Mechanical Systems and Signal Processing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Fault diagnosis in spur gears based on genetic algorithm and random forest

Mariela Cerrada [a,b,*], Grover Zurita [a,c], Diego Cabrera [a], René-Vinicio Sánchez [a], Mariano Artés [d], Chuan Li [e,a]

[a] Mechanical Engineering Department, Universidad Politécnica Salesiana, Cuenca-Ecuador
[b] Control Systems Department, Universidad de Los Andes, Mérida, Venezuela
[c] Electromechanical Engineering Department, Universidad Privada Boliviana, Cochabamba, Bolivia
[d] Mechanical Engineering Department, Universidad Nacional de Educación a Distancia, Madrid, Spain
[e] Chongqing Key Laboratory of Manufacturing Equipment Mechanism Design and Control, Chongqing Technology and Business University, Chongqing, China

A R T I C L E I N F O

Article history:
Received 14 January 2015
Received in revised form
6 June 2015
Accepted 27 August 2015

Keywords:
Gearbox
Fault diagnosis
Genetic algorithms
Random forest
Feature selection
Wavelet packets

A B S T R A C T

There are growing demands for condition-based monitoring of gearboxes, and therefore new methods to improve the reliability, effectiveness, accuracy of the gear fault detection ought to be evaluated. Feature selection is still an important aspect in machine learning-based diagnosis in order to reach good performance of the diagnostic models. On the other hand, random forest classifiers are suitable models in industrial environments where large data-samples are not usually available for training such diagnostic models. The main aim of this research is to build up a robust system for the multi-class fault diagnosis in spur gears, by selecting the best set of condition parameters on time, frequency and time–frequency domains, which are extracted from vibration signals. The diagnostic system is performed by using genetic algorithms and a classifier based on random forest, in a supervised environment. The original set of condition parameters is reduced around 66% regarding the initial size by using genetic algorithms, and still get an acceptable classification precision over 97%. The approach is tested on real vibration signals by considering several fault classes, one of them being an incipient fault, under different running conditions of load and velocity.

1. Introduction

There are constant increasing requirements for the continuous working of the transmission machines. This is why new approaches to building up fault diagnostic systems with accuracy and reliability are highly valuable. There are invaluable studies to detect gear faults by using standard diagnostic techniques, such as Cepstrum Techniques and Envelope Data Analysis with Hilbert Transforms, among other techniques [1]. In recent years, several analysis techniques for gears faults diagnosis have used Wavelet Packet Transform (WPT), in order to enhance the information that is provided by the classical statistical parameters from the vibration signal, in time and frequency domains [2–4]. In case of machine learning-based

* Corresponding author at: Control Systems Department, Universidad de Los Andes, Mérida, Venezuela.
E-mail addresses: mcerrada@ups.edu.ec, cerradam@ula.ve (M. Cerrada), gzuritav@ups.edu.ec (G. Zurita), dcabrera@ups.edu.ec (D. Cabrera), rsanchezl@ups.edu.ec (R.-V. Sánchez), martes@ind.uned.es (M. Artés), chuanli@21cn.com (C. Li).

http://dx.doi.org/10.1016/j.ymssp.2015.08.030
0888-3270/

Please cite this article as: M. Cerrada, et al., Fault diagnosis in spur gears based on genetic algorithm and random forest, Mech. Syst. Signal Process. (2015), http://dx.doi.org/10.1016/j.ymssp.2015.08.030

diagnosis, the most common approaches have been developed by using Neural Networks [5,6], Support Vectors Machines [7], Cluster Analysis [8] and Genetic Algorithms [9,10]. These approaches have been very useful for implementing condition-based maintenance (CBM), as presented in Jardine et al. [11].

