Prediction of bearing failures by the analysis of the time series

Abdenour Soualhi, Kamal Medjaher, Guy Clerc, Hubert Razik

Prediction of bearing failures by the analysis of the time series

Abdenour SOUALHI[a], Kamal MEDJAHER[b], Guy CELRC[c] and Hubert RAZIK[c]

[a]*Univ Lyon, UJM-Saint Etienne, Saint-Etienne, France*
[b]*ENIT - LGP, Tarbes, France*
[c]*Univ Lyon, Univ Lyon-1, Laboratoire Ampère, France*

Abstract

As they are a part of the energy transmission chain, bearings are considered as important mechanical components in rotating machines. This importance requires a special attention in order to avoid expensive production shutdown due to the appearance of failures. It is therefore necessary to anticipate the appearance of faults by implementing an appropriate prediction model. There exist in the literature several examples of prediction models able to estimate the remaining useful life (RUL) of bearings. These models are based on the principle of the long-term prediction without considering the degradation state of bearings or by defining a degradation threshold arbitrarily. It should be noted that, in addition to a reliable prediction model, identifying the degradation states is an important parameter in the estimation of the RUL. The presented paper proposes a particular approach based on the artificial intelligence (AI) principle. The proposed approach is composed of two model-based AI inspired by the reasoning of the human for the RUL estimation. This reasoning is modeled via the neural networks for time-series prediction. These models are called the adaptive-neuro fuzzy inference system and the neo-fuzzy neuron. The proposed approach is also composed by a third model-based AI. This model is inspired by the behavior of ants to identify the different degradation states of bearings. The combination of these two types of AI provides reliable and robust prediction results.

Keywords:Degradation states, fault prognosis, neural networks, industrial systems, pattern recognition, bearing, artificial intelligence.

1. Introduction

Rotating machines are electromechanical systems widely used in industrial sectors because of their low cost, performance and robustness. However, the environment in which they operate as well as their mode of use generates early failures. These failures are essentially mechanical failures with a rate of 42% due to bearings failure [1], [2]. These require an online monitoring tool able to anticipate the appearance of a fault before their breakdown. This anticipation is translated by the term "prognosis". The prognosis of bearings failure is an approach used to guarantee the quality and performance of the rotating machines. From the literature, it is possible to find several methods for the prediction of the degradation processes of systems. Categorizing these methods in classes is quite difficult because of a wide variety of applications. However, Vachtsevanos et al., in [3], classified prognostic methods into three main categories, prognosis based on: (i) experience, (ii) data, and (iii) model [4], [5], [6], [7], [8], [9], [10]. However, it is possible to integrate experience-based prognostic into data-driven prognosis since it handles system data. Hence, a classification of prognostic methods provided by [11] is shown in Fig. 1. A comparison was made between these categories in [12]. It appears from this comparison that the data-driven approach is suitable when the model-based approach fails in extracting a mathematical representation of the system's degradation process.

Data-driven methods are based on the analysis of features (health indicators) to predict the degradation process of the system. Most of these methods are based on the artificial neural networks (ANNs). We can cite for example the adaptive prognostic method proposed and applied to lithium-ion batteries in order to forecast their end of discharge [13]. We can also cite the research of [14] where authors developed a method based on the ANN for component health condition prognosis which is a feed forward NN model. The inputs of the ANNs are the component age values and the condition monitoring measurements at the current and previous inspection points. ANNs are modeled by a multi-input-multi-output function expressed as $y = f(x,\Psi)$ where Ψ is the parameter vector, x is the input vector and y is the output. One must underline the triple use of ANNs: feature estimation, state prediction, and the remaining useful life (RUL) estimation. RUL estimation will be retained in the rest of the paper. ANNs architecture consists of multiple layers with a usual configuration of a layer of input nodes, one or more layers of hidden nodes and an output layer of input nodes connected by weights [15].

Corresponding author

Email addresses: abdenour.soualhi@univ-st-etienne.fr, kamal.medjaher@enit.fr, guy.clerc@univ-lyon1.fr, hubert.razik@univ-lyon1.fr

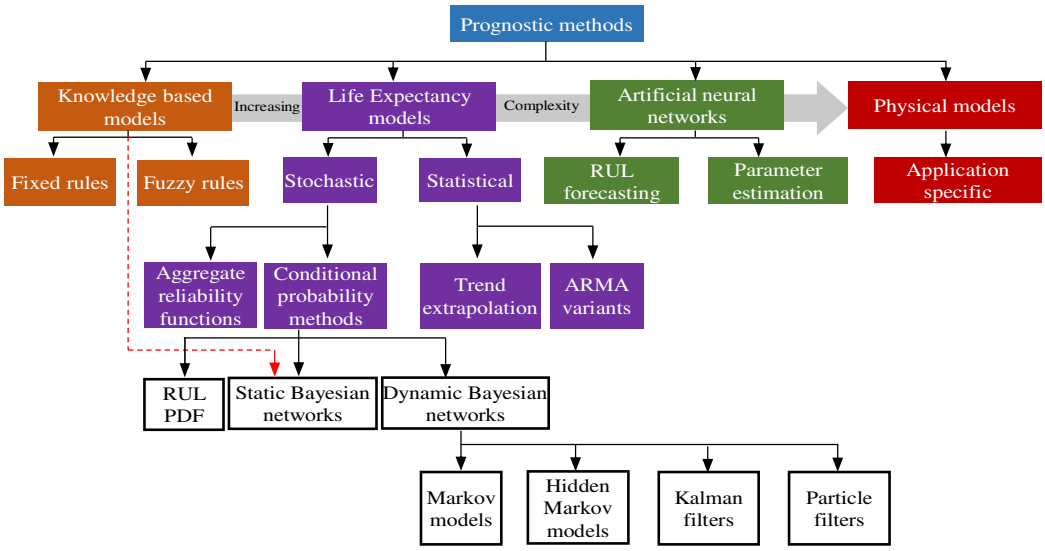

Fig. 1. Prognostic methods classification.

As we can see, the use of ANNs for the prognosis is not new. In addition, we can find in the literature several papers dealing with the problem of bearings fault prediction [4], [16], [17], [18]. The proposed paper demonstrates that it is possible to use a new approach based on a simple architecture of a neural networks coupled with a technique based on the principle of the one-step and multi-step ahead prediction to obtain a reliable estimation of the RUL. To demonstrate this, we will propose two types of neural networks: the adaptive-neuro fuzzy inference system (ANFIS) and the neo-fuzzy neuron (NFN). The RUL prediction is performed by relevant features indicating the degradation of bearings. The historical database is then analyzed in the form of time series in order to train the prediction models. The proposed approach also uses the artificial ants algorithm to identify the degradation states required to calculate the RUL. It is a new unsupervised classification method adapted to our problem of degradation state identification. A flowchart summarizing steps used in our approach is shown in Fig. 2. The artificial ants algorithm is used to analyze the historical degradation database in order to identify the different degradation states of bearings. The use of the degradation states combined with the time series prediction allows estimation of RUL with low errors.

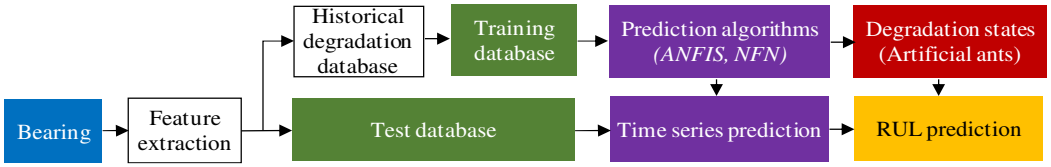

Fig. 2. Proposed approach for the prediction of bearings failure.

The two prediction methods presented in this paper allow estimation of RUL before the occurrence of a failure, and thus the prediction of a defect in the bearing. The information provided by these two methods helps technicians to make a decision about the system's reparation or its replacement before it fails. They acquire the variations of the monitored features in the time and analyze the trending in real time to give a prediction about the appearance of a defect. The International Standardization Organization (ISO) 13381-1 defines the prognosis as an estimation of the remaining life before a fault occurrence. Research conducted by [19] showed that it is relatively difficult to predict the remaining life of bearings without identifying its different degradation states. Indeed, an inherent limitation of a data-driven prediction model is its ability to predict in the long term. In other words, the long-term prediction error is very high. To correct this problem, it is necessary to define a threshold from which the prediction model will be able to give a reliable estimation of the remaining life of the bearing. This estimation is given by predicting the trend

(evolution) of the monitoring features through the prediction of the time-series (see Fig. 3). As we can see, it is more relevant to predict the evolution of the monitoring features based on a degradation threshold than starting the prediction from time $t = 0$. The difference between the time instant (t), where the prognosis begins, and the instant (t + RUL), where the fault appears allows estimation of the RUL.

At this stage, we considered that to estimate the RUL of a bearing, we need to do the following:

- Detect the different degradation states in order to limit the prediction errors;
- Propose a prediction model based on the trend analysis of the monitored features to estimate future observations.

We will begin by presenting the selected prediction models based on artificial intelligence (AI) and then we detail the proposed algorithm for the identification of the degradation states of the bearing.

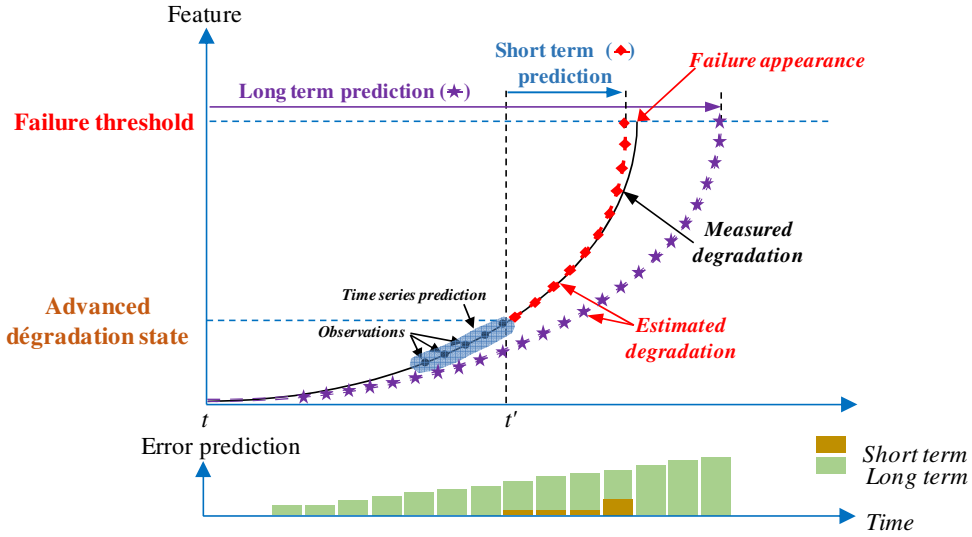

Fig. 3. Time series prediction strategy for the estimation of an degradation process.

2. Prediction models-based on the AI

Prediction by an analysis of the time series is a technique using a mathematical model constructed from the time evolution of observations (measurements) to predict the evolution of a monitored feature up to a horizon of prediction noted (p). These observations are noted respectively: $\{x_t, x_{t-r}, x_{t-2r}, x_{t-3r}\ldots x_{t-(n-1)r}\}$ where (r) represents the acquisition step and (n) represents the number of observations of the time series. An observation noted x_t corresponds to a real value of a feature (health indicator, see table 3) extracted from a signal (vibration signal in our case) recorded at time (t).

Two approaches can be adopted for the construction of a prediction model:

- The first approach proposes to predict directly the value of the observation $\hat{x}_{t+p}$ up to the horizon of prediction (p) from the series of observations $\{x_t, x_{t-r}, x_{t-2r}, x_{t-3r}\ldots x_{t-(n-1)r}\}$ (see Fig. 4). The relationship between the time-series and the output of the prediction model is expressed as follows :

$$\hat{x}_{t+p} = f\left(x_t, x_{t-r}, x_{t-2r}, x_{t-3r}, \ldots, x_{t-(n-1)r}\right) \quad (1)$$

In terms of complexity, a training for each horizon of prediction (p) is necessary.

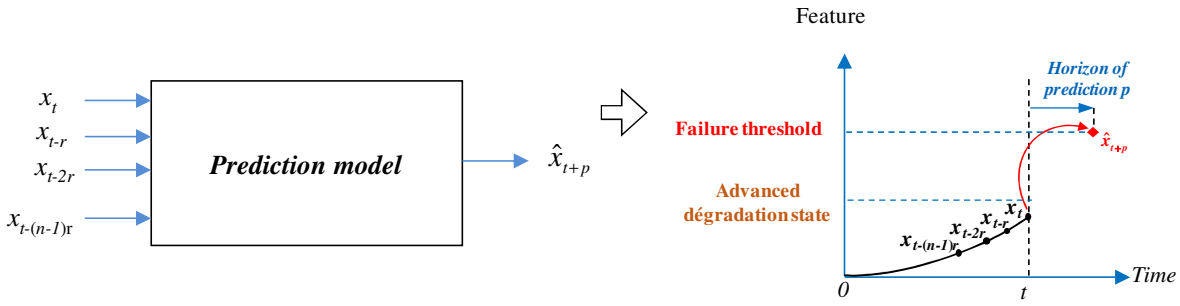

Fig. 4. One-step ahead prediction.

As we will observe from Fig. 48 of the experimental section, the prediction error for this approach is high, which leads us to opt for the second approach.

- The second approach proposes to predict directly the value of the observation $\hat{x}_{t+p}$ by cascading several prediction models to make closer estimates at one-step prediction in the future from the previous ones. The predictions of the first become the input of the second and so on (see Fig. 5). The relationship between the time-series and the output of the prediction model is expressed as follows :

$$\hat{x}_{t+p} = f\left(x_t, x_{t-r}, x_{t-2r}, x_{t-3r}, ..., x_{t-(n-1)r}\right) \tag{2}$$

In terms of complexity, only one training is necessary.

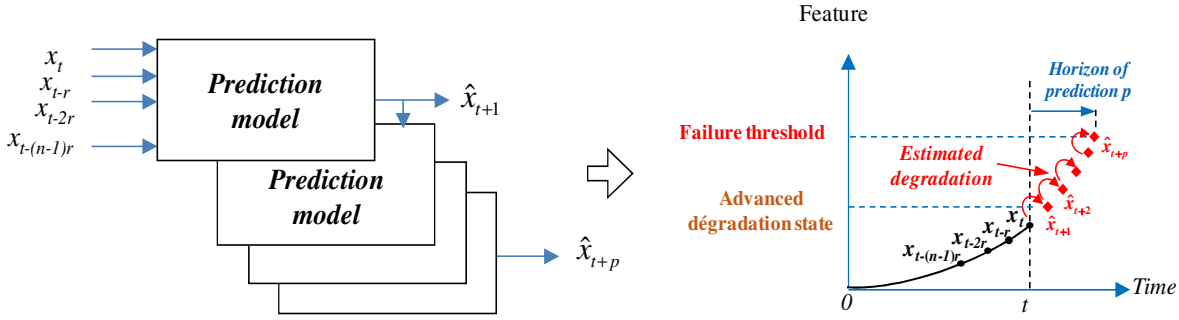

Fig. 5. Multistep ahead prediction.

