determines the probability of evolving from the current state k to the state j at $(t+1)$.

The methodology is performed in two stages. In the first stage (off line), two models must be adjusted from the historical data, that is: (i) a model based on fuzzy clustering to obtain the membership degrees of a given sample of the process to a functional state, and (ii) a system of linear equations to calculate the WFT of the prediction model. In the second stage (on line), the fuzzy clustering based model is used for identifying the current process situations and the prediction model is used for estimating the functional states expected in the future. According to the proposal in [74], WFT is a vector of weights that relates the membership degrees from the state k to the state j. In a rough sense, the vector WTP determines the transitions that are likely to happen from the state k to the state j, as an approximation of the FTP given in [72]. This section presents briefly the algorithm in [74], more details can be found in the corresponding reference.

Let S_{t+1} be a sample that is expected in the future. The one-step ahead prediction of the membership degree $\mu_j(S_{t+1})$ of the sample S_{t+1} to the state j is calculated by equation (1):

$$\mu_j(S_{t+1}) = \sum_{k=1}^{m} WFT_{kj} * \mu_k(S_t), j = 1, 2, ..., m \quad (1)$$

where WFT_{kj} is the weight of fuzzy transition from the state k to the state j, $\mu_k(S_t)$ is the membership degree of the sample S_t to the state k, and m is the number of functional states.

The development of equation (1) for each state produces equation (2), then, equation (1) can be arranged in the matrix equation (3) that allows predicting the membership degrees of a sequence of samples for each situation j at any time t, $t = 2, ..., n$ where n is the number of available samples:

$$\mu_j(S_{t+1}) = WFT_{1j} * \mu_1(S_t) + WFT_{2j} * \mu_2(S_t) + ... + WFT_{mj} * \mu_m(S_t) \quad (2)$$

$$[\mu_j(S_t)] = [\mu_k(S_{t-1})] * [WFT_{kj}], \forall j, t \quad (3)$$

With the historical information of the process, equation (3) is solved for WFT, by using the constrained Least Squares Method (LSM) subject to $WFT \geq 0$.

The change of membership degrees $\Delta\mu$ is calculated to allow including the information about the trend in membership degrees, according to the process evolution. In general, for any state j, the change $\Delta\mu$ is defined by equation (4):

$$\Delta\mu_j(S_t) = \mu_j(S_t) - \mu_j(S_{t-1}) \quad (4)$$

By considering the vectorial representation of the equation (4), the matrix equation (5) states for the estimation of ΔWFT, at any time t, $t = 2, ..., n-1$:

$$[\Delta\mu_j(S_t)] = [\Delta\mu_k(S_{t-1})] * [\Delta WFT_{kj}], \forall j, i \quad (5)$$

Then, vectors WFT and ΔWFT are calculated in the off line stage. As mentioned previously, $\mu_j(S_{t+i})$ and $\mu_j(S_{t+i-1})$ are calculated through some fuzzy classifier that is adjusted with the historical data.

In the on line stage, the initial prediction of a membership degrees $\mu_j(S_{t+1})$ is estimated by equation (6):

$$\mu_j(S_{t+1}) = \mu_j(S_{t+1}) + \Delta\mu_j(S_{t+1}) \quad (6)$$

where $\mu_j(S_{t+1})$ and $\Delta\mu_j(S_{t+1})$ are calculated by equations (3) and (5), respectively.

The information about the moment when the transition starts is considered for adjusting the initial prediction in the equation (6). In the transition between two functional states, the membership degrees of a sample to each state can have similar values, then, there is an uncertainty about the real state of the process. For this purpose, [74] includes an information index $I_D(\mu)$ that takes values in $[0, 1]$. The value 0 states for the situation of minimum certainty and 1 states for the situation of maximum certainty. Equation (7) defines this information index:

$$I_D(\mu) = \frac{\sum_i \lambda_i e^{\lambda_i}}{C\mu_M e^{\mu_M}}, i = 1, 2, ..., m \tag{7}$$

where $\lambda_i = \mu_M - \mu_i(S_t)$, $i = 1, 2...m$, $i \neq M$, $\mu_M = max(\mu_i(S_t))$, and $C = m-1$

Additionally, it is necessary to include the drift $\Lambda\mu_j(S_{t+1})$ of the predicted membership degrees to assess the change in the process trend, which is accomplished using equation (8). The final estimation of the predicted membership degree $\mu_j(S_{t+1})$ takes into account the amplification of the drift $\Lambda\mu_j(S_{t+1})$ according to the information index $I_D(\mu)$, as defined in equation (9):

$$\Lambda\mu_j(S_{t+1}) = \mu_j(S_{t+1}) - \mu_j(S_t) \tag{8}$$

$$\mu_j(S_{t+1}) = \mu_j(S_t) + \frac{1}{I_D(\mu)}\Lambda\mu_j(S_{t+1}) \tag{9}$$

Finally, the estimated class is stated by the class where $\max(\mu_j(S_{t+1}))$ is accomplished.

The values of the membership degrees estimated by the previous equations are constrained to be in the interval $[0, 1]$. A general overview of the proposed monitoring system for predicting functional states is presented in Figure 1, which is developed in two states, as previously mentioned: (i) Off-line stage for supervised training which uses historical data samples for the design of fuzzy classification model, and the calculation of vectors WFT and ΔWFT, (ii) On-line stage for monitoring the incoming samples S_{t-1} and S_t, where the obtained classification fuzzy model is used for estimating the membership degrees of each sample to each class, and for predicting the membership degrees of the future sample S_{t+1} to each class.

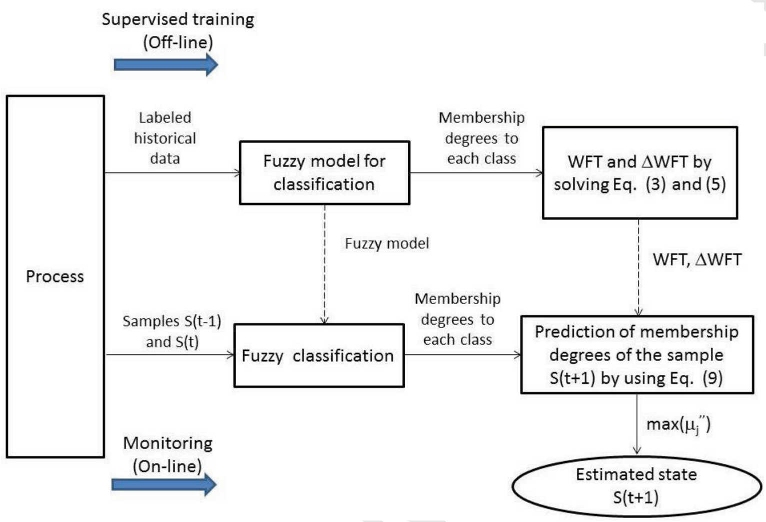


Figure 1: Monitoring system for predicting functional states

3. The proposed fuzzy approach for fault severity prediction

In a rough manner, fault severity is associated with the magnitude of the fault and it can be related to: (i) the physical size of a particular single point defect or failure mode, e.g the size of the break in a gear's tooth, the size of a hole in a ball, inner race or outer race of bearings, or (ii) a general degradation of the entire component. Figure 2 illustrates the case of three severity levels of damage on the outer race of a bearing; the severity level of single failure mode such as the break in a gear's tooth is the case study of this paper, and it is illustrated in section 4. A way to measure the severity levels is through the analysis of the magnitude value of some condition parameter extracted from the vibration signal, for example the Root Mean Square over a time interval [76], or by defining a monotonic degradation assessment index (severity index) as good as possible [77]. Along of the useful life, it is expected that a damage becomes more severe; Figure 3 shows an example of the possible behaviour of a degradation index along the time, in which the higher magnitude of the index means the larger damage in the component. The increase of the index magnitude regarding the time, determines the degradation path of the device.