Recently, Random Forest (RF), as a classification technique, has been used for fault diagnosis in several areas of engineering because it is a robust approach in case of having a large number of input attributes, a low number of available samples for learning and because the high interpretability of the tree-based models [12,13]. On the other hand, Genetic Algorithms (GA) have also been used for the optimization process in classification problems, one of them being the feature selection. In diagnosis of machinery condition, RF has been used with GA in order to improve the classification accuracy [14,15]. Yang et al. [14] have used GA to optimize two parameters of the RF algorithm: the number of trees and the split variable number at each node of the tree. In combustion engines, Karabadji et al. [15] have applied GA for selecting the best classifier based on trees for faults diagnosis in industrial fans, the outcome of the analysis was to find the classifier which appears most frequently in the population, and the classifier with the best accuracy in classification was also found.

Condition parameters for fault diagnosis are mostly related to statistical measures from signals in time and frequency domains. Moreover, parameters associated to time–frequency domain have important information on machinery condition, and they are also used in order to extend the set of condition parameters that are processed in diagnostic algorithms. These condition parameters are called *the features* in classification problems. Taking into account the availability of a large number of feature candidates for fault diagnosis, the problem of feature selection, after the feature extraction process, is still an open research area in case of using machine learning-based diagnosis. The failure nature can lead to choosing certain condition parameters and, in case of incipient fault, the best condition parameters providing good diagnostic information are not easy to identify. An incipient fault is a fault that is just beginning to show symptoms. Sometimes, minor scratches and wear on the face of the tooth cannot be seen in scheduled maintenance inspections. This is a typical example of incipient fault. This type of fault commonly appears in rotating machinery and it is related to the first stage of severe situations that can lead to loss of the device function. Detection and diagnosis of incipient faults are still research problems that should be improved in real industrial applications. On the other hand, the availability of a large number of samples for classifier training is not usual in real industrial environments, in this case, RF-based models have better performance than other classical classifiers.

This work combines different methods for proposing best features and accurate models for spur gear fault diagnosis. The problem of feature selection is addressed by using GA, and the selected features are used as input attributes for the RF-based fault classifier. Feature selection is an optimization problem where the objective is to minimize certain classification performance metric, then, the major contribution is to have an efficient feature selection mechanism that can lead to a higher diagnostic performance, regarding the use of the whole set of features.

Fig. 1 shows the overview of our approach where the vibration is the analysed signal. After the data acquisition process, the feature extraction is carried out by calculating the most common statistical parameters on time and frequency domain of the vibration signal. Additionally, WPT are applied on the vibration signal to decompose the signal and later extract more condition parameters. Finally, a wide set of condition parameters is collected, called in the following *the attributes* of the

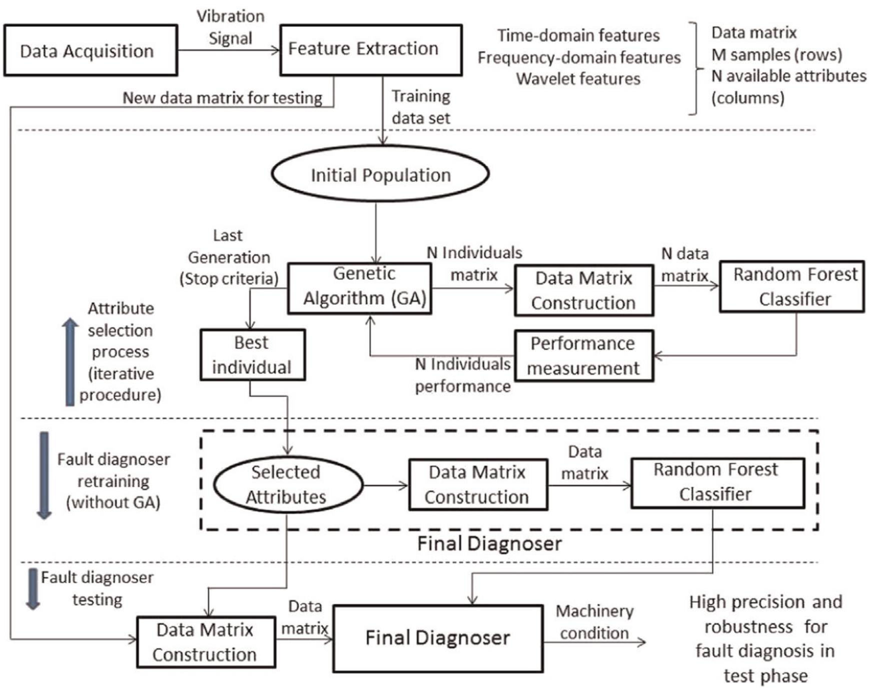


Fig. 1. An overview of the optimization approach for attribute selection in fault diagnosis.