Among estimators commonly used in the construction of prediction models, it appears that an Artificial Neural Network (ANN) is the most adapted for our approach. Indeed, the idea of using ANN for the prognosis is not new. Their use was initiated in early 1964 for climate prediction using the Widrow adaptive network [20], [21]. Their principle is based on learning the climate feature for the re-estimation of the ANN synaptic weights. The training algorithms was limited at this time, which limited the development of prediction models based on ANN. Since the 1980s, further development allowed to ANN to become an unavoidable prediction model.

Fig. 6 shows the chronological evolution of the ANN. We can deduce from this Figure that there are two types of ANN: recurrent neural networks and non recurrent neural networks [22]. The difference between these two types lies in their architecture and also in the correlation between the input and output data. ANNs are usually critiqued in two ways:

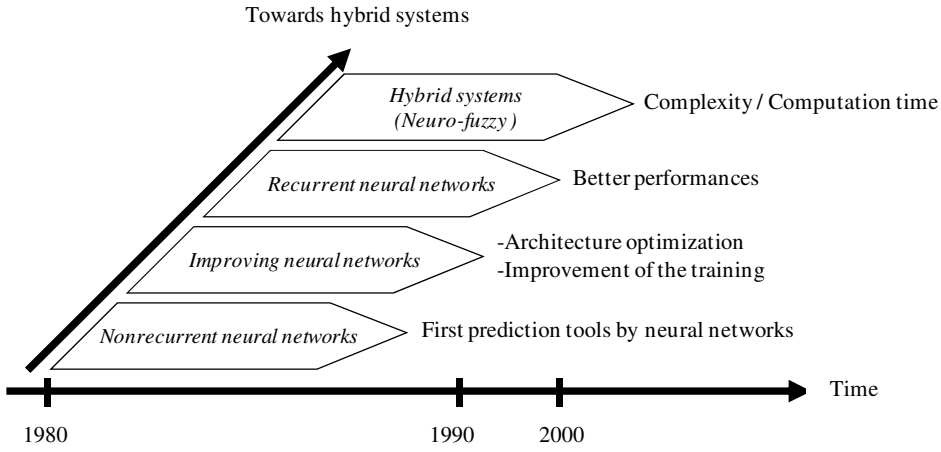

Fig. 6. Towards neuro-fuzzy systems [22].

- The physical interpretation between the input and output data is nonexistent. For this reason, ANN is called "black box";
- The architecture of the ANN requires improvements to optimize the training.

We can answer the first criticism by introducing the concept of fuzzy logic based on the rules of "If" and "Then". The goal is to replace the physical model with a linguistic understanding of the system by fuzzy rules. We will present two models based on fuzzy logic in sections 2.1 and 2.2.

We can also answer the second criticism by comparing two training techniques. The first consists of re-estimating the parameters of membership functions as well as synaptic weights. This training technique is found in the ANFIS model. The second technique is based on the multiplication of known and fixed membership functions and the training of synaptic weights. This model is called NFN.

The following section will therefore be dedicated to the presentation of the two prediction models, namely the ANFIS model and the NFN.

2.1. Time series prediction by ANFIS

As mentioned previously, a prediction model requires inputs and an output. The inputs are represented by a series of (n) observations measured up to the instant (t). These observations $\{x_t, x_{t-r}, x_{t-2r}, x_{t-3r}\ldots x_{t-(n-1)r}\}$ correspond to a series of measurement recorded respectively at time (t, t-r,t-$2r$, t-$3r$,..., t-(n-1)r) where (n) corresponds the number of measurements. Each observation is separated from another by an acquisition step (r) equal to "1" to facilitate the understanding of the architecture. The output corresponds to an observation estimated at (t + 1) (principle of the multi-step ahead prediction). The particularity of the ANFIS model lies in rules associated with the inputs. If we consider two membership functions (k = 1,2) of triangle, trapezoidal or Gaussian type (see Equation (4)) associated with each input of the model, then we obtain 2^n rules of Sugeno type of first order [23]. These rules will subsequently create a linear function linking the input observations to the estimated output. These rules are described as follows:

$$\begin{aligned}&\text{Rule}(j):\\&\text{If } \left(x_t \text{ is } A_0^j\right) \text{ and } \left(x_{t-1} \text{ is } A_1^j\right) \text{ and } \left(x_{t-2} \text{ is } A_2^j\right) \text{ and} \ldots. \text{ and } \left(x_{t-(n-1)} \text{ is } A_{(n-1)}^j\right), \quad j = 1,2,\ldots,2^n\\&\text{Then } \hat{x}_{t+p}^j = f_j(x_t, x_{t-1}, x_{t-2}, x_{t-3}, \ldots, x_{t-(n-1)}) = c_0^j\, x_t + c_1^j\, x_{t-1} + c_2^j\, x_{t-2} + c_3^j\, x_{t-3} + \ldots + c_{(n-1)}^j\, x_{t-(n-1)} + c_n^j\end{aligned} \tag{3}$$

where $A_{i\ i=0,1,2,\ldots,n-1}^j$ represents a membership function linked to the observation x_{t-i}. This membership function is denoted $\mu_k(x_{t-i})$. The parameter k is between 1 and Q where Q is the total number of membership functions associated to each input. The choice of the membership function (k) associated with each observation x_{t-i} depends on the rule (j) chosen. $\{c_0^j, c_1^j, c_2^j, \ldots, c_n^j\}$ are the coefficients of the rule (j). They are called the consequent parameters.

These parameters are estimated in the training process. Each rule (j) generates a linear function denoted f_j associated to the output $\hat{x}_{t+p}^{j}$. This output depends on the prediction strategy. For the one-step ahead prediction, $p > 1$ and for the multistep ahead prediction, $p = 1$.

The ANFIS architecture corresponding to our approach is shown in Fig. 7. We can notice from this architecture that ANFIS consists of five layers. They are identified as follows:

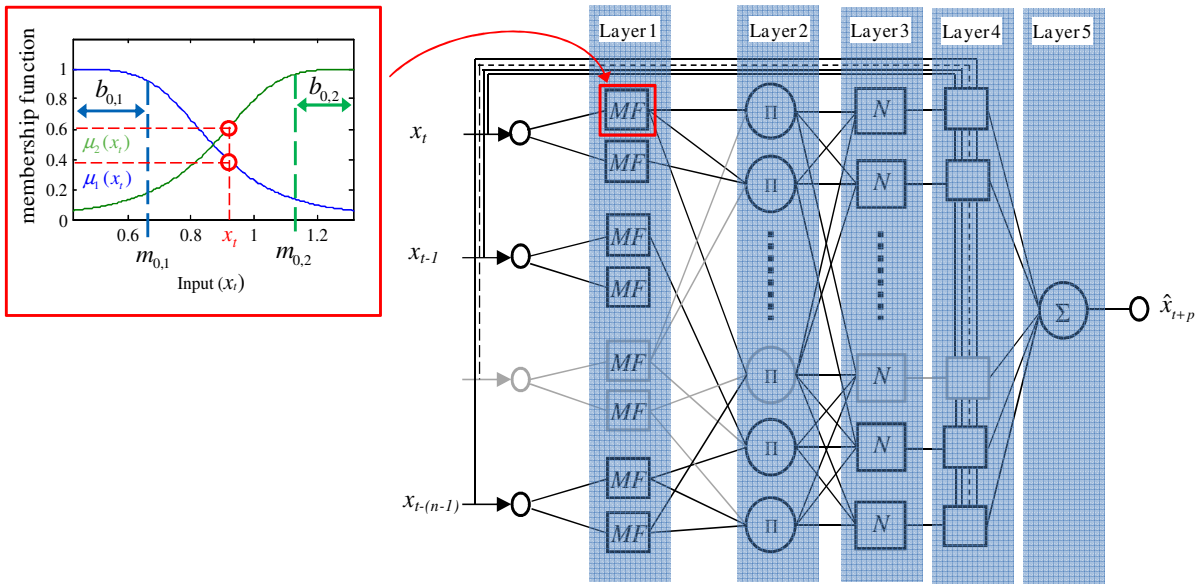

Fig. 7. Architecture of the adaptive neuro neuro-fuzzy inference system model with n-inputs and two MFs for each one.

Layer 1 : This layer is called the fuzzyfication layer. Each observation x_{t-i}, with $i = 0,1,2,...,n-1$, is associated with k-membership functions to evaluate the membership degree of this observation compared to the rule (j). The aim of each membership function is to evaluate the membership degree of the i-th input x_{t-i} compared to the fuzzy partition (k) associated with the rule (j = 1,2,...,2^n). This partition (see Fig. 7 for an example of k = 1,2) is characterized by a function named $\mu_k(x_{t-i})$ which depends on the input x_{t-i}, the mean value $m_{i,k}$ and the standard deviation $b_{i,k}$ if we consider this function as Gaussian as shown in Equation (4). This function can be represented by a trapezoid or a triangular function. The choice of the membership function will be defined by the training rate.

$$\mu_k(x_{t-i}) = \exp\left(-\left[\frac{x_{t-i} - m_{i,k}}{b_{i,k}}\right]^2\right), \quad \begin{matrix} i = 0,1,2,...,n-1 \\ k = 1,2,3,...,Q \end{matrix} \tag{4}$$

where $m_{i,k}$ and $b_{i,k}$ are the parameters of the Gaussian function named also premise parameters. Q is the total number of membership function associated to each observation input. (n) represents the number of observations of the time series.

Layer 2 : This layer is called the weighting layer where each rule (j) is labelled with the symbol Π representing a weighting by the T-norm. This T-norm makes possible to associate a weight to each rule (j) by a product between the membership functions. This weight is symbolized by ω_j and is identified by the following formula:

$$\omega_j = \mu_k(x_t) \cdot \mu_k(x_{t-1}) \cdot \mu_k(x_{t-2}) \cdot ... \cdot \mu_k(x_{t-(n-1)}) \ , \quad k = 1,2,3,...,Q \tag{5}$$

Layer 3 : This layer is called the normalization layer where each weight associated with rule (j) is normalized compared to the weights of all rules. The goal is to normalize all weights in order to evaluate the weight of each rule

(*j*). The normalization is formulated as follows:

$$\bar{\omega}_j = \frac{\omega_j}{\sum_{j=1}^{2^n} \omega_j}, \qquad j = 1,2,...,2^n \tag{6}$$

Layer 4 : This layer is called the defuzzyfication layer. As mentioned previously, each rule allows estimating an output noted $\hat{x}_{t+1}^j$. This output is derived from a function f_j depending on the input observations $\left\{x_t, x_{t-1}, x_{t-2}, x_{t-3}, ..., x_{t-(n-1)}\right\}$ as well as on the consequent parameters $\left\{c_0^j, c_1^j, c_2^j, c_3^j, ..., c_{(n-1)}^j, c_n^j\right\}$. At this stage, we try to match to each output (*j*) its corresponding weight. This correspondence is obtained by the Takagi-Sugeno approach which proposes to multiply each output by its weight as shown in the following formula:

$$\bar{\omega}_j \hat{x}_{t+p}^j = \bar{\omega}_j \cdot \left(c_0^j x_t + c_1^j x_{t-1} + c_2^j x_{t-2} + c_3^j x_{t-3} + ... + c_{(n-1)}^j x_{t-(n-1)} + c_n^j\right), \qquad j = 1,2,...,2^n \tag{7}$$

Layer 5 : This layer is called the summation layer. The output of the prediction model is obtained by summing all the outputs of the rules (*j*). The result is given as follows:

$$\hat{x}_{t+p} = \sum_{j=1}^{2^n} \bar{\omega}_j \hat{x}_{t+1}^j = \sum_{j=1}^{2^n} \bar{\omega}_j \cdot \left(c_0^j x_t + c_1^j x_{t-1} + c_2^j x_{t-2} + c_3^j x_{t-3} + ... + c_{(n-1)}^j x_{t-(n-1)} + c_n^j\right) \tag{8}$$

The prediction model presented here depends on two sets of parameters: the consequent parameters and the premise parameters. These parameters are estimated from a supervised training database. The goal is to match the input observations to the desired output and that through the parameters mentioned above. To do this, a hybrid algorithm combining the gradient descent method and the least squares estimation method is proposed [24]. The algorithm uses the gradient descent method to estimate premise parameters ($m_{i,k}^j$ and $b_{i,k}^j$) and the least squares method to identify the consequent parameters $\left\{c_0^j, c_1^j, c_2^j, c_3^j, ..., c_{(n-1)}^j, c_n^j\right\}$. To facilitate the understanding of the training process, consider that the output of the ANFIS model is expressed from one observation (n=1) and two rules (j=2):

$$\begin{aligned} \hat{x}_{t+p} &= \bar{\omega}_1 \cdot \left(c_0^1 x_t + c_1^1\right) + \bar{\omega}_2 \cdot \left(c_0^2 x_t + c_1^2\right) \\ \hat{x}_{t+p} &= \left(\bar{\omega}_1 x_t\right) c_0^1 + \left(\bar{\omega}_1 c_1^1\right) + \left(\bar{\omega}_2 x_t\right) c_0^2 + \left(\bar{\omega}_2 c_1^2\right) \end{aligned} \tag{9}$$

Where $\left(c_0^1, c_1^1, c_0^2, c_1^2\right)$ are linear consequent parameters.

The above equation shows that by fixing the premise parameters, the output of the ANFIS model is linear compared with the consequent parameters. The estimation by the least squares method can be then used to determine the optimal consequent parameters. At this step, inputs and outputs are known. Inputs are therefore propagated to layer 4. Remember that at this step, the premise parameters are fixed. Then, the estimation by the least squares method is applied. Then, the optimal consequent parameters being known, the formula below is applied to optimize the premise parameters also called weights of the ANFIS model:

$$\Delta\alpha = -\eta \frac{\partial E}{\partial \alpha} \tag{10}$$

Where η is the learning step, α is the consequent parameter to be modified and $\frac{\partial E}{\partial \alpha}$ is the gradient of the error (error = reference - effective output). This term will be calculated by the gradient retro-propagation algorithm.