Figure 2: Severity levels on the outer race of a bearing

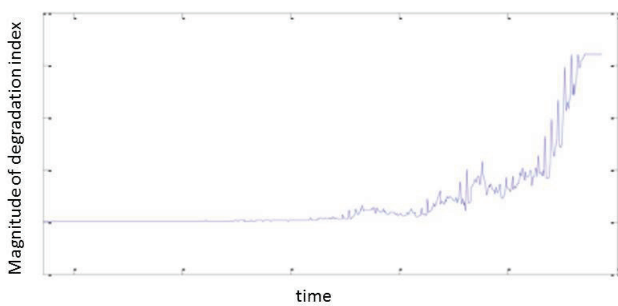


Figure 3: Monothonic degradation behaviour of a fault severity assessment index [77]

The qualitative behaviour of the degradation path, defines different scenarios identifying the degradation speed for which the devices are expected to fail completely. With enough historical information about the degraded physical condition related to some the degradation path, it is possible to predict the future condition state of a device. In this work, we do not aim at proposing a severity index, which in general is not easy to define, but consider a sequence of historical samples that are associated with certain degraded state, from low severity to high severity, to produce a fuzzy model which is able to predict the next state of the device, related to one point in the degradation path in the future. The algorithm in section 2 is applied to estimate the prediction of the fault severity in mechanical rotating devices. For this purpose, in order to apply the algorithm for fault severity evaluation in rotating machinery, we assume the following statements:

1. **Assumption 1.** There is a set of historical samples S_t of length T, e.g. a vibration signal which are measured from the mechanical device at time t, corresponding to different fault severity levels (functional condition) for a single defect point or failure mode. Each sample S_t is associated to a vector $x_t = [x_{t1}, x_{t2}, ..., x_{tp}]$, where $x_{ti}, i = 1, ..., p$, is a condition parameter extracted from the measured signal. Each sample corresponding to certain fault severity level is identified by a class C_i, $i = 1, 2, ...m$, being m the number of the severity levels under study.
2. **Assumption 2.** There is an adequate set of historical succession of samples S_t with a sample time Δt, where each sample corresponds to a fault severity level such as $S_t \in C_i$ and $S_{t+\Delta t} \in C_j$. C_i denotes the identification for the fault severity in level i, and C_j denotes the identification for the fault severity in level j. Due to the nature of fault degradation in mechanical devices, level j is more severe than level i. This succession of samples associated to different severity levels defines certain degradation path using some severity index.

Assumption 1 allows developing a fuzzy classifier in the off-line stage (training stage), to calculate the membership degree of the current sample S_t to the class C_k, in the on-line stage (monitoring stage). The vector x_t is the input feature vector to the fuzzy classifier, and the membership degrees to a class C_k are the outputs. Assumption 2 allows calculating an adequate WFT vector through equations 3 and 5 in the training stage, from the available membership degrees of historical succession of samples S_t and S_{t-1} to the class C_k calculated from the fuzzy classifier. As stated in Assumption 2, the evolution in time established by the degradation path is fitted properly by the WFT when a proper sequence of samples is used in the training stage.

As stated in [75], there are three scenarios for a degradation path measured by a fault severity assessment index: (i) good to gradual (ii) good to accelerated (iii) good to gradual to accelerated. These three scenarios are produced by a sequence in time of samples belonging to different severity levels, and hence, to different severity classes. Next figures illustrate the relationship between the succession of samples and its severity degree in a degradation path. Figure 4 shows the scenarios (ii) and (iii), where part (a) corresponds to the scenario 'good to gradual to accelerated' and part (b) corresponds to the scenario 'good to accelerated'. In case (a), there are three samples between severity level 1 and 2, two sample between severity level 2 and 3, and only one sample up to severity level 3. In case (b), there is one sample for each severity level.

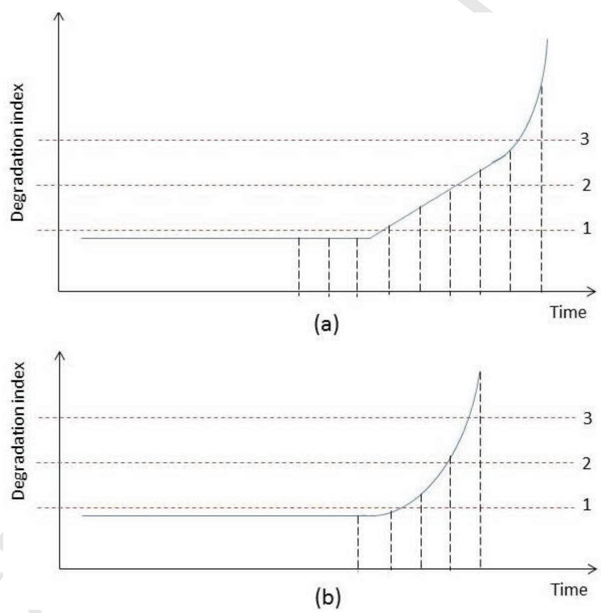


Figure 4: (a) Degradation path 'good to gradual to accelerated' (b) Degradation path 'good to accelerated'

3.1. Fuzzy model for classification

The algorithm in section 2 does not depend on the fuzzy classifier, but the used classifier has to provide a good estimation of the membership degrees of the available samples S_{t-1} and S_t to each class. Recent works developing fuzzy classification models for fault diagnosis in rotating machinery is presented in [22, 78]. In this work, the fuzzy model is designed from a set of historical samples in the form $S = (x, y)$ where $x = (x_1, x_2, .., x_n)$ is the vector of input features and $y = i \in N$ denotes the identifier of the class C_i for identifying the severity level. Each feature is a certain condition parameter calculated from the data points of a vibration signal. More details of the feature extraction process from a sample S of vibration signal will be presented in section 4. We propose two stages for obtaining the final membership function for each severity level.

In the first stage, a Mamdani-type fuzzy model is developed for classifying the severity levels. A Mamdani model is composed by a set of r fuzzy rules R^l, $l = 1, 2, ...r$ in the form of expression 10:

$$R^l : \text{IF } x_1 \text{ is } A_{1l} \text{ AND } x_2 \text{ is } A_{2l} \text{ AND...AND } x_n \text{ is } A_{nl} \text{ THEN } y \text{ is } C^l \quad (10)$$

where x_i, $i = 1, 2, ..., n$, is a input feature of the sample S, y is the output, A_{il} and C^l are fuzzy sets. Gaussian membership functions are assumed for the fuzzy sets in the previous expression 10.

The second stage uses a hierarchical clustering on the set of r fuzzy membership functions for the obtained output fuzzy sets C^l. Each proposed Gaussian membership function is described by its media μ and deviation σ, then, a sample associated to a membership function is the pair (μ, σ). Agglomerative hierarchical clustering is accomplished by linking the samples over sets of observations A and B, such as the criteria $min_{A,B}$ $\{d(a, b) : a \in A, b \in B\}$ is fulfilled. The distance d between samples a and b is calculated by using the Euclidean metric. According to the definition of the output y, the media μ of the estimated fuzzy membership functions for the output fuzzy sets C^l should be around $\mu = i \in N$, and the deviations σ depend on the training dataset. Hierarchical clustering provides the groups of most similar fuzzy membership functions for the output of the trained Mamdani fuzzy model. The centroids of each cluster are taken as the mean μ and the deviation σ of the corresponding membership function describing the set of membership functions in the cluster. The hierarchical clustering is developed with m clusters as the number of classes $C_i, i = 1, ..., m$. Once the crisp value y is proposed by the fuzzy model for classification, the membership degree of the sample S to the class C_i is calculated by evaluating the output y in the set of the Gaussian membership functions obtained by clustering, in the second stage.

3.2. WFT estimation

As the degradation process of mechanical devices is progressive, the next state from a state with a severity level C_i corresponds to a higher severity level C_j, and any state with severity level lower than C_i is not possible. In this sense, the index k in equation 1 is limited by j, this is:

$$\mu_j(S_{t+1}) = \sum_{k=1}^{j} WFT_{kj} * \mu_k(S_t), j = 1, 2, ..., m \tag{11}$$

According to equation (11), the matrix of membership degrees in equation (3) has the form 12:

$$[\mu_k(S_{1...(n-1)})] = \begin{bmatrix} \hat{A}_1 & \hat{A}_2 & ... & \hat{A}_m \end{bmatrix} \tag{12}$$

where:

$$\hat{A}_1 = \begin{vmatrix} \vec{A_1} & \vec{0} & ... & \vec{0} & \vec{0} \\ 0 & \vec{A_1} & ... & \vec{0} & \vec{0} \\ \vdots & \vdots & ... & \vdots & \vdots \\ \vec{0} & \vec{0} & ... & \vec{0} & \vec{A_1} \end{vmatrix} \quad , \quad \hat{A}_2 = \begin{vmatrix} \vec{0} & \vec{0} & ... & \vec{0} & \vec{0} \\ \vec{0} & \vec{A_2} & ... & 0 & \vec{0} \\ \vdots & \vdots & ... & \vdots & \vdots \\ \vec{0} & \vec{0} & ... & \vec{0} & \vec{A_2} \end{vmatrix} \quad , \quad ... \quad , \quad \hat{A}_m = \begin{vmatrix} \vec{0} & \vec{0} & ... & \vec{0} & \vec{0} \\ \vec{0} & \vec{0} & ... & \vec{0} & \vec{0} \\ \vdots & \vdots & ... & \vdots & \vdots \\ \vec{0} & \vec{0} & ... & \vec{0} & A_m \end{vmatrix} \tag{13}$$

and $\vec{0}$ is a vector of zero values, $\vec{A_i}$ is the following vector:

$$\vec{A_i} = \begin{bmatrix} \mu_i(S_1) \\ \mu_i(S_2) \\ \vdots \\ \mu_i(S_{n-1}) \end{bmatrix} \tag{14}$$

The matrix of membership degrees in equation (5) have the similar form of expression (12), and with $\vec{A}_i$ in the following form:

$$\vec{A}_i = \begin{bmatrix} \Delta\mu_i(S_1) \\ \Delta\mu_i(S_2) \\ \vdots \\ \Delta\mu_i(S_{n-1}) \end{bmatrix} \tag{15}$$

3.3. Monitoring and prediction system

430 The monitoring system for fault severity prediction works, in general, as depicted in Figure 1 according to [74]. However, to apply this general framework to the fault severity prediction proposed in this work, some details must be given. The details for performing the training (off-line) and monitoring (on-line) stages are illustrated in Figure 5. A brief explanation about the workflow in Figure 5 is given as follows.

