

Weighted Feature Selection with Growing Neural Networks for the FDD of Rolling Element bearings

M. Barakat, M. El Badaoui, and F. Guillet

Abstract—This paper suggests an automated approach for fault detection, diagnosis and identification of roller bearings, which is based on optimized form of growing neural networks. In our recent work, we selected features according to their classification accuracy within supervised learning stage. Since each one of selected features has different effect on classification decision, a weighted feature selection is put forward in this paper to improve the network taxonomy. This is followed by a self adaptive growing neural network that optimizes its architecture by adding or updating hidden nodes to fulfill the training requirements. This pattern recognition procedure used to recognize between signals coming from normal bearings and those generated from different industrial bearing faults. The developed approach is compared with two different types of supervised neural networks. Results demonstrate that the developed diagnostic approach can reliably separate different bearing fault conditions at various rotational speeds.

I. INTRODUCTION

Rotating machinery covers a broad range of machines and plays an important role in industry [1].

This paper is based on FDD of bearing faults using signal analysis. The objective of signal analysis is the discovery of discriminative features that allow the identification of problems in their early stages. Fault diagnosis is conducted typically in the following phases: data collection, feature extraction, selection and classification. Feature extraction is a process that extracts a set of new features from the original data into a distinct feature space [2]. In feature extraction, the time-domain vibration signal is useful to the extent of finding out the overall vibration level, for many cases, the time signals is sufficient to diagnose the health of a machine [3]. Sometimes, time signals may not exactly indicate the defect. In that case, the spectral domain obtained from Fourier analysis is used to extract frequency-domain features from measured vibration signals [4]. Generally, a large number of features are preferred for effective condition

monitoring [5]. However, this would affect the training cost and time, as well as the classification accuracy due to the presence of irrelevant or redundant features [5, 6]. So some salient features need to be selected. At present, there are some feature selection methods such as genetic algorithms [7], principle component analysis [8] etc.

At classification stage, Artificial Neural Networks (ANNs) have become the outstanding method in the recent decades exploiting their non-linear pattern classification properties, offering advantages for automatic detection and identification of machine faults. Several types of neural networks [9] have been applied in machinery fault diagnosis including a Feed Forward Neural Network (FFNN), Radial Basis Function (RBF) networks, etc. However, these approaches have limitations on generalization of results in models that can over-fit the data [10] so there is a need to use approaches with faster classification process and more accurate using the minimum number of features which primarily characterize the system conditions with an optimized structure.

Self Adaptive Growing Neural Network (SAGNN) [11] uses growing architecture in its advanced optimized learning. Its learning algorithm is incremental and has adaptive characteristics that modify the network topology during the learning process. An advanced feature selection algorithm is embedded in the network in which good features are selected [12]. Since selected features have different influences on the output classification decision, we propose to weight them according to their classification accuracy so parameters with high classification accuracy will be weighted more than those of less classification ones and thus have more effectiveness on the output decision.

The proposed algorithm offers an automatic detection and diagnosis of faults in ball bearings without the need to estimate the specific fault frequencies. To validate the classification accuracy of this technique, we compared it with other approaches. These approaches contain the following tools PCA, FFNN and RBF neural networks because they are suitable in diagnosing machinery fault [13].

Different classification approaches are applied to the fault detection, diagnosis and identification problem of machinery system with industrial bearing faults. Several impact factors are tested to evaluate the obtained results and the behavior of the selected parameters in both time and spectral domains are studied at different rotational speed.

Manuscript received January 5, 2012. This work was supported by Rhone Alpes regional council and Jean-Monnet Saint-Etienne University.

M. Barakat is with the university of Jean-Monnet Saint Etienne campus Roannais, (Phone: + 33 (0)4 77 44 81 50; fax: + 33 (0)4 77 44 89 21; e-mail: mustapha.barakat@hotmail.com; mustapha.barakat1@gmail.com).

M. El Badaoui is with the university of Jean-Monnet Saint Etienne campus Roannais, e-mail: mohamed.elbadaoui@univ-st-etienne.fr

F. Guillet is with the university of Jean-Monnet Saint Etienne campus Roannais, e-mail: guillet@univ-st-etienne.fr

I. DEVELOPED TECHNIQUE

A. Feature Extraction

Feature extraction can be viewed as finding a set of vectors that represents an observation while reducing the dimensionality. In the present study, statistical method is employed to investigate the characteristic of the system by extracting 10 statistical parameters from time domain and other 27 parameters from spectral domain. The time domain parameters are: Mean variance, kurtosis, skewness, root mean square, peak value, crest factor, clearance factor, impulse factor and shape factor. The spectral parameters are: spectral power, mean frequency, spectral kurtosis, dissymmetry coefficient, relative energy per frequency band (10 parameters), median frequency, percentiles (10 parameters), peak factor and entropy.