2.2. Time series prediction by NFN

The major disadvantage of the ANFIS model is its slow convergence during the training phase when the number of membership functions is high and/or when the number of inputs is high. To resolve this problem, we introduce a new prediction model called the "neon-fuzzy neuron" NFN. This model was introduced by Takeshi Yamakawa in 1992 [25]. Its architecture is very close to an ANN with (n) inputs. The difference lies in the membership functions

assigned to each input and the removal of the activation function. Synaptic weights and membership functions are contained in nonlinear synapses denoted SYN_i with i= 0,1, ..., n-1. Each synapse is thus connected to an input observation. With membership functions being fixed, the training is applied on the weights. One of the advantages of the NFN is its rapidity of convergence during the training phase.

The NFN model is presented in Fig. 8. We can observe from this Figure that the synaptic weights are not linear. They depend on the membership functions using the "If-Then" logic. We can also observe that each synapse has a summation point which makes possible to compute a function denoted $fi(\overline{x}_{t-ir})$ with i = 0,1,..., n-1. This function makes possible to estimate the contribution of each input by a sum between the synaptic weights multiplied by their respective membership functions. This function is then weighted with the functions of the other inputs to estimate the output of the model. This output is expressed by the following formula:

$$\hat{x}_{t+p} = \sum_{i=0}^{n-1} f_i(\overline{x}_{t-ir}) = \sum_{i=0}^{n-1}\sum_{j=1}^{h} \omega_{ji} \cdot \mu_{ji}(\overline{x}_{t-ir}) \tag{11}$$

where $\overline{x}_{t-ir}$ is the normalized i-th observation (i=0, 1, 2,…, n-1), (n) represents the number of observations of the time series, (r) represents the acquisition step, $\hat{x}_{t+p}$ is the model output. μ_{ji} is a membership function defining the membership degree of an input $\overline{x}_{t-ir}$ compared to a fuzzy partition noted (j) $_{1\le j\le h}$ where h represent the number of partitions of the membership function (see Fig. 9). ω_{ji} is a synaptic weight corresponding to the membership function.

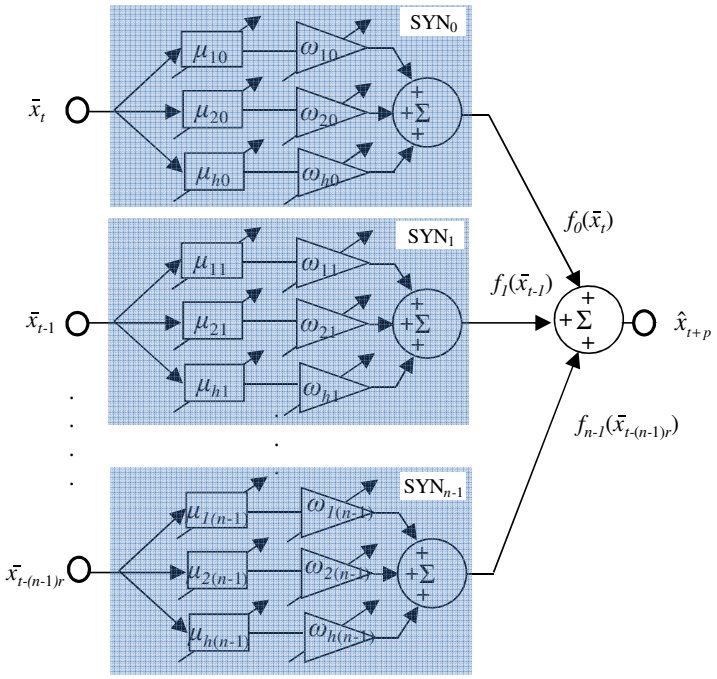

Fig. 8. Architecture of the neo-fuzzy neuron model with n-inputs and h-membership functions for each one.

The membership function used in our case is a triangular function normalized to "1" in amplitude and in input observation $0 \le \overline{x}_{t-ir} \le 1$ and $0 \le \mu_{ji}(x_{t-ir}) \le 1$. It is divided into h equivalent partitions, thus making possible to locate an input on a defined range and to calculate subsequently its membership degree. Fig. 9. allows to understand the advantage of this type of membership function.

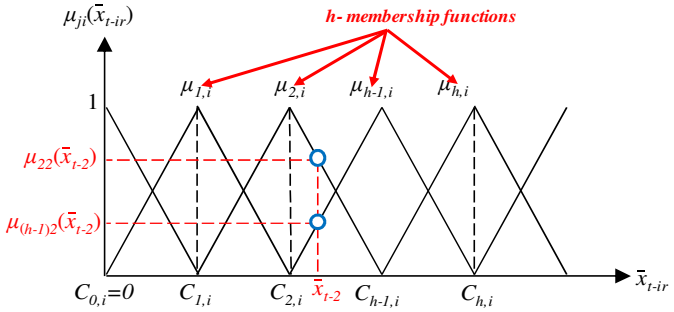

Fig. 9. Shape of the triangular membership functions in the neo-fuzzy neuron model.

For a normalized input $\bar{x}_{t-ir}$, the corresponding membership function μ_{ji} ($\bar{x}_{t-ir}$) can have only two possible values. They are given as follows:

$$\mu_{ji}(\bar{x}_{t-ir}) = \begin{cases} \dfrac{\bar{x}_{t-ir} - c_{j-1,i}}{c_{ji} - c_{j-1,i}}, & \text{if } \bar{x}_{t-ir} \in [c_{j-1,i}, c_{ji}] \\ \dfrac{c_{j+1,i} - \bar{x}_{t-ir}}{c_{j+1,i} - c_{ji}}, & \text{if } \bar{x}_{t-ir} \in [c_{ji}, c_{j+1,i}] \end{cases}, \quad \begin{matrix} 0 \le i \le n-1 \\ 1 \le j \le h \end{matrix} \tag{12}$$

where c_{ji} are the centers of the h-triangles.

We can therefore say that the input $\bar{x}_{t-ir}$ activates only two neighboring membership functions, thus reducing considerably the calculation time during the training phase. The sum of these two functions is equal to 1:

$$\mu_{ji}(\bar{x}_{t-ir}) + \mu_{j+1,i}(\bar{x}_{t-ir}) = 1 \tag{13}$$

Unlike the ANFIS model, the training phase in the NFN model concerns only the synaptic weights. The estimation of these weights is based on the principle of minimizing the quadratic error obtained by calculating the difference between the desired output and the estimated output. The goal is to minimize this difference. An algorithm is proposed to optimize the calculation time while minimizing the error. The algorithm is presented as follows:

Algorithm 1. Training steps of the NFN model

(1) Initialize ℓ to 0
(2) **For** i from 0 to n-1
(3) **For** j from 1 to h
(4) $\omega_{ji}(\ell\text{-}1)$ = initial value (0.01 in our case)
(5) **end**
(6) **end**
(7) $\hat{x}_{t+p}(\ell) \leftarrow \sum_{i=0}^{n-1} f_i(x_{t-ir}) = \sum_{i=0}^{n-1}\sum_{j=1}^{h} \omega_{ji}(\ell-1)\mu_{ji}(\bar{x}_{t-ir})$
(8) $E(\ell) = \frac{1}{2}(x_{t+p}(\ell) - \hat{x}_{t+p}(\ell))^2$
(9) **While** E(ℓ) $\geq$ defined threshold
(10) $\ell \leftarrow \ell+1$
(11) $e(\ell) = x_{t+p}(\ell) - \hat{x}_{t+p}(\ell) = x_{t+p}(\ell) - \sum_{i=0}^{n-1}\sum_{j=1}^{h} \omega_{ji}(\ell-1)\mu_{ji}(\bar{x}_{t-ir})$
(12) $\omega_{ji}(\ell) = \omega_{ji}(\ell-1) + \eta e(\ell)\mu_{ji}(\bar{x}_{t-ir})$
(13) Repeat step (7) and (8)
(14) **end**

with η: the learning parameter which determines the training speed.

2.3. Performance evaluation of the ANFIS and NFN models

The performance of a prediction model is evaluated by the following criteria:

- The root mean squared error (RMSE)

$$RMSE = \sqrt{\frac{1}{N}\sum_{p=1}^{N}(x_{t+p} - \hat{x}_{t+p})^2} \tag{14}$$

- The mean absolute percentage error (MAPE)

$$MAPE\,(\%) = \frac{1}{N}\sum_{p=1}^{N}\left|\frac{x_{t+p} - \hat{x}_{t+p}}{x_{t+p}}\right| \times 100 \tag{15}$$

where N is the total number of tested observation. x_{t+p} and $\hat{x}_{t+p}$ are the actual and the predicted values of the p-th observation test, respectively.

Most researchers use essentially the MAPE to judge the quality of a predictive model. The MAPE is preferred because it normalizes the result for any feature. On the other hand, studies indicate that a prediction problem is clearly nonlinear, and values close to zero ($x_{t+p} \approx 0$) can give erroneous results (MAPE $\nearrow \infty$). For this reason, measures based on the RMSE are suggested as they penalize large errors. The RMSE makes possible to avoid the case "not determined" from the mathematical point of view, when x_{t+p} is too close to 0.

2.4. Analysis of ANFIS and NFN performance for the prediction of chaotic time series

The purpose of this section is to test the performance of the NFN and ANFIS models for the prediction of the Mackey-Glass (MG) time series [26]. This time series is generated by the following differential equation:

$$\dot{x}(t) = \frac{0.2\,x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1\,x(t) \tag{16}$$

The time series of MG is chaotic, so there is no clearly defined period. This is a well known problem in the research community to test prediction models. To obtain the data of the time series, we applied the method of Runge Kutta of order 4 to Equation (16). We assume here that $x(0) = 1.2$, $\tau = 17$ and $x(t) = 0$ for $t > 0$. The shape of the MG time series is given in Fig. 10.

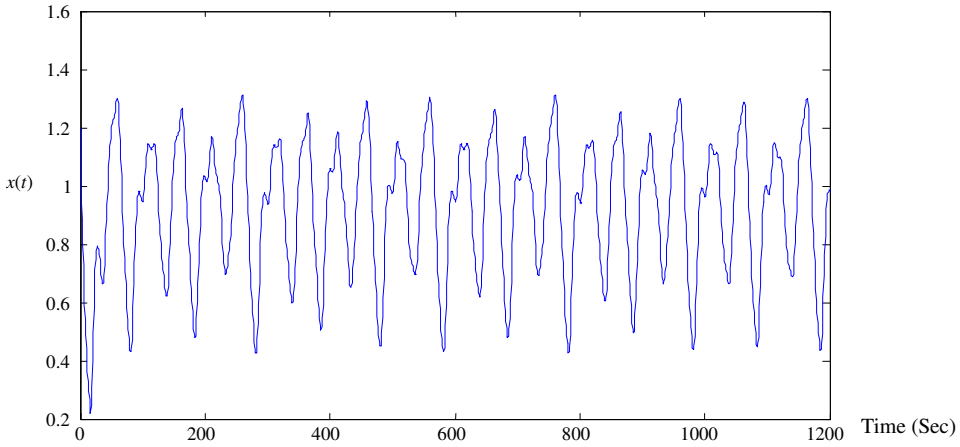

Fig. 10. Mackey-Glass time Series.

We need to use known values from the time series of MG until time (t) to predict the value at a given moment in the future at ($t+p$). The standard approach for this type of prediction is to create a time series $\{x_{t-(n-1)r}\ldots x_{t-3r}, x_{t-2r}, x_{t-r}, x_t\}$ to predict the future value x_{t+p}. Following the classical parameters of the time series prediction of the MG, we put $p = 6$, $r = 6$ and $n = 4$. At each time step t, the input data for the NFN and ANFIS are a time series of the form: $\{x_t, x_{t-6}, x_{t-12}, x_{t-18}\}$. The output of the NFN and ANFIS corresponds to the prediction $\hat{x}_{t+6}$.

We choose a data interval between 118 and 1118 seconds to train and test the NFN and ANFIS models. We obtain a database of 1000 observations. The first 500 observations are used for the training, while the remaining 500 are used for the test. For the ANFIS model, two bell-shaped membership functions are assigned to each input. This generates 16 fuzzy rules and 104 parameters to adjust (24 premise parameters and 80 consequent parameters). The initial membership functions which are equally spaced and covering the entire input space are given in Fig. 11. The membership functions after the training phase are given in Fig. 12. For the NFN, 55 equidistant membership functions of "triangle" type are assigned to each input. This gives us a total of 220 synaptic weights to adjust.

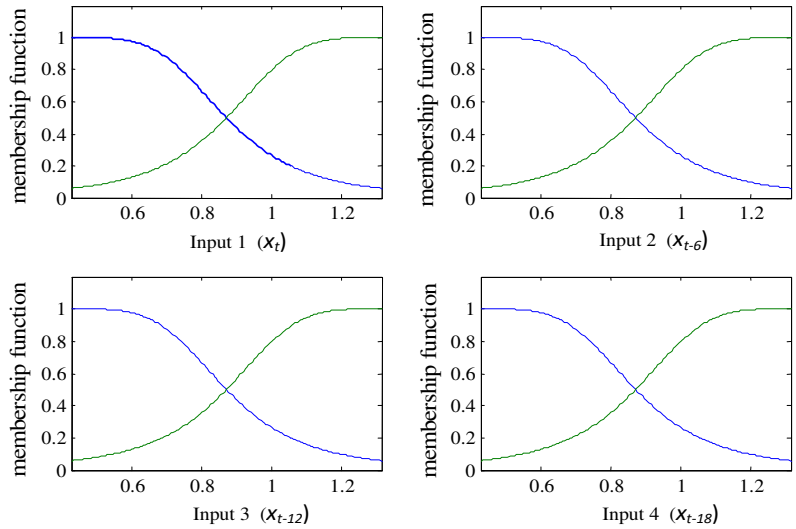

Fig. 11. Initial membership functions.