Please cite this article as: M. Cerrada, et al., Fault diagnosis in spur gears based on genetic algorithm and random forest, Mech. Syst. Signal Process. (2015), http://dx.doi.org/10.1016/j.ymssp.2015.08.030

diagnosis problem. The optimization problem is solved in each of the GA cycles, in the sense of searching the subset of attributes that performs the best diagnosis decision by the RF-based classifier. The outcome of this iterative procedure is a subset of attributes explaining the machinery condition, from the information in the available data. Once the best subset is chosen, the RF training is running again, in order to set the final diagnoser. The results in the diagnoser testing, after applying our approach to the new data, show the robustness of the model regarding the high precision in the correct classification of the machinery condition.

Finally, the capabilities of the GA-based selection process are compared with two alternative selection techniques. A simple attribute clustering regarding the correlation between attributes is proposed by using hierarchical clustering based on the *p*-values matrix (non correlation hypothesis test) that is obtained from the correlation matrix for the entire samples. The *p*-values are used as similarity measure. On the other hand, the importance variables that are proposed by the RF algorithm are also used as a selection reference regarding the GA-based selection.

This paper is organized as follows. Section 2 presents the theoretical background supporting the proposed approach. Section 3 details the experimental procedure to collect the data and the feature extraction process. Section 4 states the GA design for features selection. Section 5 discusses the results and compares the selection abilities of our approach with other selection alternatives based on hierarchical clustering and the importance variables that are provided from the RF algorithm. Finally, Section 6 presents the conclusions of this work.

2. Theoretical background

This section presents the conceptual foundations for developing this work. Section 2.1 shows the use of the WPT to decompose a signal into a set of coefficients, this is later used for proposing condition parameters in the diagnosis problem. Section 2.2 presents the techniques of classification based on decision trees including RF. Subsequently, in Section 2.3, the generalities of the GA as a technique for multi-objective optimization are presented, and some aspects for our particular case are outlined, such as the population structure. For further details about the conceptual foundations, see the references.

2.1. Wavelet packet decomposition

The Wavelet Transform (WT) is a powerful tool that has attracted great attention in several fields of the engineering, and particularly, as a powerful analysis instrument for gears fault detection and diagnosis [2–4,16,17]. The analysis presented in Yan et al. [2] provides an extensive overview of some of the latest efforts in the development and applications of WT for fault diagnosis in rotating machinery.

WT mathematically breaks long complicated signals, into smaller ones, and thus facilitate the interpretation of the signals. They are capable of revealing aspects of data that other techniques have missed, such as trends and discontinuities in higher derivative [18]. This technique often can compress or de-noises a signal without any appreciable degradation, and it provides accurate information on the localization of energy content in time and frequency [4]. The WT can be mainly classified as Continuous Wavelet Transform (CWT), Discrete Wavelet Transform (DWT), and Wavelet Packet Transform (WPT). In the WPT framework, compression and de-noising ideas are exactly the same as those developed in the wavelet framework. The only difference is that WPT offers flexible analysis, because the details as well as the approximations of the analysed signal are split [2,19]. In the following, WPT is briefly presented as an alternative for extracting features in fault diagnosis problems.