B. Feature Selection

Feature selection is a process by which a sample in the measurement space is described by a finite and usually smaller set of number classed features. The features become components of the pattern space and feature selection is regarded as a procedure to determine which variables are to be measured either first or last according to chosen algorithm. There are many potential benefits of feature selection: facilitating data visualization and data understanding, reducing the measurement and storage requirements, reducing training and time consumption, and defying the curse of dimensionality to improve prediction performance. Those are the main reasons why PCA algorithm is used as a selective tool to compare it with weighted parameter selection in Self Adaptive Growing Neural Network (SAGNN).

C. Weighted Feature Selection with SAGNN

Weighted Feature selection with Self Adaptive Growing Neural Network (SAGNN) is an advanced classification approach based on the concepts of neural networks. A feature selection algorithm was embedded in SAGNN in which only parameters that are good for classification are selected. Selected features are weighted in various values depending on tested classification accuracy factor that indicates the classification sensitivity of every selected feature on different classes. Moreover this technique is able to specify the appropriate features for each class whether it is nominal or faulty.

SAGNN Classifier

SAGNN is an input-output pair classifier. It consists of three feed forward hidden layers. The input layer receives a vector $X = (x_j) \in IR^{Q_{max}}$ of Q_{max} parameters computed from an input signal after DWT decomposition. The hidden layers F_k include v_{ki} normalized Gaussian functions for the parameters $j = 1, \dots, Q_{max}$, and for classes $k = 1, \dots, P$. The output layer $G \in IR^P$ provides the output decision. SAGNN

proposes winning zones according to y_{jlim} and σ_{jlim} that prevent the multiplication of numerous Gaussian neurons in hidden layers.

The activation of the associated hidden layer is computed according to (1) [11]:

$$F_k(X) = \left[f_{kjv}(x_j) = \frac{1}{\sigma_{kjv} \sqrt{2\pi}} e^{-\frac{(x_j - c_{kjv})^2}{2\sigma_{kjv}^2}} \right] \quad (1)$$

Local cost functions are defined to check if training parameters lie on winning zones. They are related to hidden layer functions as in (2).

$$Cost_{kj} = 1 - \frac{\max_{v=1, \dots, v_{kj}} \{f_{kjv}(x_j)\}}{y_{jlim}} ; \quad j = 1, \dots, Q_{max} \quad (2)$$

When $Cost_{kj} \geq \varepsilon$ then an additional neuron is created in the associated hidden layer, otherwise the winning neuron is updated.

Let M be the training set. When all data $X \in M$ have been processed, a global test of three stopping criteria is used to decide to continue or not adaptation and growing of the hidden layers.

- 1- $cost_error_{kj} > \varepsilon$ where ε is an arbitrary threshold related to the admissible error in the Gaussian nodes location and $cost_error_{kj}$ is defined in (3).

$$cost_error_{kj} = 1 - \frac{1}{\hat{W}_k} \sum_{X \in M} \left(\frac{\max_{v=1, \dots, v_{kj}} \{f_{kjv}(x_j)\}}{y_{jlim}} \right) \quad (3)$$

- 2- $v_{kj} \geq N_{lim}$ where N_{lim} is the maximum number of nodes used for class k and parameter j .
- 3- $iter = iter_max$ with $iter$ the epochs number (i.e. number of presentation of M to the network) and $iter_max$ the maximum acceptable number of epochs.

If criterion 1, 2 or 3 is verified for parameter j in class k , the adaption process is completed for this parameter in this class. If none of the criteria is satisfied then another epoch is carried out. The learning stage is ended when one of the criteria is verified for all the parameters for all classes [11].