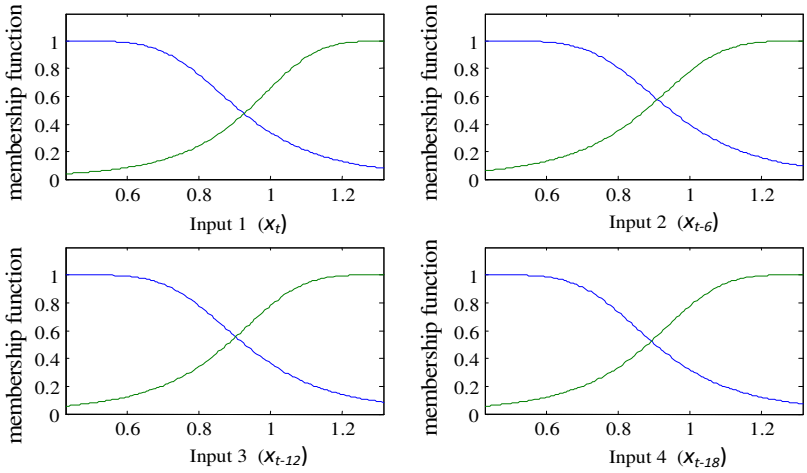

Fig. 12. Membership functions after the training phase.

Note that these membership functions do not change significantly after the learning phase. Obviously, the fitting is done by linear parameters, while the nonlinear parameters allow last adjustments.

The normalized root mean squared error (NRMSE) of observations predicted by the NFN and ANFIS models as well as other models is given in Table 1. The NRMSE is defined as the root mean squared error divided by the

standard deviation of the target time series.

Table 1. Prediction error for 500 observations test with an horizon of prediction $p = 6$

Method	$NRMSE_{test}$
Linear model prediction	0.55
Cascade correlation NN	0.06
6th-order polynomial	0.04
Genetic algorithm & fuzzy system [27]	05 membership functions ⟶ 0.049 07 membership functions ⟶ 0.042 09 membership functions ⟶ 0.038
ANFIS	0.0077
NFN	0.031

We can see from Table 1 that the proposed two models give very good results in the case of the time series of MG.

Figures 13 and 14 show the prediction result of the 500 observations test and their prediction error for the NFN and ANFIS models.

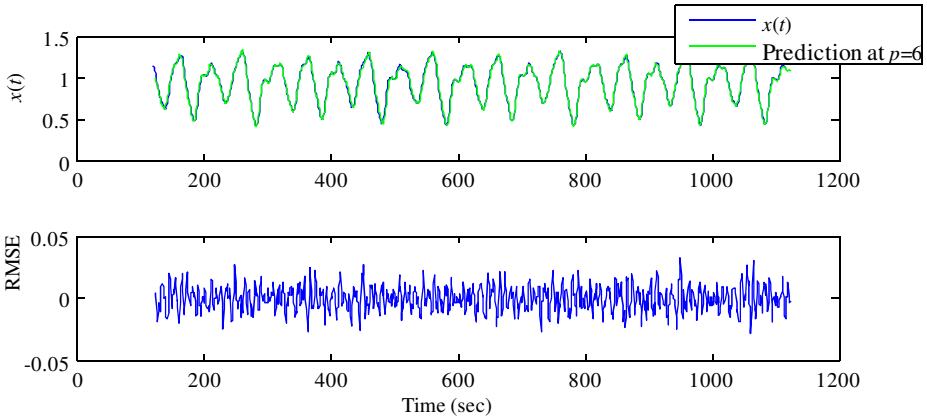

Fig. 13. Prediction results for the neo-fuzzy network. RMSE: root mean squared error.

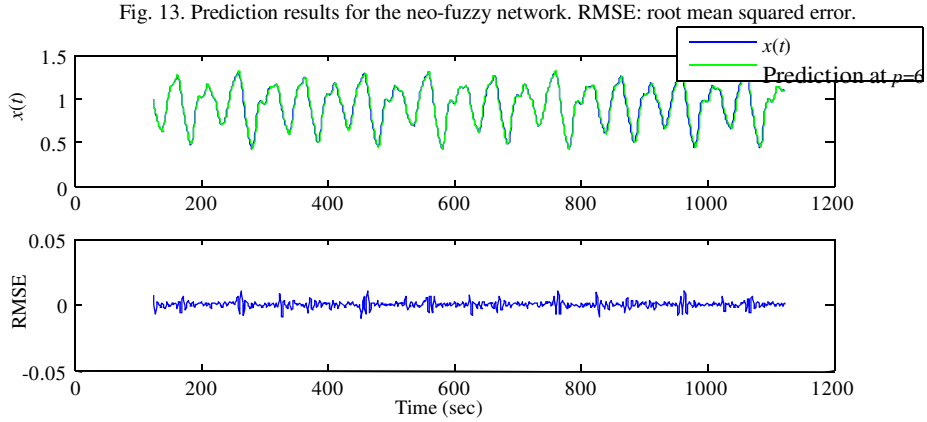

Fig. 14. Prediction results for the adaptive neuro neuro-fuzzy inference system. RMSE: root mean squared error.

From the previous Figures, we can conclude that the NFN offers great advantages for modeling complex systems thanks to the simplicity of its structure composed of a single neuron. The ANFIS needs to change the number of layers, the number of neurons in each layer and the membership function to find the structure that gives a good fit for the problem to model. In the NFN, it is necessary to modify only the number of fuzzy partitions of the input variables which makes possible to find the most appropriate structure easily.

The NFN and ANFIS models will be used for the prediction of bearing failure as we will see in the experimental section.

3. Identification of the degradation states by pattern recognition

In the introduction section, we came to the conclusion that the long-term prediction error is very high with the absence of defined degradation states. To correct this problem, it is necessary to define thresholds (degradation states) from which the prediction models will be able to give a reliable estimation of the RUL. In this section, we are interested in a method based on the classification of observations to identify degradation states. The classification consists of classifying observations by comparing them to patterns. An observation is characterized by a set of q-features. An observation is represented by a point in a q-dimensional space called the representation space defined by certain features. These features are extracted from analyses performed on the signals measured by the sensors installed on the bearing. The features being real values, an observation noted 'x_i' is defined by a vector of q-features $x_i = [x_{i1}, x_{i2}, ..., x_{iq}]$ (see Fig. 15).

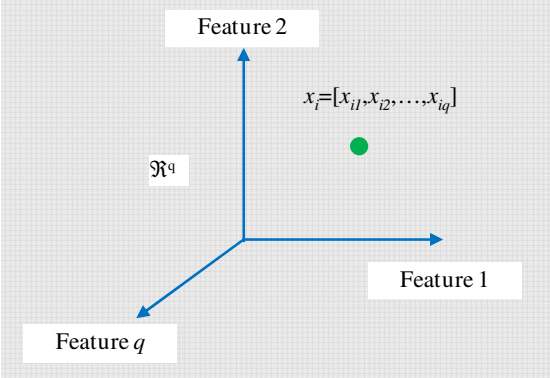

Fig. 15. Representation of an observation in a q-dimensional space.

The purpose of the classification is to associate an observation 'x_i' with a pattern called a class. A class Ω_i is composed by a set of similar observations. In the context of the prognosis, a class Ω_i represents a degradation state of the system. In the ideal case, a class is composed by identical observations represented by a single point in the representation space. Nevertheless, disturbances (measurement noise, sensor accuracy, etc.) can appear. This implies that an observation is different to another observation in the same class. The influence of noise on a set of M-classes (Ω_1, Ω_2, ..., Ω_M) representing the M-degradation states gives M-patterns projected in the representation space (see Fig. 16). The goal will be then to identify these patterns to identify the degradation states.

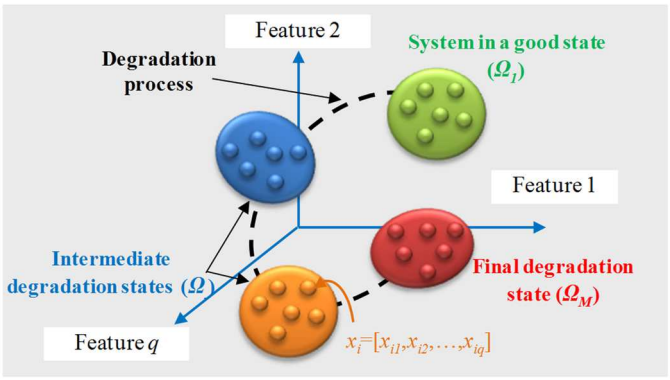

Fig. 16. Visualization of the different degradation states of a system by classes.

The classification method presented here is based on the principle of pattern recognition. As shown in Fig.17, a complete pattern recognition approach is composed of several elements.

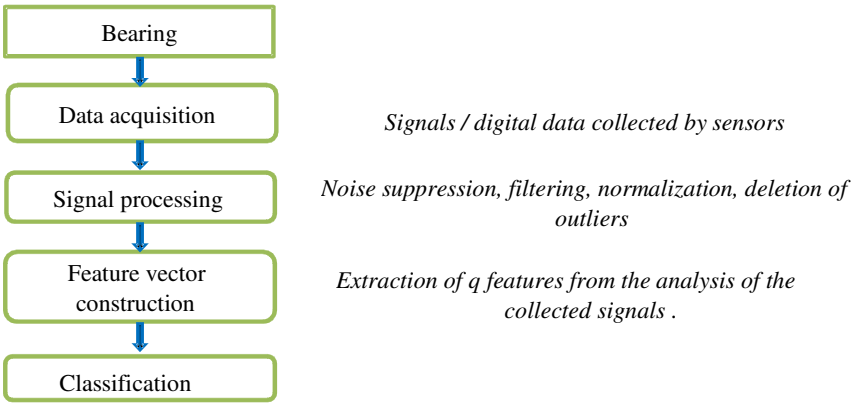

Fig. 17. Necessary step for the classification.

The identification of the different degradation states is obtained in two steps:

Construction of the feature vector: The feature vector is composed by a set of q-features obtained by using various signal processing techniques (time, frequency, time-frequency analysis) on measured signals collected from sensors placed on the bearing. More explanation on the construction of the feature vector is given in the experimental section.

Classification: Two kind of classification techniques, namely, supervised and unsupervised classification are defined. Supervised classification is based on a priori knowledge of measured observations to identify classes, while unsupervised classification is based on the similarity between observations to identify classes. Having no prior knowledge of the class composition, we will use the principle of unsupervised classification to identify the different degradation states.

3.1. Unsupervised classification by the artificial ants algorithm

We propose here a classification algorithm based on the behavior of an ants colony to construct a set of M-paths connecting the ant nest to the food source in order to classify a set of N-observations by using the notion of grouping between ants. Note that the path followed by an ant represents the class of membership of an observation.

The principle of the artificial ants algorithm is to consider a population of N-ants carrying a set of N-observations. Each observation is described by a q-feature vector. The aim is to group these ants in a set of M-classes (Ω_1, Ω_2,..., Ω_M) so that the similar observations are grouped together in the same classes while the dissimilar observations are in separate classes. The value of M is unknown. The grouping of these ants is obtained by a succession of interactions between them. Each ant has rules of behavior that it follows according to the situations that it meets:

- The nest contains the observations to be classified and corresponds to the starting point for ants going to the food source;
- The path taken by ants to the food source is a class. All paths taken by ants before arriving to the food source correspond to the M-classes;
- Searching for a path to a food source corresponds to the search for a class for observations;
- Each ant taking a path memorizes it by depositing a certain quantity of pheromone. The quantity of pheromone deposited on each path depends on the number of ants which followed this path;
- The choice of a path for an ant depends on the similarity between this ant and all ants which had taken this path.

The artificial ants algorithm can be summarized in two steps; steps 1 and 2 are an optimization of the ascendant hierarchical classification.

Step 1

Consider a population of *N*-ants noted $Ant_{i\ i\in(1,\ldots,N)}$ carrying *N*-observations. These *N*-ants (initially located in the nest) will classify the *N*-observations in a first set of *m*-classes ($m \leq N$) and doing this by trying to choose a path to a food source. This step is performed by a similarity function denoted Sim_1 (Ant_i, Ant_j). This function measures the degree of similarity between an ant (Ant_j) carrying an observation x_j located in the nest and another ant (Ant_i) carrying an observation x_i that goes to the food source by taking a path (*i*). The similarity function Sim_1 (Ant_i, Ant_j) is given as follows:

$$Sim_1\ (Ant_i, Ant_j) = [\tau_i] \cdot \left[d(x_i, x_j) \right]^{-1} \tag{17}$$

where $d(x_i, x_j)$ is the distance between the observations carried by the ants (Ant_i, Ant_j) and τ_i is the quantity of pheromone deposited on the path (*i*).

At the beginning, the quantity of pheromone deposited by each ant is equal to the following:

$$\underset{i\in(1,\ldots,N)}{\tau_i} = \xi \tag{18}$$

Once the research group has left the nest, the quantity of pheromone deposited by ants on each path is updated by the following:

$$\underset{i\in(1,\ldots,m)}{\tau_i} = n_{Ants_i} \cdot \xi \tag{19}$$

where $\xi \in [0\text{-}1]$ is a constant related to the quantity of pheromone deposited by ants and n_{Ants_i} is the number of ants which had taken the *i*-th path.

The first set of *m*-classes (initial classes) is obtained by the following algorithm:

Algorithm 2. Step 1: Behavior simulation of an ant Ant_j moving when it leaves the nest

```
(1)  Place all observations in the nest
(2)  Evaluate the quantity of pheromone of each ant by (18)
(3)  While there are ants in the nest
(4)       Find the most similar ant (Ant_j) to the ant (Ant_i) based on the similarity given by (17)
(5)       if  Sim_1 (Ant_i,Ant_j) ≥ S_Sim      % S_sim represents the similarity threshold %
(6)            The ant (Ant_j) will take the same path as the ant (Ant_i)
(7)            Update the quantity of pheromone given by (19)
(8)       end
(9)       if  no ant takes a path
(10)           Decrease the similarity threshold S_Sim
(11)      end if
(12) end
```

In the first step, the number of operations required for the *N*-ants to follow the *m*-paths is logarithmic, so a complexity of O (N log N). This represents an advantage over the hierarchical ascendant classification where the complexity is O (N^2).

At the end of this step, the *m*-paths represent the *m*-classes. An example of the behavior of *N*-ants when they leave the nest towards the food source by taking the *m*-paths is shown in Fig. 18.

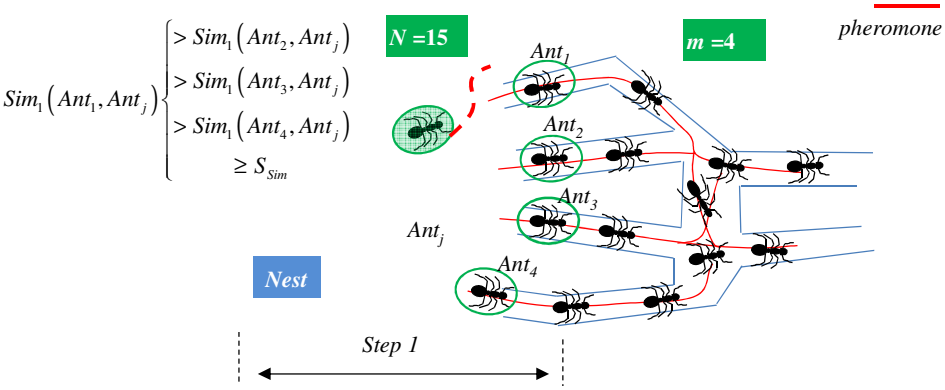

Fig. 18. Example of the behaviour of an ant (Ant_j) when it leaves the nest.