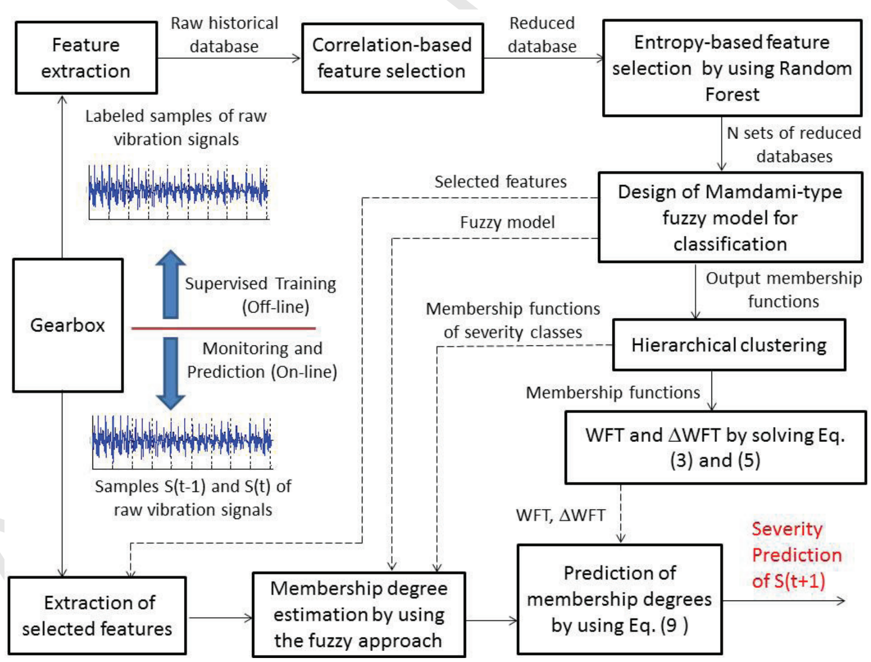


Figure 5: Fuzzy approach for fault severity prediction in helical gears from vibration signal

1. **Training stage**: in this stage, two static models must be created: (i) the fuzzy model for fault severity classification and (ii) the linear model to set the vector WFT and ΔWFT. These models correspond to the outputs of the upper blocks of the supervised training workflow in Figure 1. However, classical approaches to build classifiers for fault severity classification require a proper dataset of features extracted from the raw vibration signal. The proper dataset is built from a set of historical vibration signals, associated to severity levels that occur from the healthy state to the highest severity level, by executing the three first upper blocks of the the supervised training workflow in Figure 5:

    - *Feature extraction*: feature extraction is the process of calculating the condition parameters associated to the vibration signal. The most common condition parameters in time domain, frequency domain and time-frequency are calculated. Then, a dataset with s samples and f features is obtained, each sample is in the form $S = (x, y)$, where x is the vector of features (condition parameters), and y is the corresponding severity class, as stated in **Assumption 1**. Feature extraction will be detailed in section 4.1, according to the case study in this work.
    - *Correlation-based feature selection*: this block performs a classical correlation analysis on the historical dataset obtained in the feature extraction block, to keep the uncorrelated features according to a specified correlation threshold. Usually, this is a previous step after considering other feature selection technique. The correlation analysis in our case study is described in section 4.2.
    - *Entropy-based feature selection based on Random Forest*: a second analysis of the features selected after the correlation analysis is performed in this block by identifying the importance variables, according to the entropy value of each feature; the entropy is calculated through the Random Forest algorithm. After executing this block,

N proper datasets of samples with different reduced set of features, upper to several entropy thresholds, are available to build the fuzzy classifier. The entropy analysis of our case study is presented in section 4.2.

- *Design of Mamdani-type fuzzy model for classification*: in this block, several classic fuzzy models are developed as proposed in section 3.1, which are composed by rules in the form 10. The best fuzzy model regarding its accuracy for fault classification is obtained by analysing the N sets of reduced features obtained from the previous block.
- *Hierarchical clustering*: once the best fuzzy model is obtained with n selected features, hierarchical clustering is applied to the set of membership functions of the output fuzzy sets in each rule, in order to obtain the membership functions of each class. According to proposed in section 3.1, the output of this block are different fuzzy clusters of membership functions as number of severity classes is considered. As final result, given a sample S with a vector of selected features x, the fuzzy model gives the membership degree of S to each severity class described by each cluster.
- *WFT and ΔWFT solving*: vectors WFT and ΔWFT are solved as proposed in section 3.2, through equations (3) and (5), by considering a sequence of samples which is associated with certain degradation path. The degradation path is assumed as the real degradation performance of the machine in time, according to **Assumption 2**. The degradation path associated to certain rotating machinery is verified by experts, after analysing the historical run-to-failure behaviour. In this work, we assume that a degradation path is known and we select the proper sequence of samples associated to that degradation path, as discussed in **Assumption 2**.

The blocks *Design of Mamdani-type fuzzy model for classification* and *Hierarchical clustering* correspond to the block *Fuzzy model for classification* in Figure 1.

2. **Monitoring and Prediction stage**: the prediction system receives two successive samples S_{t-1} and S_t, which are assumed to have the same degradation path for which the prediction system was trained, and the lower blocks in Figure 5 are executed:

   - *Extraction of selected features*: for these samples, only the selected features in the training stage are extracted.
   - *Membership degree estimation*: the membership degrees $\mu_{C_i}(S_{t-1})$ and $\mu_{C_i}(S_t)$, $i = 1, ..., m$, of the two samples to each class are calculated, by using the fuzzy classification model and the output membership functions obtained by hierarchical clustering.
   - *Prediction of the membership degree*: According to the algorithm in section 2, only two successive available samples S_{t-1} and S_t are needed to predict the membership degree of the expected sample S_{t+1} to each class, based on the knowledge of the membership degrees $\mu_{C_i}(S_{t-1})$, $\mu_{C_i}(S_t)$ and vectors WFT, ΔWFT. This prediction is obtained from equation (9).

   The lower blocks in Figure 5 are, essentially, the same blocks in Figure 1, except by the block *Extraction of selected features*, which is needed to extract the input features to the block *Membership degree estimation*.

4. Experimental setup

This section shows the experimental set-up to obtain the dataset for training and testing the prediction capabilities of the proposed approach. All the experiments were carried out in the experimental test bed in Figure 6. The rotation motion of the equipment is generated by a 1.1 kW motor powered by three-phase 220 V at 60 Hz with a nominal speed of 1650 rpm. The torque motion is transmitted into a gearbox, where several gear fault severities are assembled. At the end of the gearbox shaft, the torque is transmitted to a pulley, which is part of the magnetic brake system. The magnetic brake function is to control different loads according to the measurement settings. A variable-frequency drive was used to generate different speeds. The data acquisition system was performed with the NI CompactDAQ-9191 of National Instruments and the module NI 9234, which is inserted in the DAQ slot. This device has a maximum sample frequency of 51.2 kS/s, anti-aliasing filtering, 24-bit resolution, IEPE signal coupling, and Ethernet communication. The data acquisition software was developed in our laboratory on NI LabVIEW environment.

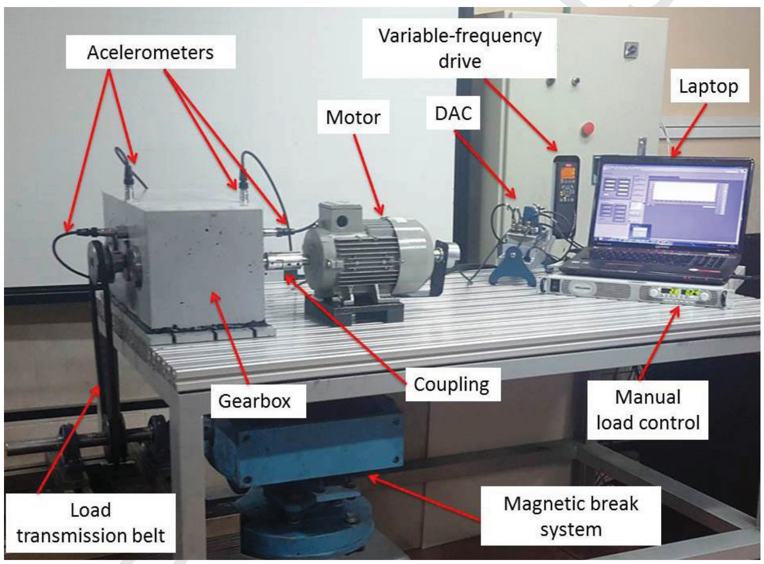


Figure 6: Experimiental test bed

Experimental settings for the signal collection are shown in Table 1 and the gearbox was configured with ten different fault severities of the tooth breaking, i.e. several sizes of the break, including the healthy condition, see Table 2. Severity level C_1 is the healthy or normal condition, and severity level C_2 to C_{10} states for the successive severity levels that simulates the degradation process of the tooth breaking. Figure 7 shows the real helical gear conditions simulated in our test bed.