Feature Selection with SAGNN

The procedure for feature selection is usually based on two basic components: a search algorithm and an evaluation criterion. Accuracy testing approach and performance factor are used as search algorithm and evaluation criterion respectively. The selection process starts after adapting and

training each type of extracted parameters in SAGNN classifier. Then an accuracy testing factor γ_{jl} is computed and tested for each class. l is an index for the tested class and $j = 1, \dots, Q_{\max}$ is the index of an extracted feature where Q_{\max} is the total number of parameters. The accuracy testing factor is defined as: [11]

$$\gamma_{jl} = \frac{\sum_{X \in M_l} \max_{k=1..P} \left\{ \max_{v=1..V_k} \{f_{kiv}(x_{lij})\} \right\}}{\sum_{X \in M_l} \max_{v=1..V_d} \{f_{liv}(x_{lij})\}} \quad (4)$$

Parameters are arranged from more to less desirable according to their gamma values and the performance rate (P_R) which is the evaluation factor is computed: [12]

$$P_R = (C_{R_j})^2 \times C_{F_j} \quad (5)$$

where C_{R_j} and C_{F_j} are the classification rate and confidence factor respectively. Starting with the most desirable parameter, the classification process of training samples is activated for all classes and the average performance rate is computed. Another desirable parameter is added, the classification process started again and the average performance rate is also calculated. This process is repeated for all parameters starting with most desirable and ending with less desirable parameter. The optimum solution will correspond to highest value of average performance rate. At this stage parameter selection is achieved, and their corresponding receptors are reserved. At the end of training and selecting process, the new approach is able to: (1) select the suitable parameters for classification and (2) identify the number and positions of neurons for each selective parameter.

Classification Stage with Weighted Outputs

Let $X = (x_j)$, $j = 1, \dots, Q^*$ be a new vector that includes the winning parameters of a testing signal to be classified where Q^* is the total number of winning parameters. During growing, selecting and adapting process of SAGNN, parameters are selected and the number of neurons are specified and localized in subspaces of different classes $k = 1, \dots, P$. The output decision vector $g_k(X)$, $k = 1, \dots, P$ of the testing signal X is calculated using (6):

$$g_k(X) = \sum_{j=1}^{Q^*} w_j * \frac{\max_{v=1..V_{kj}} \{f_{jv}(x_j)\}}{y_{j\lim}} \quad (6)$$

where $w_j = 1 - \min_{l=1..P} \{\gamma_{jl}\}$ is a weight coefficient based on accuracy testing factor. High weights correspond to efficient

testing factors so they have more influence on the output decision of the network. They are computed during training stage and used to weight the decisions in the classification process.

The output vector $g_k(X)$ represents the sensitivity and effectiveness of receptors (Gaussian nodes) in receiving samples of different parameters. It is to say that the receptors that are most sensitive to a given sample correspond to the winning class and so the class of sample X is given according to (7):

$$\xi = \arg \max_{k=1, \dots, P} \{g_k(X)\} \quad (7)$$

II. EXPERIMENTAL MANIPULATION AND DATA CLASSIFICATION

The test rig (fig. 1) consists of a variable speed electric DC motor driving a shaft rotor with cone shaped end attached to roller bearing. An accelerometer sensor is mounted on top of the bearing's house. The selected area is made smooth to ensure effective coupling. A shock absorber tape was located beneath the mechanical elements to avoid the external vibrations.

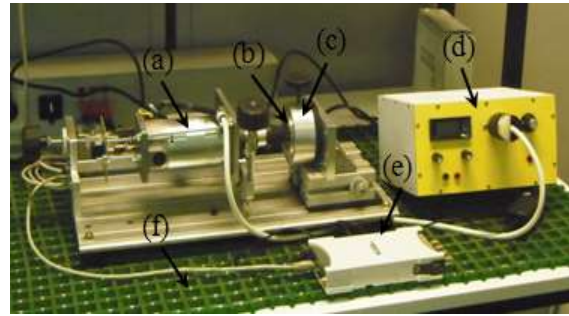


Fig. 1. Experimental machinery system, (a) electric motor, (b) tested bearing, (c) sensor position, (d) electric generator, (e) acquisition card and (f) shock absorber tape.

A variable speed DC motor (0.25 hp) with speed up to 6000 rpm is used as the basic drive. A short shaft is attached to the motor through a flexible coupling. An accelerometer of model 353815 with 10.29 mV/g sensitivity is connected to National Instruments card (NI 9234) to collect the vibration signals of 6303 Z/MT33 bearings. Signals are acquired by accelerometer collected at the card and reserved in the computer. All the computations are done on computer of 1.7 GHz frequency processor with 8 CPUs and the programming parts are carried out on Matlab of version 2011.