Let Ψ and ϕ be the selected wavelet function and its corresponding scaling function, given by the Eqs. (1) and (2), respectively, where $g(k)$ and $h(k)$ are the impulse response of the low pass filter ϕ and the high pass filter Ψ, respectively, k is a shift index:

$$\phi(t) = \sqrt{2}\sum_{k} g(k)\phi(2t-k) \tag{1}$$

$$\Psi(t) = \sqrt{2}\sum_{k} h(k)\phi(2t-k) \tag{2}$$

The mentioned filters depend on the wavelet family, and they are also known as Quadrature Mirror Filters (QMF) [19]. The decomposition of a signal by using a wavelet function is accomplished as follows:

Let V_0 be a vector space which is generated by the scaling functions $\phi(t)$ and its corresponding translations $\phi(t-k)$. The vector space V_1 is such that $V_1 \subset V_0$, and its corresponding scaling and translations function are $\phi(2t)$ and $\phi(2t-k)$, respectively. It is possible to move on from V_0 to V_{-1} and, in general, from V_j to V_{j-1} through the vectorial operation (3), where W_j is the orthogonal vector space called Wavelet Space, which is generated from the wavelet function $\Psi(t)$ and its corresponding translations $\Psi(t-k)$:

$$V_{j-1} = V_j \oplus W_j \tag{3}$$

Then, Eq. (3) states that a function which defined on V_{j-1} could be decomposed into a function that belongs to V_j and other function that belongs to W_j, without a loss of information.

Please cite this article as: M. Cerrada, et al., Fault diagnosis in spur gears based on genetic algorithm and random forest, Mech. Syst. Signal Process. (2015), http://dx.doi.org/10.1016/j.ymssp.2015.08.030

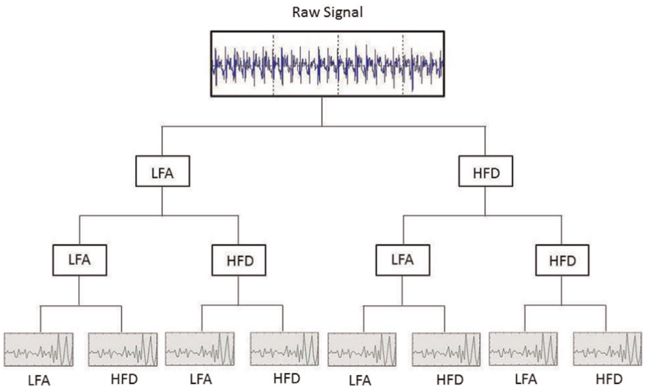


Fig. 2. Wavelet packets decomposition.

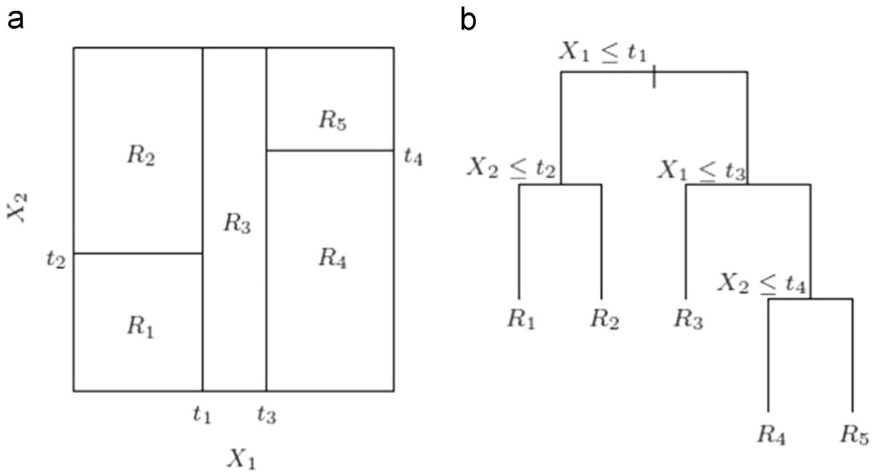


Fig. 3. (a) Binary division of the attributes space and (b) model representation as a binary tree [20].