Table 1: Experimental settings

Parameter	Value
Sampling frequency	50 kHz
Length of each sample	10 sec
Number of samples	5
Rotation Frequency (constant speed)	8 Hz, 12 Hz, 15 Hz
Range Frequency (variable speed)	0-15 Hz sin wave 0.5 Hz
	0-15 Hz square wave 0.5Hz
Load	No Load, 10 V, 30 V

Table 2: Gear fault conditions

Label	**Description**
C1	Healthy gear
C2	Tooth breaking level 1
C3	Tooth breaking level 2
C4	Tooth breaking level 3
C5	Tooth breaking level 4
C6	Tooth breaking level 5
C7	Tooth breaking level 6
C8	Tooth breaking level 7
C9	Tooth breaking level 8
C10	Tooth breaking level 9

4.1. Feature extraction

This section shows the result of the block *Feature extraction* in Figure 5. Vibration signals was collected by using four PBC IEPE accelerometers with a sensitivity of 100 mV/g, and 75 vibration samples are recorded for each severity level. Feature extraction was performed by computing statistical parameters on time domain and frequency domain, such as [79, 80]: mean, standard devia-

Figure 7: Real gear damages with different severity levels

tion, kurtosis, skewness, root mean square, crest factor and kurtosis of Teager's Energy Operator. In time domain, these parameters were computed over the entire vibration raw signal. In frequency domain, the signal was split in 80 bands and some statistical parameters were computed for each band. On the other hand, Wavelet Packet Decomposition (WPD) has been reported as useful technique for extracting features in time-frequency domain [81]. In this work, WPD was used to obtain wavelet coefficients in different levels and proportional energy operator was computed for each coefficient. This feature extraction was accomplished for each accelerometer, as depicted in Figure 8. Then, we have a dataset with 750 samples and 2652 features.

4.2. Feature selection

This section shows the result of the blocks *Correlation-based feature selection* and *Entropy-based feature selection* in Figure 5. Feature selection was accomplished at first by applying a statistical correlation analysis, by identifying the attributes that have correlation values upper to 95%. Attributes under the spec-

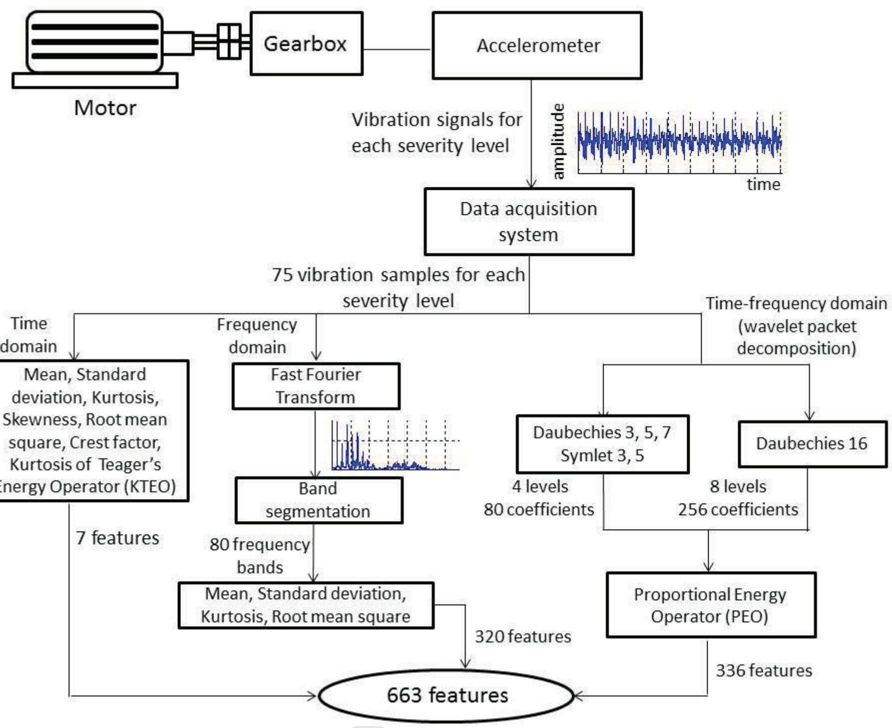


Figure 8: Feature extraction process, one accelerometer

ified threshold are kept as selected features, and the attributes associated with more than one high correlated attributes are also kept as selected features. The rest of attributes are discarded. Following this analysis, only 1083 attributes are selected. In a next step, Random Forest (RF) algorithm was applied to identify the importance variables, in order to reduce the dimensionality of the features vector for the classification problem. RF uses the Shannon entropy to calculate the information gain of each node in the tree, in order to reach an exact separation of classes along the tree. The information gain is defined as $I(S, f) = H(S) - H(S, f)$, where $H(S)$ is the entropy measure over the dataset S before selecting the feature f and $H(S, f)$ is the entropy measure after selecting the attribute f, as shown in equations 16 and 17, respectively [82, 83]:

$$H(S) = -\sum_{c \in C} p(c) log_2 p(c) \tag{16}$$

where c is a class in the set C, and the probability of each class is defined as $p(c) = n_c/N$ where n_c is the number of samples in class c and N is the cardinality of S.

$$H(S, f) = \sum_{v \in F} \frac{|S_v|}{|S|} H(S_v) \tag{17}$$

where v is a value of the attribute f, F is the set of all possible values of f, and $S_v = \{s \in S | f = v\}$

In case of our dataset, Figure 9 shows the normalized entropy values for each feature of the reduced dataset by correlation, with 1083 attributes. There are 4, 15, 44 and 146 important features with normalized entropy values upper to 0,9-0,8-0,7 and 0,6 respectively. These sets of features define four different datasets that will be used for the adjustment of the fuzzy model for classification.

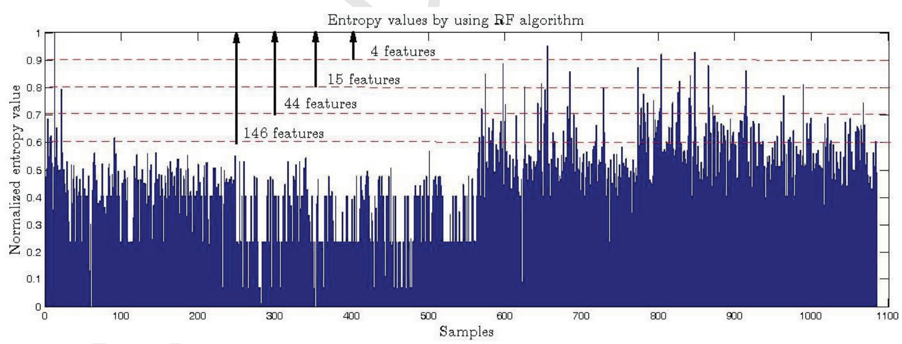


Figure 9: Entropy values using RF associated to each feature

5. Results

This section shows the results of applying the fuzzy approach for fault severity prediction in helical gears, according to the workflow in Figure 5. Results of the training and monitoring stages are presented separately.

5.1. *Training stage*

5.1.1. *Fuzzy model for classification*

As shown in Figure 5, the next step, after applying feature extraction and selection to have a set of proper historical dataset, is to adjust the classification model according to the description in section 3.1. The set of fuzzy rules in the form (10) was obtained by using the fuzzy C-means algorithm as technique for rule extraction from data [84]. Each set of reduced historical dataset defined in section 4.2 was split in training (70%) and test set (30%). Different training experiments was developed with each dataset, under different parameters values of the fuzzy C-means algorithm, such as the fuzziness index and number of clusters. The final set of selected features depends on the classification accuracy of the fuzzy model. After different training experiments, a fuzzy model was adjusted with 800 clusters, the fuzziness index value was set to 1.8 and the number of selected input features were 15, which are described in Table 3.