The motor setup is made to rotate at six different rpm speeds (1200, 1500, 1800, 2100, 2400, and 2700) in order to study their possible influences on the bearing system. The bearing faults utilized in this experiment are of industrial origin and not artificial. They include a defective outer-race and a defective roller. At every rotational speed, 100 samples are collected for each type of bearings (Nominal or defective). The duration of each sample is one second and its

sampling frequency is 12800. A total of 1800 samples are collected for all types (Nominal and defective) at different rotational speeds. A nominal sample in addition to faulty samples (outer and roller fault) at 1800 rpm are displayed in fig. 2.

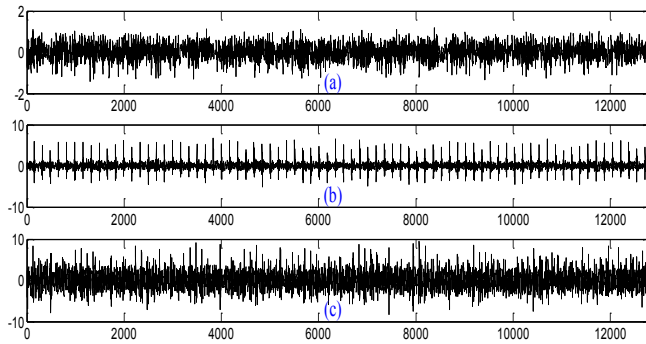


Fig. 2. (a) Nominal, (b) Outer race fault and (c) Roller race fault bearing samples at 1800 rpm.

In fault identification, faults are classified at six different rotational speeds for each of the three classes (normal, outer race and roller race bearing faults).

Parameters in both time and frequency domains are investigated. At preprocessing stage, 10 temporal and 27 spectral parameters are extracted from training data samples. Parameters from training data are selected using PCA technique and the selected parameters are used to train the two networks (FFNN and RBF) respectively. In weighted parameter selection with SAGNN, parameters are selected in an embedded process corresponding to maximum performance rate. The total number of collected training samples is 450 (25 samples at each rpm for each class). Another 1350 testing samples (75 samples at each rpm for each class) are classified and several testing factors are applied to compare and evaluate the results.

TABLE I
MOST SENSITIVE PARAMETERS CORRESPONDING TO THREE BEARING CLASSES

<i>rpm</i>	<i>Normal</i>	<i>Outer race fault</i>	<i>Roller race fault</i>
1200	Dissymmetry coefficient	Dissymmetry coefficient	Median frequency
1500	Dissymmetry coefficient	Kurtosis	8 th Relative energy band
1800	5 th Decil	Root mean square	2 nd Decil
2100	7 th Decil	Shape factor	Spectral power
2400	Spectral kurtosis	10 th Relative energy band	Variance
2700	Dissymmetry coefficient	Shape factor	3 rd Decil

Table I illustrates that various parameters are selected at different rotational speeds for the same class. This insures that the most sensitive parameter for a certain class changes according to rotational speed of the machine. Results show that parameters behave in different way depending on type of fault and speed of rotation. Dissymmetry coefficient from

A. FDD using Weighted Feature Selection with SAGNN

37 temporal and spectral parameters ($Q = 37$) are extracted from the measured training samples and used in selection, adaptation and training of the SAGNN. $P = 18$ is the number of output classes (6 Normal + 6 outer race fault + 6 roller race fault).

The maximum number of neurons used for achieving the training process of each parameter is 70 nodes. The global cost error threshold is $\epsilon = 0.001$ in order to assure that all parameters are laid in the winning zones within this error value. The coefficient which determines how the limited Gaussian functions are peaked or flat relative to maximum wideness vector D is $\alpha = 0.05$. In addition the epochs number or the maximum accepted iteration number used is $iter_max = 10$. At this stage, the coefficients of the network are identified and the growing and adaptation process in the training step is initiated.

During the training process, gamma matrix is computed. According to gamma matrix, the algorithm is able to specify if a certain parameter is suitable to distinguish a certain class or not. After that, parameters are arranged and listed form more to less desirable according to their gamma values. Starting with the most desirable parameter, the performance rate is computed. Another desirable parameter is added to the process and the performance rate is also calculated. This process is repeated for all parameters starting with most desirable and ending with less desirable one. Parameters of highest performance rate correspond to optimum solution. 14 parameters out of 37 are selected. The selected parameters correspond to optimum solution and are the only parameters used in classification.

According to testing accuracy factor, the parameters that are most sensitive to each one of the 18 classes and though having the ability to separate them precisely are summed up in table I.

spectral domain is a good parameter for fault detection decision. 4 out of 10 decil coefficients are selected; this reflects how the accuracy of decils is.