Step 2

Once the nest is empty, the *N*-ants continue their paths and try to optimize the classification of *N*-observations in a second set of classes ($M < m$) and doing this by choosing a path to the food source.

The goal of this step is to regroup similar classes to reduce their number. This is done by a second similarity function denoted Sim_2 ($Ants_i$, $Ants_j$). This function measures the degree of similarity between a group of ants ($Ants_j$) taking the path (i) and another group of ants ($Ants_i$) located in the path (j) to the food source. It returns a value between 0 (when i and j are totally different) and 1 (when i and j are identical). This similarity is formulated as follows:

$$Sim_2\left(Ants_i, Ants_j\right) = 1 - \frac{[\tau_i]^e \cdot \left[d\left(g_i, g_j\right)\right]^f}{\sum_{A\in(1,\ldots,m)} [\tau_A]^e \cdot \left[d\left(g_A, g_j\right)\right]^f} \tag{20}$$

with

- $d(g_i,g_j)$: the distance between the gravity center of the observations carried by the groups of ants ($Ants_i$, $Ants_j$);
- $f \in [0, 1]$ and $e \in [-1, +1]$ are two parameters defining respectively the importance of the pheromone deposited by ants on the paths (i) and (j) and the distance between them. τ_A corresponds to the quantity of pheromones deposited by ants in the A-th path. g_A corresponds to the gravity centre of the observations carried by ants following the A-th path.

The role of the parameters e and f is as follows: if $e = 0$, this corresponds to a classical similarity between two groups of ants based on the Mahanalobis distance. If $e > 0$, this corresponds to a similarity between two groups of ants ($Ants_i$, $Ants_j$) discriminating the group of ants ($Ants_i$) taking a path accumulating an important quantity of pheromone. This constrains the ant group ($Ants_j$) to explore other paths by seeking (if possible) a better similarity with other groups of ants. If $e < 0$, this corresponds to a similarity between two groups of ants ($Ants_i$, $Ants_j$) which, unlike the previous situation, favours the group of ants ($Ants_i$) taking a path with an important quantity of pheromone. This accelerates the classification process. If $f = 0$, only the quantity of pheromone is taken into account. Therefore, a compromise between the distance and the quantity of pheromone is a good approach to optimize classification.

The quantity of pheromone deposited on each path taken by the groups of ants is updated by the following:

$$\underset{i\in(1,\ldots,M)}{\tau_i^{t+1}} = n_{Ants_i} \cdot (1-\varepsilon) \cdot \tau_i^t \tag{21}$$

where $\varepsilon \in [0, 1]$ is the evaporation factor. It is used to prevent an unlimited accumulation of pheromone on paths and in our case, to limit the weight of each class. n_{Ants_i} is the number of ants taking the i-th path. g_i, g_j are the gravity centers of the observations carried by the group of ants $Ants_i$, $Ants_j$ respectively.

The obtained M-classes must be evaluated by a criterion noted J. This criterion allows evaluating the compactness and separability of the obtained M-classes. It is formulated as follows :

$$J = trace\left(\Sigma_W^{-1} \cdot \Sigma_B\right) \tag{22}$$

where Σ_W is intra-class variance matrix and Σ_B is the inter-class variance matrix.

$$\Sigma_w = \frac{1}{N}\sum_{i=1}^{M}\sum_{j=1}^{n_i}\left(x_{ij} - g_i\right)\cdot\left(x_{ij} - g_i\right)' \tag{23}$$

$$\Sigma_B = \frac{1}{N}\sum_{i=1}^{M}\left(g_i - g\right)\cdot\left(g_i - g\right)' \tag{24}$$

with:
- g: gravity center of all observations;
- M: number of classes;
- g_i: gravity center of the class Ω_i
- x_{ij}: j-th observation of the class Ω_i
- n_i: number of observations of the class Ω_i
- N: total number of the set of observations.

An increase of the criterion J between the instant (t), where two classes (Ω_i, Ω_j) (characterized by two paths taken by two groups of ants ($Ants_i$, $Ants_j$)) are about to be grouped, and the instant (t +1) where the two classes are grouped allows approving the grouping. In the opposite case, the two groups of ants ($Ants_i$, $Ants_j$) will look for other paths to satisfy the criterion J. The second set of M-classes is obtained by the following algorithm:

Algorithm 2. Step 2: Behavior simulation of two group of ants $Ants_i$, $Ants_j$ moving to the food source	
(13)	Evaluate the criterion J (t) of paths taken by ants by the formula (22)
(14)	Evaluate the quantity of pheromone deposited on each path taken by ants by using formula (21)
(15)	Find two similar groups of ants $Ants_i$, $Ants_j$, using two different paths by using formula (20)
(16)	Group two groups of ants $Ants_i$, $Ants_j$ on the same path
(17)	Evaluate the criterion J (t + 1) of the path taken by the two groups of ants $Ants_i$, $Ants_j$
(18)	**if** J (t + 1) > J (t)
(19)	**While** J (t + 1) - J (t) > 0
(20)	Validate the path taken by the two groups of ants
(21)	J (t) ← J (t+1)
(22)	Update the quantity of pheromone of this path by using formula (19)
(23)	Repeat steps 15 to 17
(24)	**end**
(25)	**end**

An example of the behaviour of two groups of ants towards the food source by taking the M-paths is shown in Fig. 19.

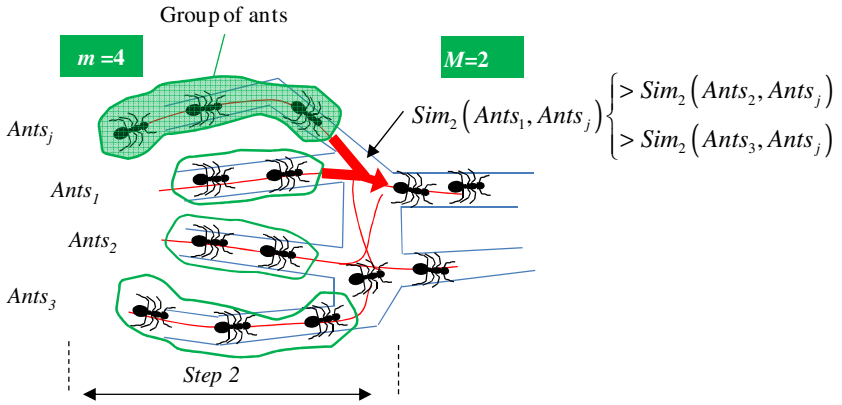

Fig. 19. Example of the behaviour of two group of ants ($Ants_i$, $Ants_j$) if they use two similar paths.

In the second step, the number of operations necessary for the *N*-ants to join the *M*-paths is polynomial. The complexity is in $O(m^2)$ with $M \ll N$.

4. Experimentation

The degradation states used to validate the prognostic models are obtained from the degradation database of four roller bearings provided by the University of Cincinnati. The aim of the prognosis in this case is to detect the different degradation states and to estimate the RUL before the appearance of the next degradation state.

The degradation process of bearings was carried out on a test bench designed by Rexnord Corp. The test bench was composed of four ZA-2115 roller bearings placed on a shaft, as shown in Fig. 20. The characteristics of these bearings are detailed in Table 2.

The rotation speed of the tested bearings was fixed at 2000 rpm. The aging was obtained by a spring mechanism generating a radial load of 26.6 KN on the tested bearings. An accelerometer was placed on the vertical axis of each bearing. The vibration signals were recorded every 10 minutes with a sampling frequency of 20 kHz.

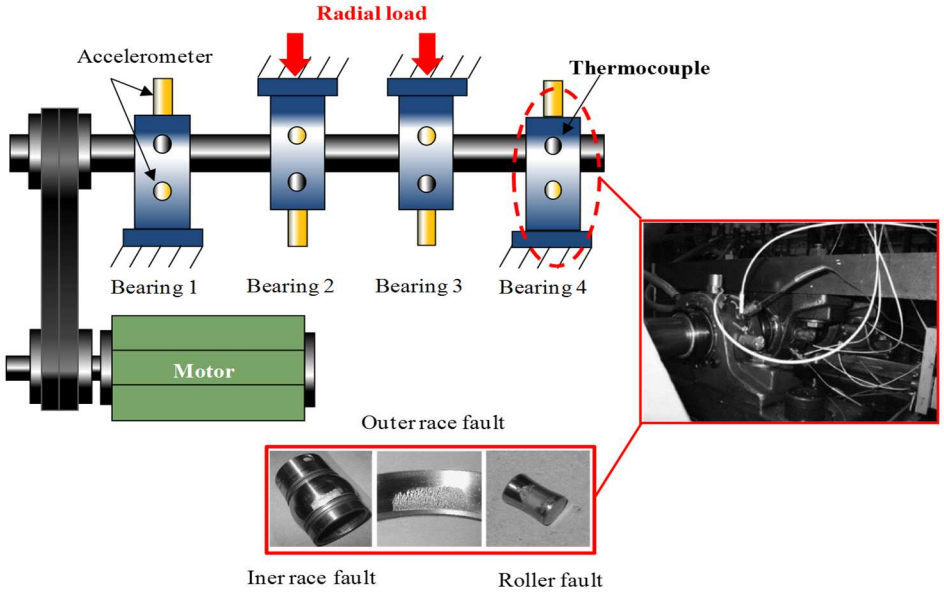

Fig. 20. Aging roller bearings test bench [28].

Table 2. Bearings characteristics

Bearing diameter (mm)	8.4
Roller number	16
Primitive diameter (mm)	71.62
Contact angle (°)	15.17

Construction of the feature vector : As mentioned previously, features can be obtained by using time, frequency and time-frequency analysis. In our case, it was sufficient to analyze the temporal characteristic of the vibration signals of bearings (1, 2, 3 and 4) to extract relevant features. In this case, we used statistical features. These features will be used to identify the different degradation states of bearings. These features were summarized in Table 3.

Table 3. Statistical features extracted from the vibration signal

Energy		Peak amplitude		Mean value		Crest factor	
$E=\sum_{\ell=1}^{N} x_\ell^2$	γ_1	$S_{peak}=\sup_{1\le\ell\le N}\lvert x_\ell\rvert$	γ_2	$\bar{x}=\frac{1}{N}\sum_{\ell=1}^{N} x_\ell$	γ_3	$\mathrm{F_c}=\frac{S_{crête}}{S_{eff}}$	γ_4
Power		**Root mean square (RMS) value**		**Kurtosis**		**Factor K**	
$P=\frac{1}{N}\sum_{\ell=1}^{N} x_\ell^2$	γ_5	$S_{RMS}=\sqrt{\frac{1}{N}\sum_{\ell=1}^{N}(x_\ell)^2}$	γ_6	$s_{kurt}=\frac{\frac{1}{N}\sum_{\ell=1}^{N}(x_\ell-\bar{x})^4}{S_{eff}^4}$	γ_7	$\mathrm{F}_K=\mathrm{S}_{peak}\cdot\mathrm{S}_{RMS}$	γ_8

with:

- $x(t)$: acquired signal;
- N: number of samples composing the signal;
- f_{ech}: sampling frequency;
- $x_\ell = x(t= \ell/f_{\mathrm{ech}})$, with ℓ nonzero positive integer less than or equal to N.

Fig. 21 shows the evolution of bearings degradation thanks to the features extracted from the vibration signals acquired from accelerometers placed on the vertical axis. Note that bearings (1, 2, 3 and 4) were tested during 35 days.

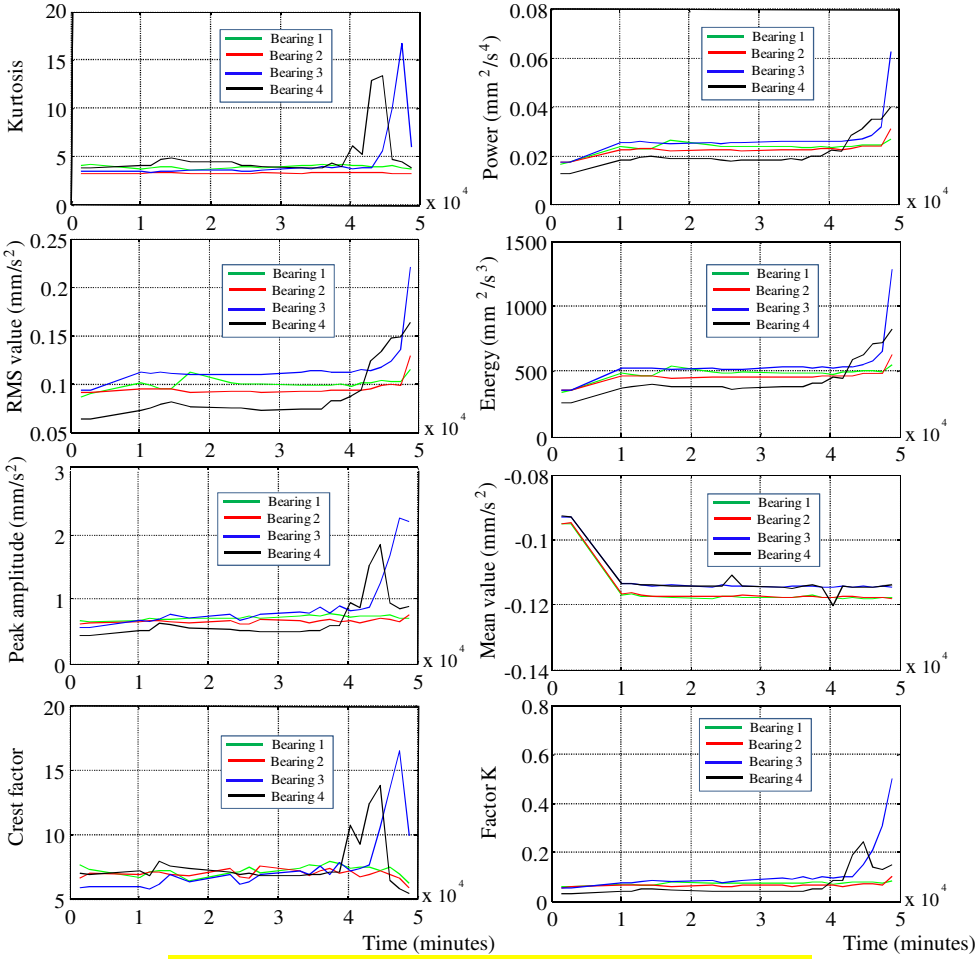

Fig. 21. Evolution of statistical features extracted from the vibration signal.