Let $x(t)$ be the discretized time signal. Then, this signal $x(t)$ is decomposed in the Low Frequency Approximation (LFA) with the filter $\phi(t)$ by using Eq. (4), and in the High Frequency Detail (HFD) with the filter $\Psi(t)$ by using Eq. (5), where $\downarrow 2$ is the sub-sampling operation for the displacement from V_j to V_{j-1}:

$$LFA = (x*\phi)\downarrow 2 \tag{4}$$

$$HFD = (x*\Psi)\downarrow 2 \tag{5}$$

In the wavelet packet decomposition, this process is repeated recursively for each resulting signal LFA and HFD, until obtaining the required level of decomposition. As a result, the raw signal in discrete time domain is decomposed in a tree with the signals LFA and HFD. The wavelets coefficients are the result in the last level of the tree, as illustrated in Fig. 2, with 3 levels. In this work, the wavelet coefficients in level 6 are calculated for the signal vibration and their energies are used as condition parameters for fault diagnosis in order to extend the classical set of condition parameters on time and frequency domains.

2.2. Decision trees and random forest

Decision Tree (DT) is a powerful technique for classification and prediction problems, that is based on the partition of the attribute space, by using a iterative procedure of binary partition providing a highly interpretable model [20–22].

Consider a data set with m samples of a couple $(x_i, y_i), i = 1, \ldots, m$, x_i is a vector of n attributes, i.e., $x_i = (x_{i1}, x_{i2}, \ldots, x_{in})$, and y_i is the logistic classification variable. Let C be an identifier denoting the class C, then, $y_i = C$ indicates that the vector of attributes x_i is associated to the class C. The algorithm for selecting attributes and partition follows a greedy procedure [20]. From all the data, an attribute (variable) j and a partition point s are selected, and the pair of semiplanes R_1 and R_2 is set as described in the following equation:

$$R_1(j,s) = \{X | X_j \leq s\}, \quad R_2(j,s) = \{X | X_j > s\} \tag{6}$$

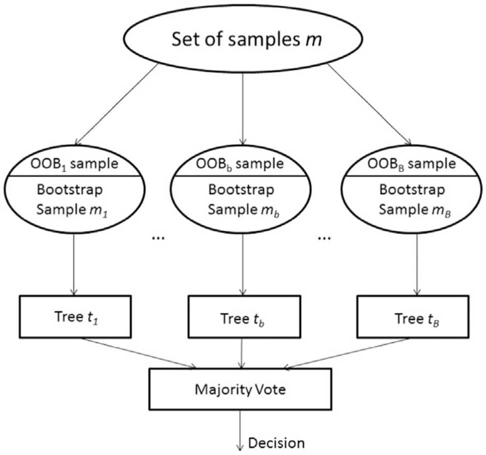


Fig. 4. Structure of a RF.

Let $\hat{p}_{kc}$ be the proportion of observations of class C in the region R_k, regarding the total of N_k observations into this region. The DT algorithm seeks the best selection (j, s) that solves the maximization problem in the following equation:

$$\arg\max_{c} \hat{p}_{kc} \tag{7}$$

where $\hat{p}_{kc} = (1/N_k)\sum x_i \in R_k I(y_i = C)$, I is a membership indicator of the attribute vector to that region. The expression $\hat{p}_{kc}$ is a homogeneity measure of the child nodes, also called the impurity function. Other impurity functions are defined by the misclassification error, the Gini index and the cross-entropy or deviance [12,20].

The iterative procedure splits the attribute space into r disjoints regions R, as far as the stop criterion is reached. The class C is assigned to node k of the tree, which represents the region R_k, that is, $C(k) = \arg\max_c \hat{p}_{kc}$. This procedure searches throughout all possible values of all attributes among the samples.

The classification model is a binary tree, where the mother node represents the initial partition on the domain of the selected attribute, and the child nodes represent successive partitions on the remaining attribute domains to define each region. The leafs denote the regions that best represent the classification model. Fig. 3 shows an example of this iterative procedure on a two-dimensional space of attributes X_1 and X_2, where the successive partitions are defined by the pairs (x_1, t_1), (x_2, t_2), (x_1, t_3), (x_2, t_4). This binary partition is repeated until the best regions are reached, in order to approximate the model output in Eq. (6). The stop criterion could be the number of nodes, or some classification cost. More details of partition rules for selecting the pairs (j, s) can be reviewed in [12,20].