4 temporal parameters (Shape factor, kurtosis, variance, and root mean square) and other 10 spectral parameters $Q^* = 14$ are selected for classification. The weighted

coefficients provide weight for each selected parameter according to its accuracy testing factor and thus affect the

output classification decision. The weights for selected parameters are displayed in table II.

TABLE II
OUTPUT WEIGHTS OF WINNING PARAMETERS

Parameters	Dissymmetry coefficient	Kurtosis	Median frequency	10 th Relative energy band	3 rd Decil	5 th Decil	Variance
Weights	1	1	1	1	0.99	0.9	0.87

Parameters	Root mean square	Shape factor	2 nd Decil	Spectral power	Spectral kurtosis	7 th Decil	8 th Relative energy band
Weights	0.87	0.85	0.8	0.78	0.68	0.57	0.55

The fourteen selected parameters are extracted from the testing samples of machinery system for classification. Testing samples of 18 different classes are classified by

weighted parameter selection with SAGNN approach of weighted outputs and classification results are displayed in table III.

TABLE III
CONFUSION MATRIX OF THE 18 CLASSES USING WEIGHTED PARAMETER SELECTION WITH SAGNN

	rpm	Normal						Outer						Roller					
Normal	1200	75	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1500	0	75	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1800	0	0	75	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2100	0	0	0	14	5	56	0	0	0	0	0	0	0	0	0	0	0	0
	2400	0	0	0	0	72	3	0	0	0	0	0	0	0	0	0	0	0	0
	2700	0	0	0	0	0	75	0	0	0	0	0	0	0	0	0	0	0	0
Outer	1200	0	0	0	0	0	0	75	0	0	0	0	0	0	0	0	0	0	0
	1500	0	0	0	0	0	0	0	75	0	0	0	0	0	0	0	0	0	0
	1800	0	0	0	0	0	0	0	0	75	0	0	0	0	0	0	0	0	0
	2100	0	0	0	0	0	0	0	0	0	75	0	0	0	0	0	0	0	0
	2400	0	0	0	0	0	0	0	0	0	0	75	0	0	0	0	0	0	0
	2700	0	0	0	0	0	0	0	0	0	0	0	75	0	0	0	0	0	0
Roller	1200	0	0	0	0	0	0	0	0	0	0	0	0	72	3	0	0	0	0
	1500	0	0	0	0	0	0	0	0	0	0	0	0	0	75	0	0	0	0
	1800	0	0	0	0	0	0	0	0	0	0	0	0	0	0	70	5	0	0
	2100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	64	10	0
	2400	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	73	0
	2700	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	70

A total of 90 samples are misclassified. All fault samples are well detected and well diagnosed however some roller element bearing faults are misidentified. Weighted parameter selection with SAGNN offers the best classification results in comparison to the previous mentioned approaches.

B. Results evaluation and discussions

In this section, several impact factors are figured out to analyze the classification results and understand their classification accuracy. These factors are used to evaluate the classification results obtained by the three different

classification approaches (PCA-FFNN, PCA-RBF and weighted parameter selection with SAGNN).

Five impact factors are chosen to recognize the performance of the three approaches.

- 1- Non detection occurs when the fault is not detected.
- 2- False alarm happens when machine operates normally and false alarm is generated.
- 3- Diagnosis error takes place if the fault is detected but misdiagnosed.
- 4- Identification error arises if the fault is well detected and diagnosed but misidentified.

5- Classification rate is the percentage of well classified training samples.

Obtained results are exhibited in table IV.

TABLE IV
IMPACT FACTORS OF DIFFERENT TECHNIQUES

Classifiers	Non detection	False alarm	Diagnosis error	Identification error	Classification rate
<i>PCA-RBF</i>	0.11 %	0 %	0.88 %	6.44 %	87.62 %
<i>PCA-FFNN (10, 18)</i>	0 %	0 %	2.55 %	11.55 %	83.33 %
<i>PCA-FFNN (10, 16, 18)</i>	0.88 %	0 %	5.66 %	10.55 %	80.37 %
<i>Weighted parameter selection with SAGNN</i>	0 %	0 %	0 %	2.88 %	93.33 %

Table IV shows that weighted parameter selection with SAGNN offers best results in non detection, false alarm, diagnosis error, identification error and classification rate.