Fig. 21 shows the evolution of the RMS value, power, energy and the factor K of vibration signals recorded during the test period. After two days of test (2880 minutes), the first signs of aging were observed in the bearing 4. These signs are due to a phenomenon called "breaking-in phase" of the roller bearing: the lubricant present in the bearing will be gradually evacuated and the imperfections of the metal will gradually disappear. This is an important phenomenon that must be observed in order to avoid a misinterpretation of the rolling degradation process.

After 30 days of test (43200 minutes), the first signs of degradation were observed. After that, the RMS value, energy and power continue to increase as degradation increases. We can observe that the factor K is not a good feature for tracking bearing degradation. This is shown in Fig. 21, where the factor K for bearing 4 begins to decrease from 44640 minutes. The same remark can be said for the Kurtosis, crest factor and peak amplitude. A decrease in the value of these features is observed for bearings 3 and 4 after 44640 and 47520 minutes of test.

The energy of the vibration signal is considered as the image of power over time. We will choose the power in our case. We finally keep three features: the mean value, the power and the RMS. It can also be noted that these three parameters give a better representation of the degradation of the bearings (see Fig. 24).

4.1. Classification (identification of degradation states) :

The artificial ants algorithm was used to identify the unknown *M*-degradation states from the historical data of bearings (1, 2 and 3). A detailed analysis of the different monitoring features can help us to choose the most relevant ones. Based on Fig. 21, we chose the root mean square (RMS), the mean value and the power of the vibration signal. The rest of the features were not selected because when the degradation increased, the value of these decreased. For the selected features (γ_3, γ_5, γ_6), several observations disturbed by the noise can appear. Thus, a class that groups observations of the same degradation state can be represented by a geometric pattern of observations. If the features are well chosen, each degradation state can be represented by a compact class grouping similar observations in the representation space as shown in Fig. 24.

Step 1

Fig. 22 shows the evolution of the criterion *J* compared to the similarity threshold values (S_{sim}). We can differentiate from this Figure two very distinct areas indicating the evolution of the criterion *J*. In the first area, S_{sim} decreased from 132 to 43. This decrease corresponded to a grouping of observations varying between 24 classes for $J = 132$ and 9 classes for $J = 43$. The intra-class variance in this case is close to 0, which gives a high value for *J*. On the other hand, if the number of classes decreases, the criterion *J* decreases because the intra-class variance increases. This observation is made in the second area with a criterion $J \in [5,8\text{-}43]$. A similarity threshold set at 0.4 generates a number of *m*-classes equal to 9, sufficient to optimize the classification.

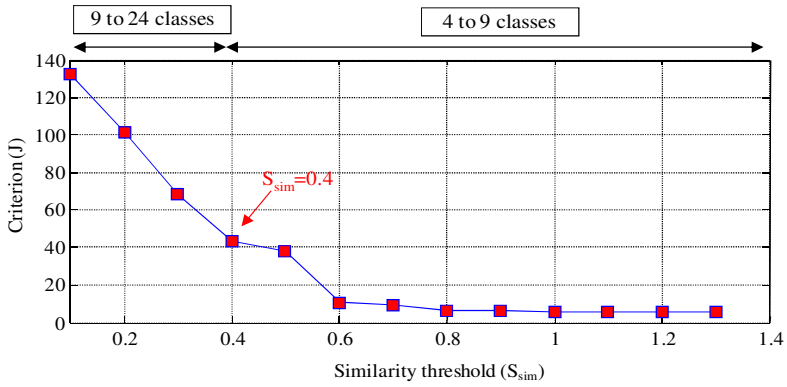

Fig. 22. Evolution of the criterion *J* compared to the similarity threshold.

Step 2

The second step allowed grouping the previous classes into a small number of *M*-classes. The parameter ε was set to 0.01 in order to limit the weight of each class. The parameter ξ was arbitrarily set to 1. The highest value of criterion *J* will define the best parameters *e* and *f*, such that: for a *J* around 99%, $e = -0.4$ and $f = 0.6$ (see Fig. 23).

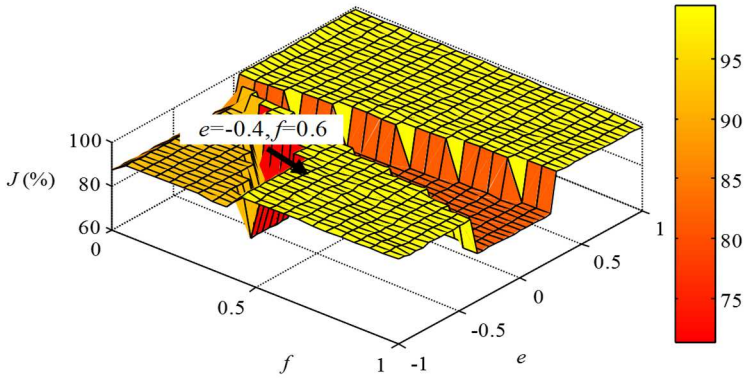

Fig. 23. Determination of the parameters e and f from the criterion J.

The determination of the parameters e and f allowed to the second step of the artificial ants algorithm to obtain a second set of M-classes, with $M = 4$. A principal component analysis (PCA) was carried out in order to change the orientation axis of the features allowing a better visualization of the classes. Fig. 24 shows the position of the M-classes of the test set in the 3-dimensional subspace after the PCA.

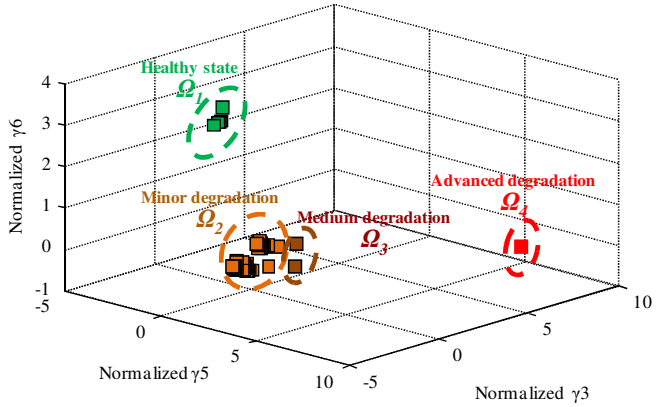

Fig. 24. Representation of the classes obtained by the artificial ants algorithm in the 3-dimensional subspace after the PCA.

As shown in Fig. 24, the artificial ants algorithm identified four classes ($M = 4$), noted respectively Ω_1, Ω_2, Ω_3, and Ω_4. The first class (Ω_1) included observations belonging to bearings 1, 2 and 3 acquired during the two first days of testing. Considering that the bearings were in a good state at the beginning of the experimentation, this class represented bearings in a healthy state. The last class (Ω_4) corresponded to the last degradation state observed in the last day of the test. This degradation is due to a defect located in the inner race (see Fig. 25). This class corresponded to an advanced degradation state. Two other classes were identified by the artificial ants clustering. These classes corresponded to the intermediate degradation states. Finally, we obtained four classes corresponding respectively to the healthy state, minor degradation state, medium degradation state and advanced degradation state.

Fig. 25. Inner race fault observed in bearing 3.

The acquired database corresponded to the historical degradation process of bearings 1, 2, 3 and 4. This historical database was composed of 1944 observations. 1458 observations were used as a training database (486 observations for each of bearings 1, 2 and 3). 486 observations corresponding to bearing 4, which were different and independent of the training database, were used for the validation of the proposed prediction models. The statistical

features (γ_3, γ_5, γ_6) of the bearings 1, 2 and 3 are used to estimate the parameters of the NFN and ANFIS.

4.2. Estimating the RUL before the next degradation states of bearing 4

The ANFIS and the NFN models were used to estimate the RUL before the next degradation states of the bearing 4. For this, we will predict the evolution of the three selected features (γ_3, γ_5, γ_6) between the time instant (t), where the imminence of the next degradation state (75% of the degradation severity) is detected, and the time instant ($t + p$), where the appearance of this degradation state is confirmed (see Fig. 26). The choice of 75% of the degradation severity was justified because it was possible to obtain earlier predictions which can be an advantage for developing a maintenance strategy.

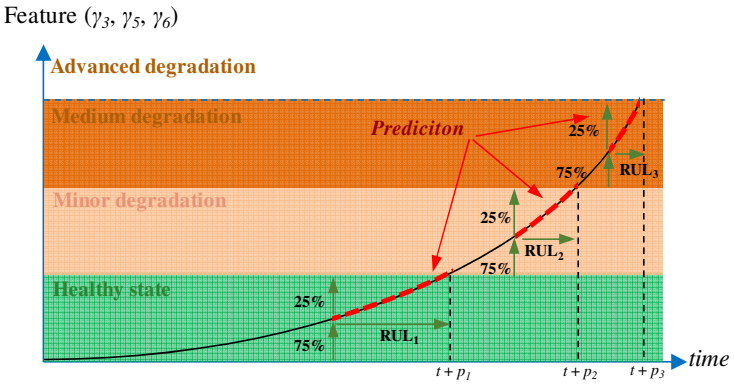

Fig. 26. Illustration of the estimation of the RUL after the detection of the imminence of the next degradation state

As presented in section 2, the prediction of x_{t+p} noted $\hat{x}_{t+p}$ can be obtained by two approaches:

4.2.1. Approach (1): Multistep ahead prediction

The first approach predicts the evolution of the features until the horizon p from the previous predictions. Since the value of p is not known, it will be determined by performing a diagnostic on the predicted p-values. Table 4 summarizes the different parameters implemented in the ANFIS and the NFN models.

Table 4. Configuring ANFIS and NFN

ANFIS	
Number of inputs (length of the time series) (n)	3
Acquisition step (r)	1
Number of membership functions	2
Membership function type	Gaussian
T- norm	Product
Inference type	Linear Sugeno
Training	
Number of observations	1944
Number of iterations	100
Learning / test database	1458/486
Learning Algorithm	LSE+ gradient descent
NFN	
Number of inputs (length of the time series) (n)	3
Acquisition step (r)	1
Number of membership functions	10 to 16
Membership function type	Triangular
Training	
Number of observations	1944
Number of iterations	100
Learning / test database	1458/486
Learning algorithm	LSE

Results

Based on the rule of 75% of threshold of each degradation state, Three starting points for the prediction models were identified:

- Minor degradation → t = 5553 minutes;
- Medium degradation → t = 42667 minutes;
- Advanced degradation → t = 49910 minutes.

Fig. 27 shows that the first degradation state is diagnosed by the artificial ants algorithm after 9176 minutes of test. The first starting point being fixed at 5553 minutes of test, the objective is to obtain a RUL_1 equal to 3623 minutes. Fig. 28 and 29 show that the second and the third degradation states are diagnosed by the artificial ants algorithm after 42667 and 49910 minutes of test respectively. The second and third starting points being fixed at 42667 and 49910 minutes of test, the objective is to obtain a RUL_2 equal to 193 minutes and a RUL_3 equal to 392 minutes.

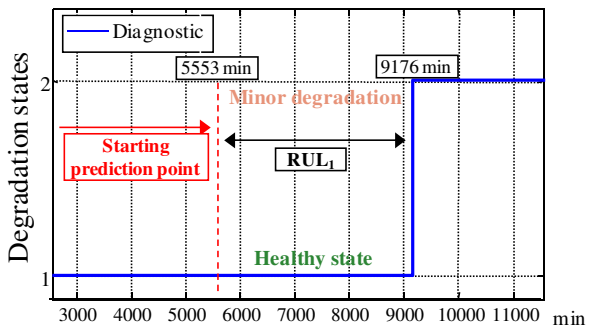

Fig. 27. Identification of the start and end points for the calculation of the RUL_1

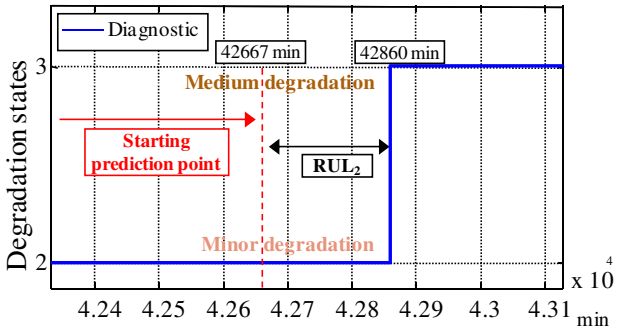

Fig. 28. Identification of the start and end points for the calculation of the RUL_2

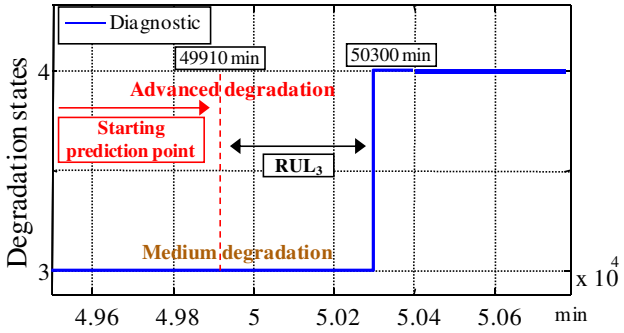

Fig. 29. Identification of the start and end points for the calculation of the RUL_3

- Minor degradation ⟶ t = 5553 minutes

Since bearing 4 started with a healthy state, it is now necessary to predict the minor degradation. Fig. 30, 32 and 34 show the evolution of RMSE for each of the monitored features (γ_3, γ_5, γ_6).

- **Power of the vibration signal (γ_3)**

A time series $\{x_t, x_{t-1}, x_{t-2}\}$ of three temporal observations is used as inputs in the ANFIS model. The obtained outputs give satisfactory results for a short-term prediction (up to 6000 minutes). The obtained RMSE increased between 6×10^{-5} and 8.83×10^{-4}. For an horizon of prediction up to 7500 minutes, the obtained RMSE increased between 8.83×10^{-4} and 0.0109. The prediction results in the case of a medium-term prediction were not satisfactory

(see Fig. 30).

The NFN, like the ANFIS model, has a set of three time series. The RMSE varied between 8.26×10^{-5} and 8×10^{-4} for a short-term prediction and between 1.83×10^{-4} and 0.00375 for a medium and a long-term predictions. We can conclude that this architecture gives satisfactory results for short- and long-term predictions (see Fig. 31).