One of the problems of tree-based techniques is the high variance, therefore the bagging technique is applied to improve this issue. The bagged classifier is composed of a set of decision trees which are built from the random subsets of available data samples, and the estimated class is proposed from this set of classifiers. Let c and $f_{bag}(x_i)$ be the class and the classifier proposing the class c for the input x_i, respectively. The bagged classifier selects the class with the largest number of "votes" that are proposed for each classifier $f_{bag}(x_i)$, according to the following equation:

$$\hat{C}_{bag}(x_i) = \arg\max_{c} \hat{f}_{bag}(x_i) \tag{8}$$

where $\hat{C}_{bag}(x_i)$ is the estimated class C for the input x_i, and $\hat{f}_{bag}(x_i)$ is the vector $p_c(x_i)$, that indicates the proportion of estimators proposing the class c.

Random Forest (RF) is a technique based on trees, which proposes a modified approach of the bagging technique, to build a non-correlated trees collection T_b, $b = 1,\ldots,B$, with low bias (low error on the training data) and low variance (low error on the test data) by averaging their predictions [23]. The process to decrease the variance, by reducing the correlation between the trees, is accomplished through the random selection of the input variables and the random selection with replacement of samples from the data set of size m. The selected variables and samples are used to grow every tree in the forest (bootstrap sample). This random selection has shown that around 2/3 of the data are chosen, then, the training set m_b for each classifier is, in general, $m_b \subset m$.

The RF algorithm for the classification problem is summarized in Hastie et al. [20] and Fig. 4 illustrates the structure of a RF according to this algorithm. The complement set $OOB_b = m - m_b$ of each tree T_b is the Out-Of-Bag sample, and it is used as the cross validation set for the tree, i.e, the test set that is used during the process to grow the tree (training process). A performance measure of the RF is the OOB-error that is defined as the average of the classification error associated to each tree T_b using the OOB_b sample. Once the RF is obtained, the decision for classifying a new sample x_i is according to the following equation:

$$\hat{C}^B_{rf}(x_i) = majority\ vote\{\hat{C}_b(x_i)\}^B_1 \tag{9}$$

where $\hat{C}_b(x_i)$ is the class that is assigned by the tree T_b.

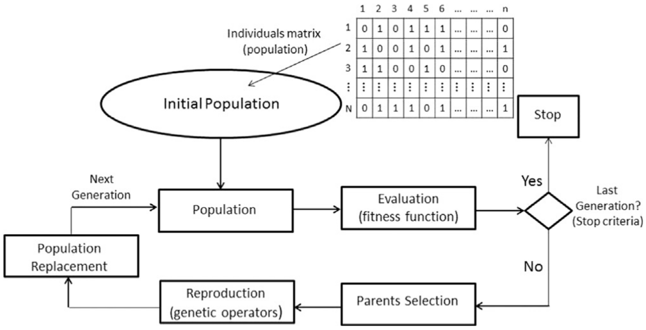


Fig. 5. General cycle of the GA.

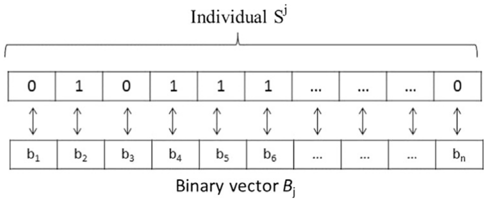


Fig. 6. Individual encoding.

RF is implemented by different packages of software [13,24], and its use in fault diagnosis for rotating machinery has not been widely explored in the literature. This work aims to exploit its properties in this research area.

2.3. Genetic algorithms

Genetic Algorithms (GA) are probabilistic search algorithms that emulate the biological evolution of a population, by applying genetic operators that allow the recombination of the individuals, in order to strengthen their performance with regard to a quality function. GA have been popularly used for solving NP-complete optimization problems, and it constitutes the central paradigm of the Evolutionary Computing (EC) [25]. The basic algorithm that determines the operation of the EC algorithms, including GA, is shown in Fig. 5. In the following, three key components are described in a general manner: individuals encoding, initial population and fitness function. The details about the use of GA in our work are described later in Section 4.