For the three techniques, no false alarm is generated. PCA-RBF provides better classification rate from PCA-FFNN of one and two hidden layers architecture however PCA-FFNN (10, 18) has a preference in detecting all faults.

The major drawback of the mentioned techniques is that they must be trained on all fault scenarios and hence if an unexpected fault occurs it will be necessary to train the classifier with new data. A new class can be easily trained in weighted parameter selection with SAGNN without altering the previous data because this classifier reserves a hidden sub-layer for each class and hence an extra hidden sub-layer can be added to the network.

I. CONCLUSION

In this paper, a weighted parameter selection with Self Adaptive Growing Neural Network (SAGNN) is developed to detect, diagnose and identify bearing faults of industrial nature. Selected features are weighted according to their testing accuracy factor that relates the classification accuracy of desired class with the rest classes. This technique is compared with two neural network classifiers, (1) Radial Basis Function (RBF) and (2) Feed Forward Neural Network (FFNN) in which Principal Component Analysis (PCA) is used to reduce the dimension of extracted features at the preprocessing stage of the compared techniques.

The three techniques are applied to machinery system of industrial bearing faults. Several impact factors are used to check the detection, diagnosis and identification accuracy of these approaches. Results show that the proposed technique offers the best classification performance and thus it is a powerful classification tool that optimizes the number of selected features and the number of neurons related to each selected one. Moreover this approach is capable to recognize the parameters that are most sensitive to each of target classes.

Second aim from this paper was to study the behavior of parameters in time and frequency domains at different rotational speeds. Classification results show that if a

parameter is most sensitive to a certain class at a specific rpm then this parameter may not remain the most sensitive when the rotational speed changes.

REFERENCES

- [1] Han, T, Yang, BS, Choi, WH, and Kim, JS, "Fault diagnostic system of induction motors based on neural network and genetic algorithm using stator current signals," *International Journal of Rotating Machinery*, pp.1-13, 2006.
- [2] N. Wyse, R. Dubes, and A.K. Jain, "A critical evaluation of intrinsic dimensionality algorithms," *Pattern Recognition in Practice* 1980, pp. 415-425.
- [3] B. Samanta, K. R. Al-Balushi, and S. A. Al-Araimi, "Artificial neural networks and genetic algorithm for bearing fault detection," *Soft Computing*, vol.10, 2006, pp. 264-271
- [4] B. Samanta, and K. R. Al-Balushi, "Artificial neural network based fault diagnostics of rolling element bearings using time domain features," *Mechanical Systems and Signal Processing*. vol.17, (2003), pp. 317-328
- [5] A. Malhi, R. Gao, PCA-based feature selection scheme for machine defect classification, *IEEE Transaction on Instrumentation and Measurement* 53, 2004, pp.1517-1525.
- [6] A. Widodo, B.-S. Yang, and T. Han, "Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors," *Expert Systems with Applications* 32 (2), 2007, pp. 299-312.
- [7] Samanta, B, "Artificial neural networks and genetic algorithms for gear fault detection," *Mechanical Systems and Signal Processing*, 2004, 18 , pp. 1273-1282.
- [8] L. Liu, J. Kang, J. Yu, Z. and Wang, "A comparative study on unsupervised feature selection methods for text clustering," *Proceeding of NLP-KE'05*, 2005, pp. 597-601.
- [9] Jang, J.-S.R., C.-T. Sun, and E. Mizutani., "Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence. MATLAB curriculum series," *Upper Saddle River, NJ: Prentice Hall*. xxvi, 1997, pp. 614.
- [10] B. Samanta, K.R. Al-Baulshi, and S.A. Al-Araimi, "Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection," *Engineering Applications of Artificial Intelligence* 16, 2003, pp.657-665.
- [11] Barakat M., Druaux F., Lefebvre D., Khalil M., and Mustapha O., "Self Adaptive Growing Neural Network classifier for Faults Detection and Diagnosis," *Neurocomputing journal Elsevier*, 74 2011, pp. 3865-3876.
- [12] Barakat M., Lefebvre D., Khalil M., Druaux F., and Mustapha O., "Parameter Selection Algorithm with Self Adaptive Growing Neural Network Classifier for Diagnosis Issues," *International journal of machine learning and cybernetics*, 2012, DOI: 10.1007/s13042-012-0089-5.
- [13] Baillie, D. and J. Mathew, "Diagnosing Rolling Element Bearing Faults with Artificial Neural Networks," *Acoustics Australia*, 1994, pp. 79-84.