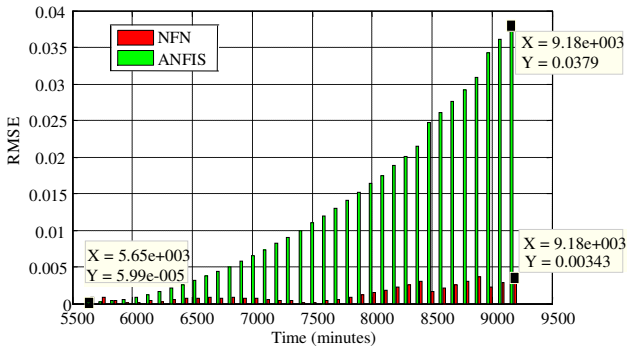

Fig. 30. Evolution of the RMSE for the power of the vibration signal.

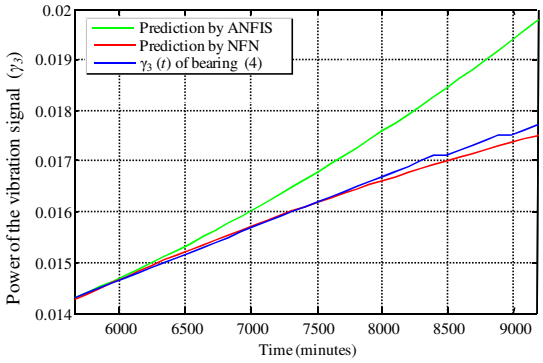

Fig. 31. Evolution of the prediction given by ANFIS and NFN for the power of the vibration signal.

- **RMS (γ_5)**

The ANFIS model with three inputs gave satisfactory results for the short and medium-term predictions (up to 7500 minutes). The obtained RMSE increased between 5×10^{-4} and 4.42×10^{-3}. For the long-term (up to 9175 minutes), RMSE increased until 9.54×10^{-3} (see Fig. 32). RMSE obtained by NFN varied between 6.94×10^{-5} and 3.98×10^{-4} for a short and medium-term predictions. In the long-term, RMSE increased until 0.0446 (see Fig. 32). These results showed that the architecture of NFN gave satisfactory results for short-term and even for the medium-term predictions (see Fig. 33).

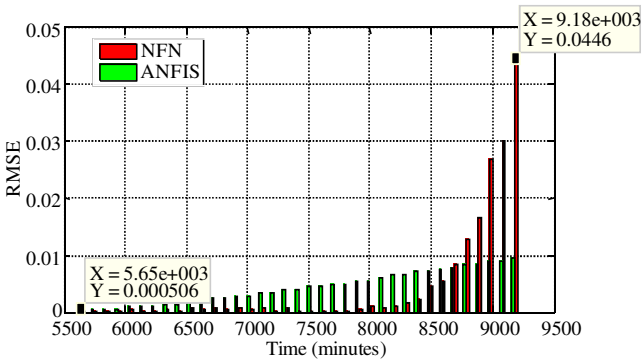

Fig. 32. Evolution of the RMSE for the RMS of the vibration signal.

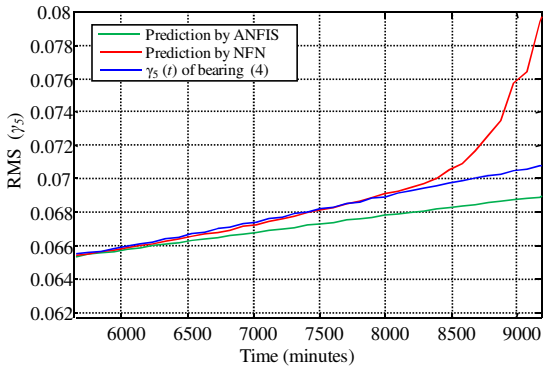

Fig. 33. Evolution of the prediction given by ANFIS and NFN for the RMS of the vibration signal.

- **Mean value of the vibration signal (γ_6)**

The ANFIS model gave satisfactory results for short-term predictions. For the medium and long-term prediction, the increase in RMSE indicated a poor prediction of the degradation process of bearing 4 (see Fig. 34 and 35). Results obtained by NFN were less precise than those obtained by ANFIS, either in the short or long term predictions.

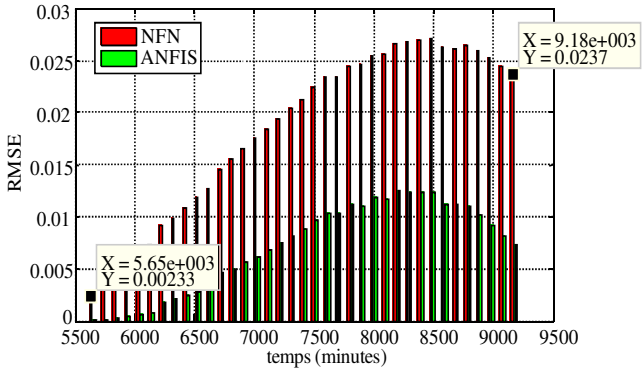

Fig. 34. Evolution of the RMSE for the mean value of the vibration signal.

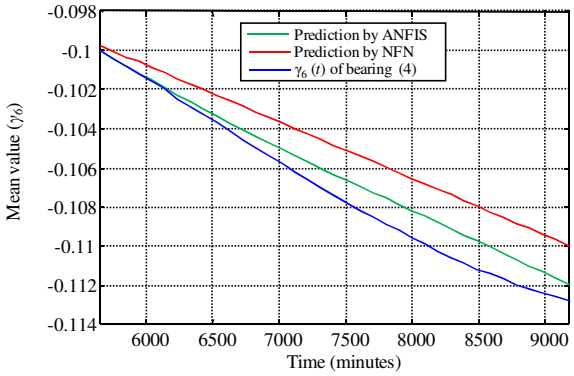

Fig. 35. Evolution of the prediction given by ANFIS and NFN for the mean value of the vibration signal.

The classification of the predicted observations from the statistical features (γ_3, γ_5, γ_6) allowed estimation of the appearance of the first degradation state after 9274 minutes of test. Based on the first starting point (minor degradation → t = 5553 minutes), the RUL_1 is equal to 3732 minutes (2 days, 14 hours and 9 minutes). The first degradation state was diagnosed after 9176 minutes of test which represents a RUL_1 of 3623 minutes (2 days, 12 hours and 14 minutes). We can deduce that the first approach combined with NFN gave a good prediction result. We can also give the same remark for the ANFIS which estimated the appearance of the first degradation state after 9372 minutes of test corresponding to RUL_1 equivalent to 3819 minutes (2 days, 15 hours and 36 minutes).

- Medium degradation state⟶ t= 42667 minutes

The first degradation state being diagnosed, it is now necessary to predict the evolution of the monitored features generated by the appearance of the medium degradation (see Fig. 37, 39 and 41). Fig. 36, 38 and 40 show the evolution of RMSE for each of the monitoring features (γ_3, γ_5, γ_6).

- **Power of the vibration signal (γ_3)**

The RMSE obtained with the ANFIS model did not exceed 0.0014. The RMSE obtained by the NFN model were slightly higher than the ANFIS model and did not exceed 0.0029.

- **RMS (γ_5) and the mean value (γ_6) of the vibration signal**

We can also give the same remarks for the RMS value. The RMSE for the two features do not exceed 0.0016 for the NFN model and 0.00083 for the ANFIS model.

The classification of the predicted observations from the statistical features (γ_3, γ_5, γ_6) allowed estimation of the appearance of the second degradation state after 42860 minutes of test. Based on the second starting point (medium degradation → t = 42667 minutes), the RUL_2 is equal to 193 minutes (3 hours and 12 minutes). The second degradation state was diagnosed after 42860 minutes which represents a RUL_2 of 193 minutes. We can deduce that the first approach combined with NFN gave an excellent prediction result. We can also give the same remark for the ANFIS which estimated the appearance of the second degradation state after 42860 minutes of test.

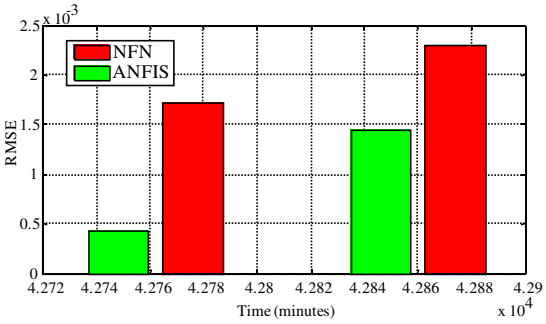

Fig. 36. Evolution of the RMSE for the power of the vibration signal.

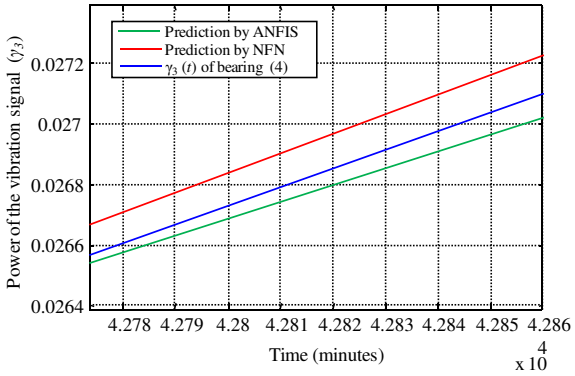

Fig. 37. Evolution of the prediction given by ANFIS and NFN for the power of the vibration signal.

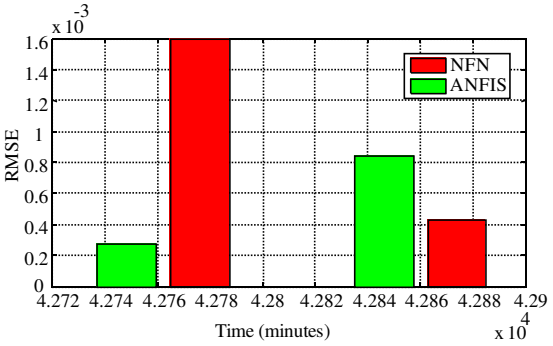

Fig. 38. Evolution of the RMSE for the RMS of the vibration signal.

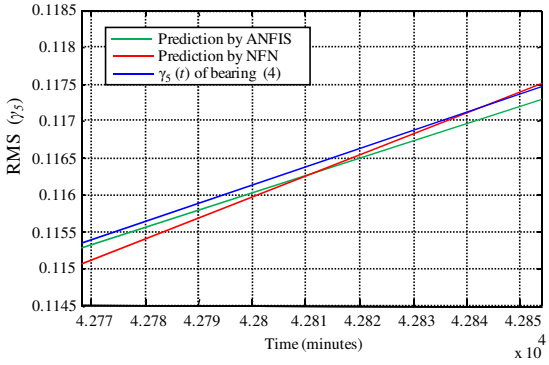

Fig. 39. Evolution of the prediction given by ANFIS and NFN for the RMS of the vibration signal.

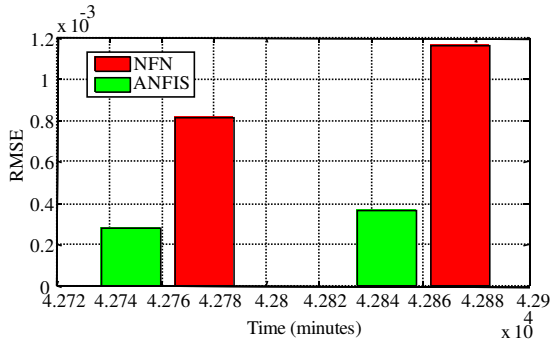

Fig. 40. Evolution of the RMSE for the mean value of the vibration signal.

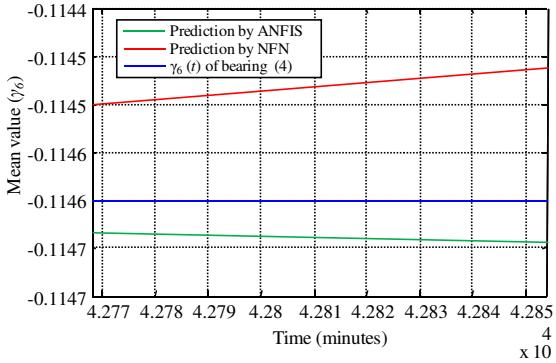

Fig. 41. Evolution of the prediction given by ANFIS and NFN for the mean value of the vibration signal.

- Advanced degradation state⟶ t= 49910 minutes

As in the previous case, the short-term prediction was very satisfactory, whether for the ANFIS or for the NFN model. The RMSE for ANFIS did not exceed 0.0012; and for NFN, the RMSE did not exceed 0.00034. Fig 42, 44 and 42 show the evolution of the RMSE for each of the monitoring features (γ_3, γ_5, γ_6). Fig. 43, 45 and 47 show the evolution of the predicted features before the appearance of the advanced degradation state.

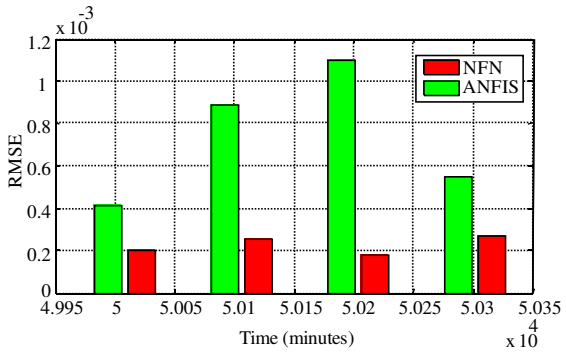

Fig. 42. Evolution of the RMSE for the power of the vibration signal.

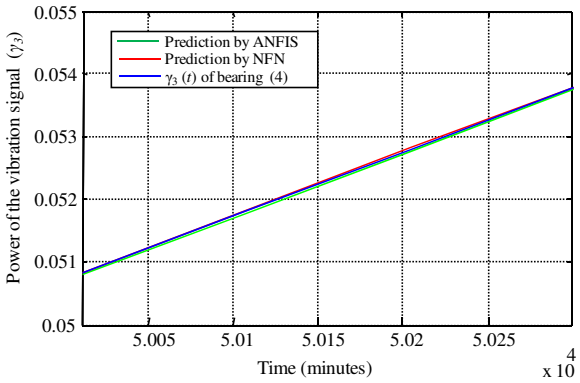

Fig. 43. Evolution of the prediction given by ANFIS and NFN for the power of the vibration signal.