1. *Individuals encoding*: Each individual is defined by the structure and the content. The structure refers to the format and the content refers to information that is provided by the individual in each component of its structure. Most of applications using GA propose a binary encoding, then, an individual S^j is a binary vector B_j of length n, where the numbers 1 or 0 for each b_i are associated to the meaning of the individuals information, see Fig. 6. In our case, binary encoding is used to define if a feature is selected as input attribute for the fault diagnoser. This is explained later in Section 4.
2. *Initial population*: This is composed of individuals of the first generation of possible parents. The initial population P is a set $P = \{S^j\}, j = 1, \ldots, N$. Given the binary representation of the individual S^j, the population is encoded as a binary matrix M where the files $j = 1$,..., N are the randomly selected individuals, as shown in Fig. 5. After starting the GA cycles, the population has the same structure but the content is modified according to the individual evolution.
3. *Fitness function*: The fitness function f, or evaluation function, is a function to measure each individual performance as solution of the optimization problem that is solved by the GA. In general, f is a map, $f: P \to \Re$, that assigns to an individual $S^j \in P$ a real value regarding the optimization problem. In order to avoid the dispersion of the function values, sometimes a fitness scaling is performed [26]. In case of our classification problem, the fitness function will be based on classification metrics such as OOB-error and F-score. The details for defining the fitness and scaling functions are given later in Section 4.

The mechanisms for performing the remaining components are fully discussed in the literature. In case of parent selection, there are different selection mechanisms, most of them based on the relative probability of selection p_j of an individual S^j [27,28]. The most popular genetic operators for reproducing the individuals are the crossover and mutation, however other operators also exist and they can be more appropriate in specific problems [29,30]. Finally, the population replacement strategies must aim to maintain the diversity in the population as well as to improve the evaluation of the fitness function for the individuals in the new population. In this work, direct and elitist replacement are combined [31].

Please cite this article as: M. Cerrada, et al., Fault diagnosis in spur gears based on genetic algorithm and random forest, Mech. Syst. Signal Process. (2015), http://dx.doi.org/10.1016/j.ymssp.2015.08.030

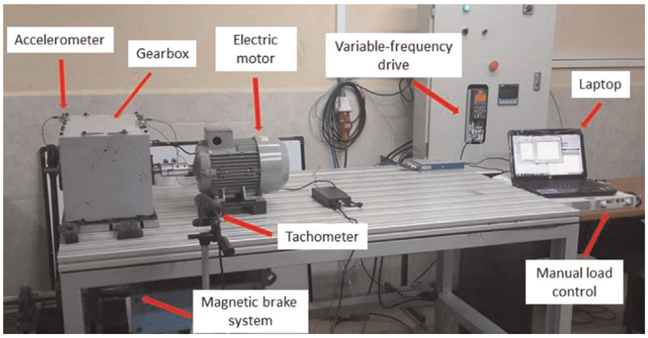


Fig. 7. Vibration analysis laboratory at the Universidad Politécnica Salesiana of Cuenca, Ecuador.

Table 1
Simulated gear faults.

Label	Description
1	Normal
2	Gear tooth breakage 10%
3	Pinion pitting
4	Pinion with face wear 0.5 mm
5	Gear misalignment
6	Gear tooth breakage 50%
7	Gear tooth breakage 100%

3. Measurement procedure and feature extraction

This section details all the experiments for measuring the raw vibration signal and then, the feature extraction process to build the data matrix for the next attribute selection process. The experiments were carried out in the experimental test bed that is shown in Fig. 7.

The rotation motion of the equipment is generated by a 1.1 kW motor powered by three-phase 220 V at 60 Hz with a nominal speed of 1650 rpm. The torque motion is transmitted into a gearbox, where several gear fault configurations are assembled. At the end of the gearbox shaft, the torque is transmitted to a pulley, which is part of the magnetic brake system. The magnetic brake function is to control different loads according to the measurement settings. A variable-frequency drive was used to generate different speeds. Tables 1 and 2 show the details on the different gear faults under study, where the fault number 4 is an incipient fault.