Table 3: Selected features

Feature	**Domain**	**Sensor**	**Associated equation**
Mean	Time	1	Mean=$\frac{\sum_{i=1}^{N} v_i}{N}$; where v is a vector of vibration data points, $N = \lvert v \rvert$, and v_i is the value of ith data point
PEO of coefficients 2, 10, 26, 48, 64, 140, 174, 246	Time-frequency using wavelet Daubechies 16	4	PEO=$\frac{E_c}{E_s}$, $\sum_{n_c} \text{PEO} = 1$ where E_c is the energy of the wavelet coefficient E_s is the energy of the raw vibration signal, n_c is the number of wavelet coefficients, and $E_{c,s} = N \sum_{i=1}^{N} v_i^2$
PEO of coefficients 76, 153	Time-frequency using wavelet Daubechies 16	2	
PEO of coefficients 131, 158, 204	Time-frequency using wavelet Daubechies 16	3	
PEO of coefficient 160	Time-frequency using wavelet Daubechies 16	1	

The classification model has a RMSE=2,6784E-04 for estimating the class label $C = 1, ..., 10$ in the training phase, and a RMSE=0.0735 with the data in test set. Hierarchical clustering was applied on the set of the output membership functions and ten clusters have been identified, as illustrated in Figure 10. Finally, the centroids of each cluster define the membership function of each severity class, as shown in Figure 11. With these membership functions, we have 100% of accuracy for classification with the training dataset and 87,11% of accuracy with the test dataset.

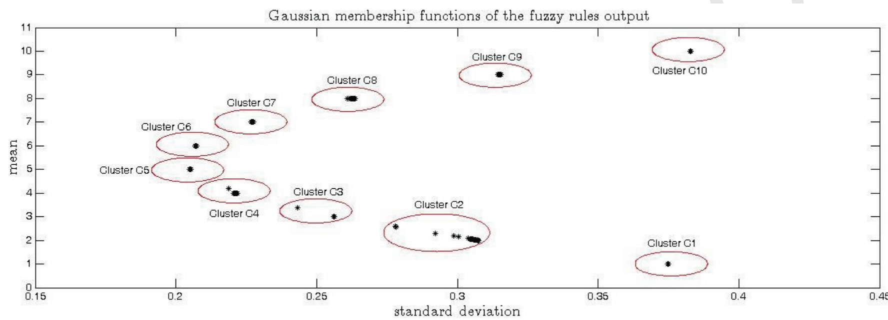


Figure 10: Clusters of the output membership functions

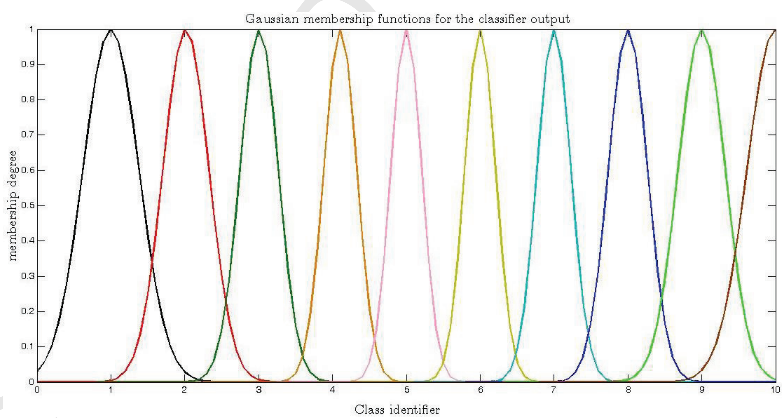


Figure 11: Membership functions of each severity class

5.1.2. WFT solving

According to the workflow in Figure 5, once the fuzzy model for classification has been developed, the next step is the estimation of the vectors WFT and ΔWFT according to equations 3 and 5. Given 10 severity classes, 100 weights for each vector WTF and ΔWFT have to be computed. In order to fit the time evolution of the degradation process, an adequate set of training samples must be used (see **Assumption 2**, section 3). Training data for calculating WFT is strongly dependant on the degradation path that has the degradation process of the mechanical rotating device. This data must fit the real trend in the time. In this sense, the training data is arranged properly according to the severity levels, in order to simulate these two scenarios: (i) good to accelerated path, and (ii) good to gradual to accelerated path. In case (i), 15 successive samples from severity level C1 to C10 was arranged properly, as illustrated in Figure 12. In the same manner, 27 successive samples from severity level C1 to C10 was arranged as illustrated in Figure 13. These samples are taken from the training set.

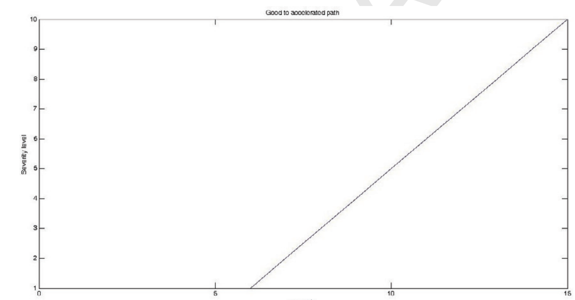


Figure 12: Good to accelerated simulated path

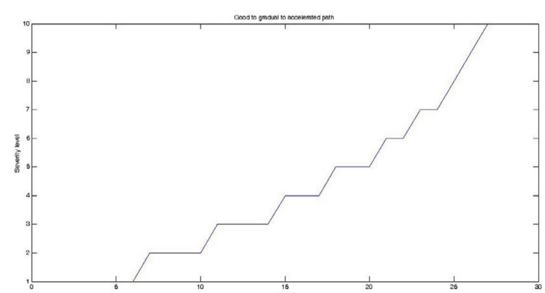


Figure 13: Good to gradual to accelerated simulated path

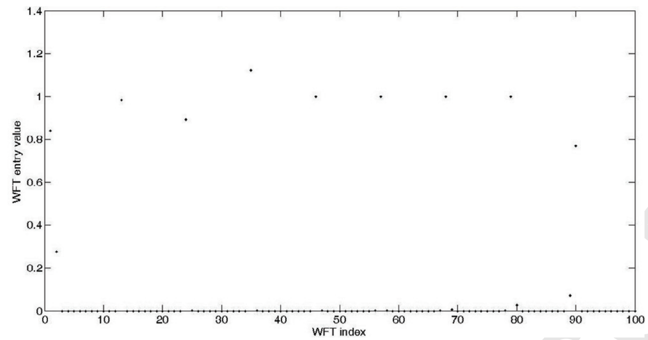
Figure 14 shows the values obtained for the vector WFT in each scenario. A brief analysis of this results shows that in case of the scenario (i), the weight values of the transition from C_i to C_j, being C_j more severe than C_i, are high with regard to the weight values in case of the scenario (ii), as expected. For example, the weight of the transition from C1 to C2 is 0,2767 in case (i) and it is decreased to 0,2228 in case (ii); the weight of remaining in the state C1 is around 0,839 in both cases. On the other hand, the weight of the transition from the state C2 to C3 is 0,9826, and the weight of remaining in the same state C2 is zero, in case (i). Conversely, the weight of remaining in the same state C2 is 0.7453 in case (ii), and the weight of the transition from C2 to C3 decreases to 0,2491, as expected for this degradation path. Finally, note that the transition from the state C_j to C_i is not possible, and the weight values are zero.

5.2. Monitoring and prediction stage

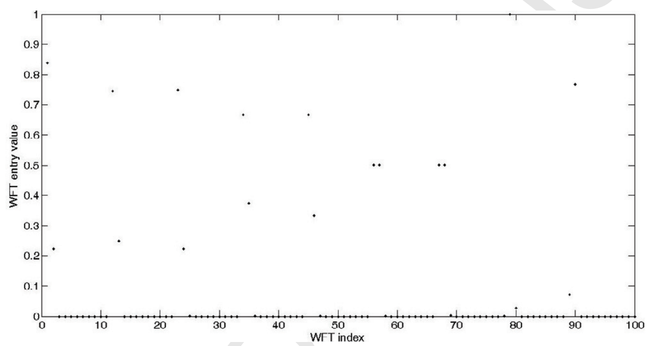
Finally, as proposed by the monitoring system, in the on line stage, we suppose two available samples S_{t-1} and S_t; the memberships degrees of each sample to each class are calculated by using the fuzzy model obtained in section 5.1.1, and we predict the next condition S_{t+1} from the knowledge of the vector WTF and ΔWTF computed in section 5.1.2. Next sections show and discuss the results of each scenario.

5.2.1. Good to accelerated path

In this scenario, 15 successive samples from severity level C1 to C10 was arranged properly. After six samples in C1, the simulated scenario takes one sample per severity level, as illustrated in Figure 12. Once the training phase is accomplished, the linear model is ready for predicting the membership degrees for the next state, given the measurements in the current time t and the previous time $t-1$, in the monitoring stage. For testing the prediction capabilities, the scenario was simulated with data from the test set that have a sequence of samples similar to the training set. Three successive samples were taken for



(a)



(b)

Figure 14: (a) WFT for degradation path, scenario (i) (b) WFT for degradation path, scenario (ii)

testing.

The first case is shown in Table 4, where samples S_{t-1}, S_t and S_{t+1} belong to the true classes C4, C5 and C6. In that case, the fuzzy classifier assigns the correct class of the samples S_{t-1} and S_t. The maximum of membership degree for the sample S_{t+1} given by predictor vector $\mu_{Cj}(S_{t+1})$ is 1, and this prediction match to the maximum of the estimated membership degree given by the fuzzy classifier. The sample S_{t+1} is correctly assigned to severity class C6. This prediction is according to the trained sequence where successive samples belong to successive severity levels.