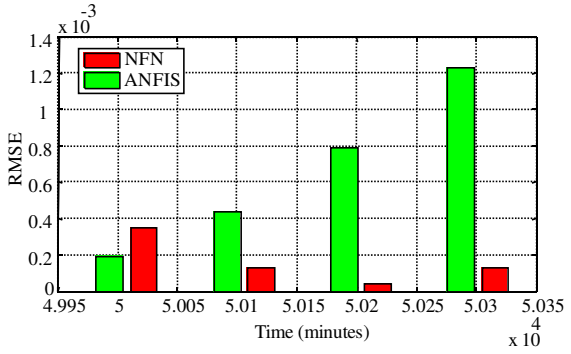

Fig. 44. Evolution of the RMSE for the RMS of the vibration signal.

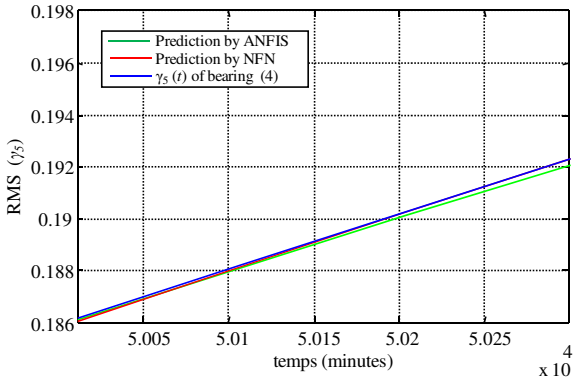

Fig. 45. Evolution of the prediction given by ANFIS and NFN for the RMS of the vibration signal.

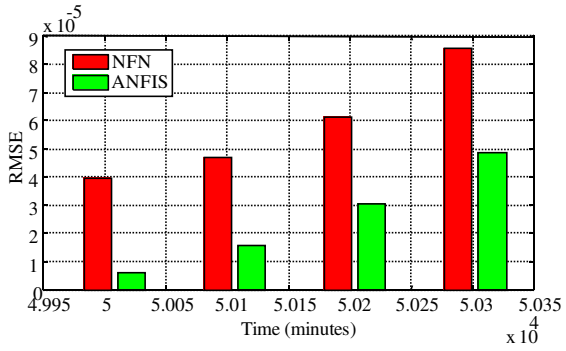

Fig. 46. Evolution of the RMSE for the mean value of the vibration signal.

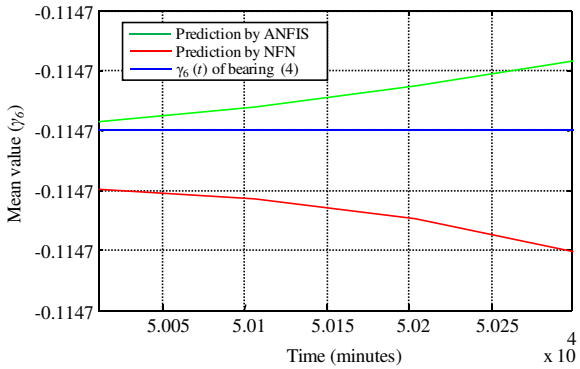

Fig. 47. Evolution of the prediction given by ANFIS and NFN for the mean value of the vibration signal.

The classification of the predicted observations from the statistical features (γ_3, γ_5, γ_6) allowed estimation of the appearance of the third degradation state after 50302 minutes of test. Based on the third starting point (advanced degradation → t = 49910 minutes), the RUL_3 is equal to 392 minutes (6 hours and 31 minutes). The third degradation state was diagnosed after 50302 minutes which represents a RUL_3 of 392 minutes. We can deduce that the first approach combined with NFN gave an excellent prediction result. We can also give the same remark for the ANFIS which estimated the appearance of the third degradation state after 50302 minutes of test.

4.2.2. Approach (2): One-step ahead prediction

Approach (2) consists of predicting the evolution of the observation from the time series $\{x_t, x_{t-r}, x_{t-2r}, \ldots, x_{t-(n-1)r}\}$ acquired until the time instant (t). The parameters of ANFIS and NFN are the same as those defined in Table 4 with $n = 3$ and $r = 1$. The following Tables correspond to the prediction of the average power of the vibration signal with several horizons of prediction (p = 1, 2, 10, 20, 30, 40).

Since bearing 4 started with a healthy state, it is now necessary to predict the minor degradation. Table 5, shows the evolution of the RMSE of the average power of the vibration signal (γ_3) for different horizons of prediction (p) for the ANFIS and NFN models and thus for:

- Minor degradation ⟶ t = 5553 minutes;
- Medium degradation ⟶ t = 42667 minutes;
- Advanced degradation ⟶ t = 49910 minutes.

Table 5. RMSE of γ_3 for different horizons of prediction (p)

Horizon of prediction (p)	$t+p$ (minutes)	RMSE ANFIS (02 membership functions)	RMSE NFN (10 membership functions)
Minor degradation ⟶ t = 5553 minutes			
1	5651	6×10^{-5}	8×10^{-4}
2	5749	0.0057	0.0033
10	6533	0.1812	0.0123
20	7513	0.4426	0.0159
30	8493	1.1043	0.027
40	9473	1.0716	0.056
Medium degradation state⟶ t= 42667 minutes			
1	42765	4.29×10^{-4}	1.72×10^{-3}
2	42863	0.0282	0.013
Advanced degradation state⟶ t= 49910 minutes			
1	50008	4×10^{-4}	2×10^{-4}
2	50106	0.0242	0.001
3	50204	0.088	0.002
4	50302	0.2435	0.003
Simulation time		47 seconds	8 seconds

4.2.3. Comparison between the two approaches

From the previous results, we can observe that the predictions results for p = 1, for the three degradation states, are the same as those for Approach (1). However, as the horizon of prediction increased, the value of the RMSE increased. This is shown in Fig.48 which indicates the RMSE obtained by Approach (1) and (2) from the time instant t = 5553 minutes. This leads us to conclude that approach (1) is best suited for short and long-term predictions.

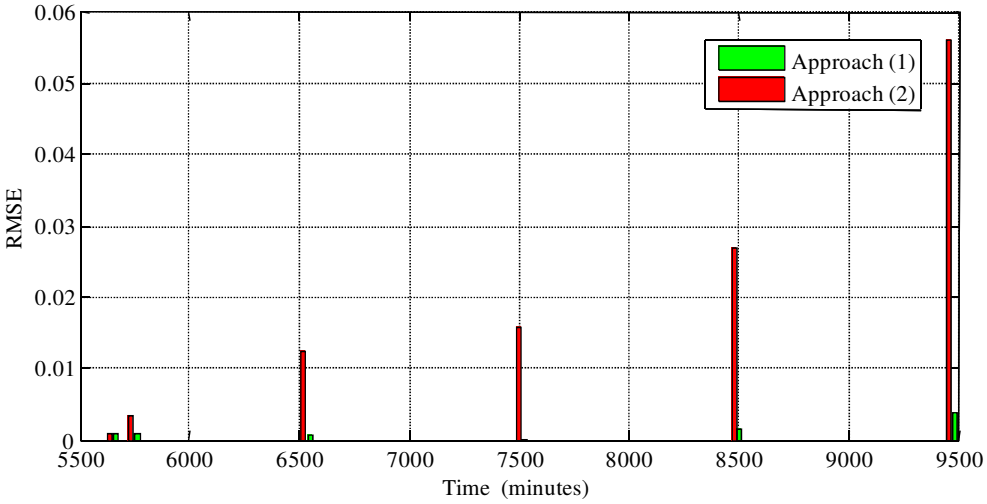

Fig. 48. Evolution of the root mean squared error (RMSE) for the power of the vibration signal.

5. Conclusion

As mentioned previously, the prognosis of bearings failure corresponds to the estimation of their remaining useful life. In this context, three models based on the principle of AI were proposed. The ANFIS and NFN models estimate the evolution of the degradation process, while the artificial ants algorithm identifies the different degradations states. A comparative study between different prediction models was done in the Mackey-Glass time series problem and in real data corresponding to the degradation of roller bearings. The obtained results showed the efficiency of the NFN and ANFIS models to obtain low estimation errors in short and long term predictions. In addition to these results, we arrived at the conclusion that NFN is a tool offering great advantages for the modeling of complex systems thanks to the simplicity of its structure that consists of a single neuron. While in the ANFIS model, it is necessary to change the number of layers, the number of neurons in each layer, and the membership function to find the structure that gives a good fit for the problem to model. The NFN and ANFIS models can be used for short-term prediction and as we have observed in the experimental part, for long-term prediction. The artificial ants algorithm makes possible to highlight classes grouping similar observations. This improves classification and allows the identification of the degradation states. A comparative study, not presented in this paper, was done to evaluate the performance of several classification algorithms (K-means, hidden Markov models, ANFIS, multi-layer feed-forward network, ...) in the 'IRIS' problem. 'IRIS' corresponds to a database used to test the classification models. The IRIS problem is about dividing the set of iris flowers in different groups based on the flower features. Those features are the length and width of a sepal and the length and width of a petal. The obtained results showed that the artificial ants algorithm obtained the best classification results.

References

[1] P. F. ALBRECHT, J. C. APPIARIUS, R. M. MCCOY, E. L. OWEN AND D. K. SHARMA, Assessment of the reliability of motors in utility applications - updated. *IEEE Transactions on Energy Conversion EC-1*, 1 (1986), 39–46.

[2] I. C. Report, Report of large motor reliability survey of industrial and commercial installations, part i. *IEEE Transactions on Industry Applications IA-21*, 4 (July 1985), 853–864.

[3] G. VACHTSEVANOS, F. LEWIS, M. ROEMER, A. HESS AND B. WU, *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. John Wiley & Sons, 2006.

[4] A. K. MAHAMAD, S. SAON, AND T. HIYAMA, Predicting remaining useful life of rotating machinery based artificial neural network. *Computers & Mathematics with Applications 60*, 4 (2010), 1078 – 1087.

[5] Y. LI, T. KURFESS AND S. LIANG, Stochastic prognostics for rolling element bearings. *Mechanical Systems and Signal Processing 14*, 5 (2000), 747 – 762.

[6] F. DALVAND, M. KANG, S. DALVAND, AND M. PECHT, Detection of generalized-roughness and single-point bearing faults using linear prediction-based current noise cancellation. *IEEE Transactions on Industrial Electronics 65*, 12 (2018), 9728–9738.

[7] L. CAO, Z. QIAN AND Y. PEI, Remaining useful life prediction of wind turbine generator bearing based on emd with an indicator. In *Prognostics and System Health Management Conference* (2018), 375–379.

[8] A. SOUALHI, K. MEDJAHER AND N. ZERHOUNI, Bearing health monitoring based on hilbert-huang transform, support vector machine, and regression. *IEEE Transactions on Instrumentation and Measurement 64*, 1 (2015), 52–62.

[9] X. JIN, Z. QUE, Y. SUN, Y. GUO AND W. QIAO, A data-driven approach for bearing fault prognostics. In *IEEE Industry Applications Society Annual Meeting* (2018), 1–8.

[10] R. LIU, B. YANG AND A. G. HAUPTMANN, Simultaneous bearing fault recognition and remaining useful life prediction using joint loss convolutional neural network. *IEEE Transactions on Industrial Informatics* (2019), 1–1.

[11] J. Sikorska, M. Hodkiewicz and L. Ma, Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing 25*, 5 (2011), 1803 – 1836.

[12] A. SOUALHI, G. CLERC, H. RAZIK AND F. RIVAS, Long-term prediction of bearing condition by the neo-fuzzy neuron. In *9th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives* (2013), 586–591.

[13] C. Sbarufatti, M. Corbetta, M. Giglio and F. Cadini, Adaptive prognosis of lithium-ion batteries based on the combination of particle filters and radial basis function neural networks. *Journal of Power Sources 344*, (2017), 128 – 140.

[14] Z. Tian, T. Jin, B. Wu, and F. Ding, Condition based maintenance optimization for wind power generation systems under continuous monitoring. *Renewable Energy 36*, 5 (2011), 1502 – 1509.

[15] Q. Miao, L. Xie, H. Cui, W. Liang and M. Pecht, Remaining useful life prediction of lithium-ion battery with unscented particle filter technique. *Microelectronics Reliability 53*, 6 (2013), 805 – 810.

[16] K. KAPLAN, M. KUNCAN AND ERTUNÇ, H. M. Prediction of bearing fault size by using model of adaptive neuro-fuzzy inference system. In *23nd Signal Processing and Communications Applications Conference* (2015), 1925–1928.

[17] P. DANG, H. ZHANG, X. YUN AND H. REN, Fault prediction of rolling bearing based on arma model. In *International Conference on Computer Systems, Electronics and Control* (2017), 725–728.

[18] G. YUHAI, L. SHUO, H. LINFENG AND W. LIYONG, Research on failure prediction using dbn and lstm neural network. In *57th Annual Conference of the Society of Instrument and Control Engineers of Japan* (2018), 1705–1709.

[19] D. A. TOBON-MEJIA, K. MEDJAHER, N. ZERHOUNI, AND G. TRIPOT, Hidden markov models for failure diagnostic and prognostic. In *Prognostics and System Health Managment Confernece* (2011), 1–8.

[20] B. WIDROW AND S. D. STEARNS, *Adaptive Signal Processing*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1985.

[21] M. Hu, *Application of the Adaline System to Weather Forecasting*. Department of Electrical Engineering, Stanford University., 1964.

[22] R. GOURIVEAU, M. EL KOUJOK AND N. ZERHOUNI, Spécification d'un système neuro-flou de prédiction de défaillances à moyen terme. In *Rencontres Francophones sur la Logique Floue et ses Applications* 1 (2007), 65–72.

[23] T. TAKAGI AND M. SUGENO, Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics 15*, 1 (1985), 116–132.

[24] J. S. R. JANG, Anfis: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics 23*, 3 (1993), 665–685.

[25] T. YAMAKAWA, E. UCHINO, T. M., AND KUSANAGI, H. A neo fuzzy neuron and its applications to system identification and prediction of the system behaviour. In *2nd Int. Conf. Fuzzy Logic Neural Networks* (1992), 477-483.

[26] M. MACKEY AND L. GLASS, Oscillation and chaos in physiological control systems. *Science (New York, N.Y.) 197* (1977), 287–9.

[27] K. DAIJIN, AND K. CHULHYUN, Forecasting time series with genetic fuzzy predictor ensemble. *IEEE Transactions on Fuzzy Systems 5*, 4 (1997), 523–535.

[28] H. QIU, J. LEE, J. LIN AND G. YU, Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *Journal of Sound and Vibration 289*, 4 (2006), 1066 – 1090.