Other case is presented in Table 5, where the available samples S_{t-1} and S_t belong to class C1, and S_{t+1} belongs to class C2. We compute the prediction

Table 4: Clasification and prediction. Good to accelerated path, case 1

Sample	μ_{C1}	μ_{C2}	μ_{C3}	μ_{C4}	μ_{C5}	μ_{C6}	μ_{C7}	μ_{C8}	μ_{C9}	μ_{C10}
S_{t-1}	1,13E-14	1,12E-09	5,61E-04	0,9000	7,99E-06	7,65E-21	2,46E-38	1,64E-50	3,75E-55	5,90E-54
S_t	1,94E-25	7,20E-21	1,08E-13	2,93E-04	1,0000	8,70E-06	1,66E-17	8,24E-29	1,23E-35	9,26E-38
S_{t+1}	2,38E-39	1,20E-36	5,44E-30	1,29E-16	7,20E-06	1,0000	6,40E-05	3,33E-13	2,33E-20	2,01E-24
$\mu_{Cj}(S_{t+1})$	0	0	0	0	0	1,0000	0,0001	6,55E-10	0	0

for the next sample S_{t+1}. The fuzzy classifier assigns the samples S_{t-1} and S_t to the correct class with membership degrees 0,3386 and 0,6497, respectively. According to the trained sequence the next sample could be in C1, but given two successive samples in this class the expected result could be in the next class C2. The predicted vector $\mu_{Cj}(S_{t+1})$ shows that the predicted class is C2 with a maximum membership degree of $0,6090$. The predicted membership degree to C1 is $0,5547$, then the other expected severity level for the next sample is C1 as previously mentioned. Higher severity levels have membership degrees very close to zero, as expected.

Table 5: Clasification and prediction. Good to accelerated path, case 2

Sample	μ_{C1}	μ_{C2}	μ_{C3}	μ_{C4}	μ_{C5}	μ_{C6}	μ_{C7}	μ_{C8}	μ_{C9}	μ_{C10}
S_{t-1}	0,3386	0,3250	1,20E-07	1,20E-29	6,66E-62	1,07E-100	3,76E-125	2,08E-130	1,00E-121	2,02E-106
S_t	0,6497	0,0977	1,02E-09	1,94E-34	2,44E-69	4,66E-110	1,27E-134	1,00E-138	1,98E-128	1,41E-111
S_{t+1}	0,00054	0,3670	0,1006	8,16E-13	4,66E-34	3,13E-64	2,52E-87	1,25E-96	3,64E-94	4,92E-85
$\mu_{Cj}(S_{t+1})$	0,5547	0,60907	0	0,0002	5,30E-05	2,53E-12	5,97E-10	9,01E-08	6,25E-08	3,35E-06

5.2.2. Good to gradual to accelerated path

In this scenario, 27 successive samples from severity level C1 to C10 was arranged properly, as illustrated in Figure 13. For testing the prediction capabilities, the scenario was simulated with data from the test set, and three successive samples were taken. Table 6 shown the case when the available samples S_{t-1} and S_t belong to class C7. According to the trained sequence, the expected class to the sample S_{t+1} should be C8. The maximum value of the predicted membership degree for the sample S_{t+1} permits assigning the expected severity level correctly in class C8, however the second largest predicted membership degree is for the class C7 and the third one is for the class C9. These results shows it is possible that the next sample remains in the same severity level C7. Finally, the predicted membership degree to the class C9 means that there is still a possibility of evolving to a higher severity, according to the time evolution of the simulated scenario. Note that the evolution to classes C8, C9 and C10 are in the accelerated path, then these results agree with the trained scenario.

Table 6: Clasification and prediction. Good to gradual to accelerated path, case 1

Sample	μ_{C1}	μ_{C2}	μ_{C3}	μ_{C4}	μ_{C5}	μ_{C6}	μ_{C7}	μ_{C8}	μ_{C9}	μ_{C10}
S_{t-1}	2,79E-56	7,23E-57	1,00E-52	9,63E-38	3,12E-21	9,65E-06	0,9999	0,0007	1,75E-09	4,36E-14
S_t	2,71E-56	7,00E-57	9,64E-53	9,28E-38	3,03E-21	9,51E-06	0,9999	0,0007	1,78E-09	4,42E-14
S_{t+1}	2,00E-76	9,15E-82	3,60E-82	6,66E-68	5,09E-47	6,28E-21	6,48E-05	1,0000	0,0066	1,19E-06
$\mu_{Cj}(S_{t+1})$	2,28E-56	1,13E-56	7,22E-53	6,19E-38	2,02E-21	4,75E-06	0,4999	0,5005	0,0033	2,05E-05

Other case is illustrated in Table 7. Samples S_{t-1} and S_t belong to class C2, and sample S_{t+1} belongs to class C3. These classes are correctly assigned by the fuzzy classifier. The predicted vector of membership degrees μ_{Cj} shows that sample S_{t+1} belongs to class the class C2. Note that according to the simulated scenario, there are four successive samples in the severity class C2, then, in case of having two previous samples in the class C2, the predicted class C2 can occur for sample S_{t+1}. The second highest predicted membership degree is for the class C3, and this could be expected according to the trained scenario. This case is slowly modified in Table 8 in which samples S_{t-1} and S_t belong to class C2 and C3, respectively. The predicted class given by the predicted vector of membership degrees μ_{Cj} is C3, with a large value of membership degree. This is according to the trained scenario in which class C3 belongs to the gradual path with four samples after having the previous sample in the severity level C2.

Table 7: Clasification and prediction. Good to gradual to accelerated path, case 2

Sample	μ_{C1}	μ_{C2}	μ_{C3}	μ_{C4}	μ_{C5}	μ_{C6}	μ_{C7}	μ_{C8}	μ_{C9}	μ_{C10}
S_{t-1}	2,00E-02	0,9945	0,0010	1,72E-19	1,52E-45	1,29E-79	1,71E-103	3,40E-111	4,03E-106	2,35E-94
S_t	5,37E-04	0,3670	0,1006	8,16E-13	4,66E-34	3,13E-64	2,52E-87	1,25E-96	3,64E-94	4,92E-85
S_{t+1}	6,76E-07	0,0066	0,9994	3,63E-06	2,43E-21	3,00E-46	7,51E-68	9,47E-79	2,74E-79	2,52E-73
$\mu_{Cj}(S_{t+1})$	0,0004	0,2698	0,1694	0,0233	3,91E-08	5,40E-19	0,00	0,00	0,00	0,00

Table 8: Clasification and prediction. Good to gradual to accelerated path, case 3

Sample	μ_{C1}	μ_{C2}	μ_{C3}	μ_{C4}	μ_{C5}	μ_{C6}	μ_{C7}	μ_{C8}	μ_{C9}	μ_{C10}
S_{t-1}	0,0005	0,367	0,1006	8,16E-13	4,66E-34	3,13E-64	2,52E-87	1,25E-96	3,64E-94	4,92E-85
S_t	6,76E-07	0,0066	0,9994	3,63E-06	2,43E-21	3,00E-46	7,51E-68	9,47E-79	2,74E-79	2,52E-73
S_{t+1}	6,98E-07	0,0067	0,9991	3,45E-06	2,18E-21	2,56E-46	6,31E-68	8,05E-79	2,39E-79	2,26E-73
$\mu_{Cj}(S_{t+1})$	5,68E-07	0,0048	0,7494	0,2231	1,74E-06	2,31E-12	3,76E-05	0,0005	0,0004	1,16E-05

6. Conclusions

This paper applies a fuzzy model based approach to fault severity prediction. The approach estimates the Weights of Fuzzy Transitions (WFT) in order to predict the membership degrees to the defined functional states, given two previously monitored samples. In our case, the functional states are related to different fault severity levels of one failure mode, that is tooth breaking in helical gears.

The approach does not depend on the fuzzy model to estimate the membership degrees to each severity class, and it is assumed that a good fuzzy model is available. This work combines a Mamdani-type fuzzy model and hierarchical clustering, in order to estimate the membership functions of each severity class. Results show adequate RMSE values in the assignment of the correct class according to the estimated membership degrees.

The approach aims at incorporating the time evolution, then, the adjustment of the linear equations to calculate the WTF and ΔWFT is highly dependant on the temporal behaviour of the training set in the off line stage. In our case, due to the nature of the simulated failure mode and its expected time evolution, two well known degradation paths in rotating machinery have been constructed.

The results shows that the approach was able to predict the next expected severity level, according to the behaviour of the two scenarios of degradation path. This approach is simple and easy for on line implementation, which makes it a possible condition monitoring system for rotating machinery.

Acknowledgements

This work was sponsored by The Ministry of Higher Education, Science, Technology and Innovation (SENESCYT) of the Republic of Ecuador and the Prometeo program, Authors thank their support. We also want to express our thanks to the GIDTEC research group of the Universidad Politécnica Salesiana for supporting the accomplishment of this research.

References

[1] R. B. Randall, Vibration-Based Condition Monitoring, Wiley, West Sussex, United Kingdom, 2011.

[2] M. Feldman, Hilbert transform in vibration analysis, Mechanical Systems and Signal Processing 25 (3) (2011) 735 – 802.

[3] Y. Lei, J. Lin, Z. He, M. J. Zuo, A review on empirical mode decomposition in fault diagnosis of rotating machinery, Mechanical Systems and Signal Processing 35 (1–2) (2013) 108 – 126.

[4] Z. Feng, M. Liang, F. Chu, Recent advances in time-frequency analysis methods for machinery fault diagnosis: A review with application examples, Mechanical Systems and Signal Processing 38 (1) (2013) 165 – 205.

[5] J. Chen, Z. Li, J. Pan, G. Chen, Y. Zi, J. Yuan, B. Chen, Z. He, Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review, Mechanical Systems and Signal Processing 70–71 (2016) 1 – 35.

[6] A. Rai, S. Upadhyay, A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings, Tribology International 96 (2016) 289 – 306.

[7] B.-S. Yang, D.-S. Lim, A. C. C. Tan, VIBEX: an expert system for vibration fault diagnosis of rotating machinery using decision tree and decision table, Expert Systems with Applications 28 (4) (2005) 735 – 742.

[8] S. Ebersbach, Z. Peng, Expert system development for vibration analysis in machine condition monitoring, Expert Systems with Applications 34 (1) (2008) 291 – 299.

[9] Y. Lei, Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery, Butterworth-Heinemann (Elsevier), Oxford, United Kingdom, 2016.

[10] Z. Zhang, Y. Wang, K. Wang, Intelligent fault diagnosis and prognosis approach for rotating machinery integrating wavelet transform, principal component analysis, and artificial neural networks, The International Journal of Advanced Manufacturing Technology 68 (1) (2013) 763–773.

[11] J. Chu, Y. Niu, A novel hybrid intelligent method for fault diagnosis of the complex system, International Journal of Hybrid Information Technology 9 (3) (2016) 331–340.

[12] L. Yi, Study on a novel hybrid intelligent fault diagnosis method based on improved DE and RBFNN, International Journal of Database Theory and Application 9 (8) (2016) 159–170.

[13] F. Jia, Y. Lei, J. Lin, X. Zhou, N. Lu, Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data, Mechanical Systems and Signal Processing 72–73 (2016) 303 – 315.

[14] C. Li, R.-V. Sánchez, G. Zurita, M. Cerrada, D. Cabrera, Fault diagnosis for rotating machinery using vibration measurement deep statistical feature learning, Sensors 16 (6) (2016) 895.

[15] X. Zhang, W. Chen, B. Wang, X. Chen, Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization, Neurocomputing 167 (2015) 260 – 279.

[16] X. Zhang, B. Wang, X. Chen, Intelligent fault diagnosis of roller bearings with multivariable ensemble-based incremental support vector machine, Knowledge-Based Systems 89 (2015) 56 – 85.

[17] Z. Liu, X. Chen, Z. He, Z. Shen, LMD method and multi-class RWSVM of fault diagnosis for rotating machinery using condition monitoring information, Sensors 13 (7) (2013) 8679.

[18] Z. Liu, W. Guo, J. Hu, W. Ma, A hybrid intelligent multi-fault detection method for rotating machinery based on RSGWPT, KPCA and twin SVM, ISA Transactions (2016) –doi:http://dx.doi.org/10.1016/j.isatra.2016.11.001.

[19] Y. Lei, Z. He, Y. Zi, X. Chen, New clustering algorithm-based fault diagnosis using compensation distance evaluation technique, Mechanical Systems and Signal Processing 22 (2) (2008) 419 – 435.

[20] Y. Lei, Z. He, Y. Zi, Q. Hu, Fault diagnosis of rotating machinery based on a new hybrid clustering algorithm, The International Journal of Advanced Manufacturing Technology 35 (9) (2008) 968–977.

[21] C. Li, J. Valente de Oliveira, R.-V. Sánchez, M. Cerrada, G. Zurita, D. Cabrera, Fuzzy determination of informative frequency band for bearing fault detection, Journal of Intelligent & Fuzzy Systems 30 (6) (2016) 3513–3525.

[22] C. Li, J. Valente de Oliveira, M. Cerrada, F. Pacheco, D. Cabrera, V. Sanchez, G. Zurita, Observer-biased bearing condition monitoring: From fault detection to multi-fault classification, Engineering Applications of Artificial Intelligence 50 (2016) 287 – 301.

[23] A. S. Raj, N. Murali, Early classification of bearing faults using morphological operators and fuzzy inference, IEEE Transactions on Industrial Electronics 60 (2) (2013) 567–574.

[24] F. Harrouche, A. Felkaoui, Automation of fault diagnosis of bearing by application of fuzzy inference system (FIS), Mechanics & Industry 15 (6) (2014) 477–485.

[25] H. Wang, P. Chen, Fuzzy diagnosis method for rotating machinery in variable rotating speed, IEEE Sensors Journal 11 (1) (2011) 23–34.

[26] H. Sun, K. Li, P. Chen, H. Wang, X. Ping, Y. Cao, A sequential fuzzy diagnosis method for rotating machinery using ant colony optimization and possibility theory, Journal of Mechanical Science and Technology 28 (4) (2014) 1189–1201.

[27] J. Yan, L. Lu, D. Zhao, G. Wang, Diagnosis of bearing incipient faults using fuzzy logic based methodology, in: Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference on, Vol. 3, IEEE, 2010, pp. 1229–1233.

[28] L. Zhang, G. Xiong, H. Liu, H. Zou, W. Guo, Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference, Expert Systems with Applications 37 (8) (2010) 6077 – 6085.

[29] J. Latuny, R. Entwistle, A bearing fault classifier using artificial neuro-fuzzy inference system (ANFIS) based on statistical parameters and daubechies wavelet transform features, in: Advances in Applied Mechanics Research, Conference Proceedings - 7th Australasian Congress on Applied Mechanics, ACAM 2012, 2012, pp. 165–174.

[30] H. M. Ertunc, H. Ocak, C. Aliustaoglu, ANN and ANFIS-based multi-staged decision algorithm for the detection and diagnosis of bearing faults, Neural Computing and Applications 22 (1) (2013) 435–446.

[31] J. Zhang, W. Ma, L. Ma, A fault diagnosis method based on ANFIS and bearing fault diagnosis, in: Proceedings - 2014 International Conference on Information Science, Electronics and Electrical Engineering, ISEEE 2014, Vol. 2, 2014, pp. 1274–1278.

[32] K. Li, P. Chen, S. Wang, An intelligent diagnosis method for rotating machinery using least squares mapping and a fuzzy neural network, Sensors (Switzerland) 12 (5) (2012) 5919–5939.

[33] G. Marichal, M. Artés, J. G. Prada, O. Casanova, Extraction of rules for faulty bearing classification by a neuro-fuzzy approach, Mechanical Systems and Signal Processing 25 (6) (2011) 2073 – 2082.

[34] E. Zio, G. Gola, A neuro-fuzzy technique for fault diagnosis and its application to rotating machinery, Reliability Engineering & System Safety 94 (1) (2009) 78 – 88.

[35] D. Cabrera, F. Sancho, R. V. Sánchez, G. Zurita, M. Cerrada, C. Li, R. E. Vásquez, Fault diagnosis of spur gearbox based on random forest and wavelet packet decomposition, Frontiers of Mechanical Engineering 10 (3) (2015) 277–286.

[36] M. Cerrada, G. Zurita, D. Cabrera, R. V. Sánchez, M. Artés, C. Li, Fault diagnosis in spur gears based on genetic algorithm and random forest, Mechanical Systems and Signal Processing 70–71 (2016) 87 – 103.

[37] M. Cerrada, R. V. Sánchez, D. Cabrera, G. Zurita, C. Li, Multi-stage feature selection by using genetic algorithms for fault diagnosis in gearboxes based on vibration signal, Sensors 15 (9) (2015) 23903–23926.

[38] F. Pacheco, J. V. de Oliveira, R.-V. Sánchez, M. Cerrada, D. Cabrera, C. Li, G. Zurita, M. Artés, A statistical comparison of neuroclassifiers and feature selection methods for gearbox fault diagnosis under realistic conditions, Neurocomputing 194 (2016) 192 – 206.

[39] M. Cerrada, R. V. Sánchez, F. Pacheco, D. Cabrera, G. Zurita, C. Li, Hierarchical feature selection based on relative dependency for gear fault diagnosis, Applied Intelligence 44 (3) (2015) 687–703.

[40] C. Li, R. V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, R. E. Vásquez, Multimodal deep support vector classification with homologous features

and its application to gearbox fault diagnosis, Neurocomputing 168 (2015) 119 – 127.

[41] G. N. Marichal, M. L. Del Castillo, J. López, I. Padrón, M. Artés, An artificial intelligence approach for gears diagnostics in AUVs, Sensors 16 (4) (2016) 529.

[42] Y. Lei, M. J. Zuo, Z. He, Y. Zi, A multidimensional hybrid intelligent method for gear fault diagnosis, Expert Systems with Applications 37 (2) (2010) 1419 – 1430.

[43] Y. Lei, M. J. Zuo, Gear crack level identification based on weighted K nearest neighbor classification algorithm, Mechanical Systems and Signal Processing 23 (5) (2009) 1535 – 1547.

[44] D. Wang, K nearest neighbors based methods for identification of different gear crack levels under different motor speeds and loads: Revisited, Mechanical Systems and Signal Processing 70–71 (2016) 201 – 208.

[45] Z. Li, W. Ding, A novel fault diagnosis method for gear transmission systems using combined detection technologies, Research Journal of Applied Sciences, Engineering and Technology 6 (18) (2013) 3354–3358.

[46] L. Gao, Z. Ren, W. Tang, H. Wang, P. Chen, Intelligent gearbox diagnosis methods based on SVM, wavelet lifting and RBR, Sensors 10 (5) (2010) 4602–4621.

[47] Z. Liu, W. Guo, Z. Tang, Y. Chen, Multi-sensor data fusion using a relevance vector machine based on an ant colony for gearbox fault detection, Sensors 15 (9) (2015) 21857–21875.

[48] J.-H. Zhong, P. K. Wong, Z.-X. Yang, Simultaneous-fault diagnosis of gearboxes using probabilistic committee machine, Sensors 16 (2) (2016) 185–203.

[49] C. Rajeswari, B. Sathiyabhama, S. Devendiran, K. Manivannan, A gear fault identification using wavelet transform, rough set based GA, ANN and C4.5 algorithm, Procedia Engineering 97 (2014) 1831 – 1841.

[50] S. Gupta, S. K. Abraham, V. Sugumaran, M. Amarnath, Fault diagnostics of a gearbox via acoustic signal using wavelet features, J48 decision tree and random tree classifier, Indian Journal of Science and Technology 9 (33).

[51] C. Li, R.-V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, R. E. Vásquez, Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals, Mechanical Systems and Signal Processing 76–77 (2016) 283 – 293.

[52] T. Toutountzakis, C. K. Tan, D. Mba, Application of acoustic emission to seeded gear fault detection, {NDT} and E International 38 (1) (2005) 27 – 36.

[53] A. Heng, S. Zhang, A. C. Tan, J. Mathew, Rotating machinery prognostics: State of the art, challenges and opportunities, Mechanical Systems and Signal Processing 23 (3) (2009) 724 – 739.

[54] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, D. Siegel, Prognostics and health management design for rotary machinery systems—reviews, methodology and applications, Mechanical Systems and Signal Processing 42 (1–2) (2014) 314 – 334.

[55] I. El-Thalji, E. Jantunen, A summary of fault modelling and predictive health monitoring of rolling element bearings, Mechanical Systems and Signal Processing 60–61 (2015) 252 – 272.

[56] M. Jouin, R. Gouriveau, D. Hissel, M.-C. Péra, N. Zerhouni, Particle filter-based prognostics: Review, discussion and perspectives, Mechanical Systems and Signal Processing 72–73 (2016) 2 – 31.

[57] M. Orchard, G. Kacprzynski, K. G. ebel, B. Saha, G. Vachtsevanos, Advances in uncertainty representation and management for particle filtering

applied to prognostics, in: International Conference on Prognostics and Health Management, IEEE, 2008, pp. 1–6.

[58] E. Myötyri, U. Pulkkinen, K. Simola, Application of stochastic filtering for lifetime prediction, Reliability Engineering and System Safety 91 (2) (2006) 200 – 208.

[59] A. Lorton, M. Fouladirad, A. Grall, A methodology for probabilistic model-based prognosis, European Journal of Operational Research 225 (3) (2013) 443 – 454.

[60] M. Gobbato, J. B. Kosmatka, J. P. Conte, A recursive bayesian approach for fatigue damage prognosis: An experimental validation at the reliability component level, Mechanical Systems and Signal Processing 45 (2) (2014) 448 – 467.

[61] J. Yu, Machine health prognostics using the bayesian-inference-based probabilistic indication and high-order particle filtering framework, Journal of Sound and Vibration 358 (2015) 97 – 110.

[62] Z. Shen, Z. He, X. Chen, C. Sun, Z. Liu, A monotonic degradation assessment index of rolling bearings using fuzzy support vector data description and running time, Sensors 12 (8) (2012) 10109–10135.

[63] M. Amar, I. Gondal, C. Wilson, Fuzzy logic inspired bearing fault-model membership estimation, in: 2013 IEEE Eighth International Conference on Intelligent Sensors, Sensor Networks and Information Processing, IEEE, 2013, pp. 420 – 425.

[64] W. He, Q. Miao, M. Azarian, M. Pecht, Health monitoring of cooling fan bearings based on wavelet filter, Mechanical Systems and Signal Processing 64–65 (2015) 149 – 161.

[65] P. Dempsey, A. Afjeh, Integrating oil debris and vibration gear damage detection technologies using fuzzy logic, Journal of the American Helicopter Society 49 (2) (2004) 109–116.

[66] W. Wang, J. Vrbanek, An evolving fuzzy predictor for industrial applications, IEEE Transactions on Fuzzy Systems 16 (6) (2008) 1439 – 1449.

[67] W. Q. Wang, M. Golnaraghi, F. Ismail, Prognosis of machine health condition using neuro-fuzzy systems, Mechanical Systems and Signal Processing 18 (4) (2004) 813 – 831.

[68] C. Chen, G. Vachtsevanos, M. E. Orchard, Machine remaining useful life prediction: An integrated adaptive neuro-fuzzy and high-order particle filtering approach, Mechanical Systems and Signal Processing 28 (2012) 597 – 607.

[69] C. Chen, B. Zhang, G. Vachtsevanos, Prediction of machine health condition using neuro-fuzzy and bayesian algorithms, Instrumentation and Measurement, IEEE Transactions on 61 (2) (2012) 297–306.

[70] L. Zadeh, Probability measures of fuzzy events, Journal of Mathematical Analysis and Applications 23 (2) (1968) 421 – 427.

[71] M. Beer, A summary on fuzzy probability theory, in: 2010 IEEE International Conference on Granular Computing (GrC), IEEE, 2010, pp. 5–6.

[72] R. Du, K. Yeung, Fuzzy transition probability: a new method for monitoring progressive faults. part 1: the theory, Engineering Applications of Artificial Intelligence 17 (5) (2004) 457 – 467.

[73] R. Du, K. Yeung, Fuzzy transition probability: A new method for monitoring progressive faults. part 2: Application examples, Engineering Applications of Artificial Intelligence 19 (2) (2006) 145 – 155.

[74] C. V. Isaza, H. O. Sarmiento, T. Kempowsky-Hamon, M.-V. LeLann, Situation prediction based on fuzzy clustering for industrial complex processes, Information Sciences 279 (2014) 785 – 804.

[75] C. K. R. Lim, D. Mba, Switching Kalman filter for failure prognostic, Mechanical Systems and Signal Processing 52–53 (2015) 426 – 435.

[76] W. Yang, R. Court, Experimental study on the optimum time for conducting bearing maintenance, Measurement 46 (8) (2013) 2781 – 2791.

[77] Z. Shen, Z. He, X. Chen, C. Sun, Z. Liu, A monotonic degradation assessment index of rolling bearings using fuzzy support vector data description and running time, Sensors 12 (8) (2012) 10109–10135.

[78] J. Hang, J. Zhang, M. Cheng, Application of multi-class fuzzy support vector machine classifier for fault diagnosis of wind turbine, Fuzzy Sets and Systems 297 (2016) 128 – 140.

[79] Y. Lei, Z. He, Y. Zi, A new approach to intelligent fault diagnosis of rotating machinery, Expert Systems with Applications 35 (4) (2008) 1593 – 1600.

[80] Y. Qu, B. Van Hecke, D. He, J. Yoon, E. Bechhoefer, J. Zhu, Gearbox fault diagnostics using AE sensors with low sampling rate, Journal of Acoustic Emission 31 (2013) 67–90.

[81] R. Yan, R. X. Gao, X. Chen, Wavelets for fault diagnosis of rotary machines: A review with applications, Signal Processing 96, Part A (2014) 1 – 15, time-frequency methods for condition based maintenance and modal analysis.

[82] R. Genuer, J.-M. Poggi, C. Tuleau-Malot, Variable selection using random forests, Pattern Recognition Letters 31 (14) (2010) 2225 – 2236.

[83] K. Gold, A. Petrosino, Using information gain to build meaningful decision forests for multilabel classification, in: 2010 IEEE 9th International Conference on Development and Learning, IEEE, 2010, pp. 58 – 63.

[84] S. Guillaume, Designing fuzzy inference systems from data: An interpretability-oriented review, IEEE Transactions on Fuzzy Systems 9 (3) (2001) 426–443.