The vibration analyser and digital balancer Digivibe MX-300 were used to collect the data, the data acquisition software was implemented by using the Digivibe MX 5.14. In order to record the vibration signal, the accelerometer with a sensitivity of 330 mV/g was vertically allocated on the gearbox. The measurements were conducted in different speeds (300, 600, 900, 1200 and 1500 rpm) and different break loads about 10%, 50% and 90%, regarding the motor maximum power. The total number of measured signals, with different measurement settings, was up to 1050 signals (7 gear faults, 3 loads, 5 speeds, 10 measurement repetitions for each case on 2 s, with 1 s interval between each repetition). The sampling frequency was 11,025 Hz in order to have the adequate shape of the Fast Fourier Transform (FFT). Each raw data of one measured signal was stored into a PC, and the whole set of 1050 raw signals needs no more than 52.5 MB of memory.

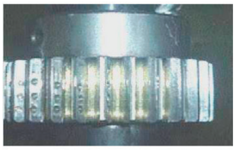
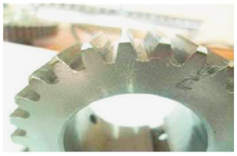
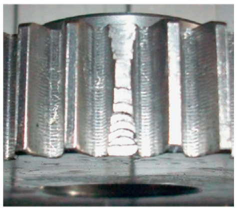
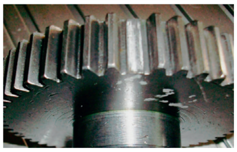
Once the raw signals are stored in the PC, condition parameters are calculated for each raw signal as presented in the following sections where the feature extraction is described. At the end of this data pre-processing, a matrix with 1050 rows (samples) and 359 columns (attributes) is obtained and ready to be post-processed by the GA and RF. The pre- and post-data processing were performed in Matlab®.

Figs. 8 and 9 show some samples of typical time and frequency signals, where they illustrate the normal running conditions and fault conditions that are referenced on Table 1. The frequency signal in normal condition highlights strong energy content between 100 Hz and 1000 Hz. By comparing the frequency signal in normal condition with regard to pinion with face wear or pitting, their spectrum looks similar with slight amplitude differences. This situation points out that a clear classification could be hard to find in the presence of incipient faults.

3.1. *Condition parameters on time and frequency domain*

Nine classical condition parameters were obtained by statistical analysis from each collected raw data on time domain, such as mean, variance, standard deviation, skewness, kurtosis, root mean square (RMS), maximum value, crest factor and energy.

Table 2
Gear faults.

Image	Description
	Gear tooth breakage 10%
	Pinion pitting
	Pinion with face wear 0.5 mm
	Gear misalignment
	Gear tooth breakage 50%
	Gear tooth breakage 100%

Moreover, the FFT was applied for each raw data on time domain, and ten condition parameters were calculated, for each signal on frequency domain, such as mean, variance, skewness, kurtosis, central frequency, root mean square (RMS), standard deviation and other three condition parameters as in Eqs. (10), (11) and (12), that are denoted CP_1, CP_2 and CP_3 [32]:

$$CP1 = \frac{\sum_{k=1}^{K} (f_k - FC)^3 s(k)}{K(STDF)^3} \tag{10}$$

$$CP2 = \frac{STDF}{FC} \tag{11}$$

$$CP3 = \frac{\sum_{k=1}^{K} (f_k - FC)^{1/2} s(k)}{K\sqrt{STDF}} \tag{12}$$

where $s(k)$ is the spectrum for $k=1,\ldots,K$, K is the number of the spectrum lines, f_k is the frequency value of the k-th spectrum line, STDF is the standard deviation for frequency and FC is the frequency center, that are defined in the following equations:

$$STDF = \sqrt{\frac{\sum_{k=1}^{K} (f_k - FC)^2 s(k)}{K}} \tag{13}$$

Please cite this article as: M. Cerrada, et al., Fault diagnosis in spur gears based on genetic algorithm and random forest, Mech. Syst. Signal Process. (2015), http://dx.doi.org/10.1016/j.ymssp.2015.08.030

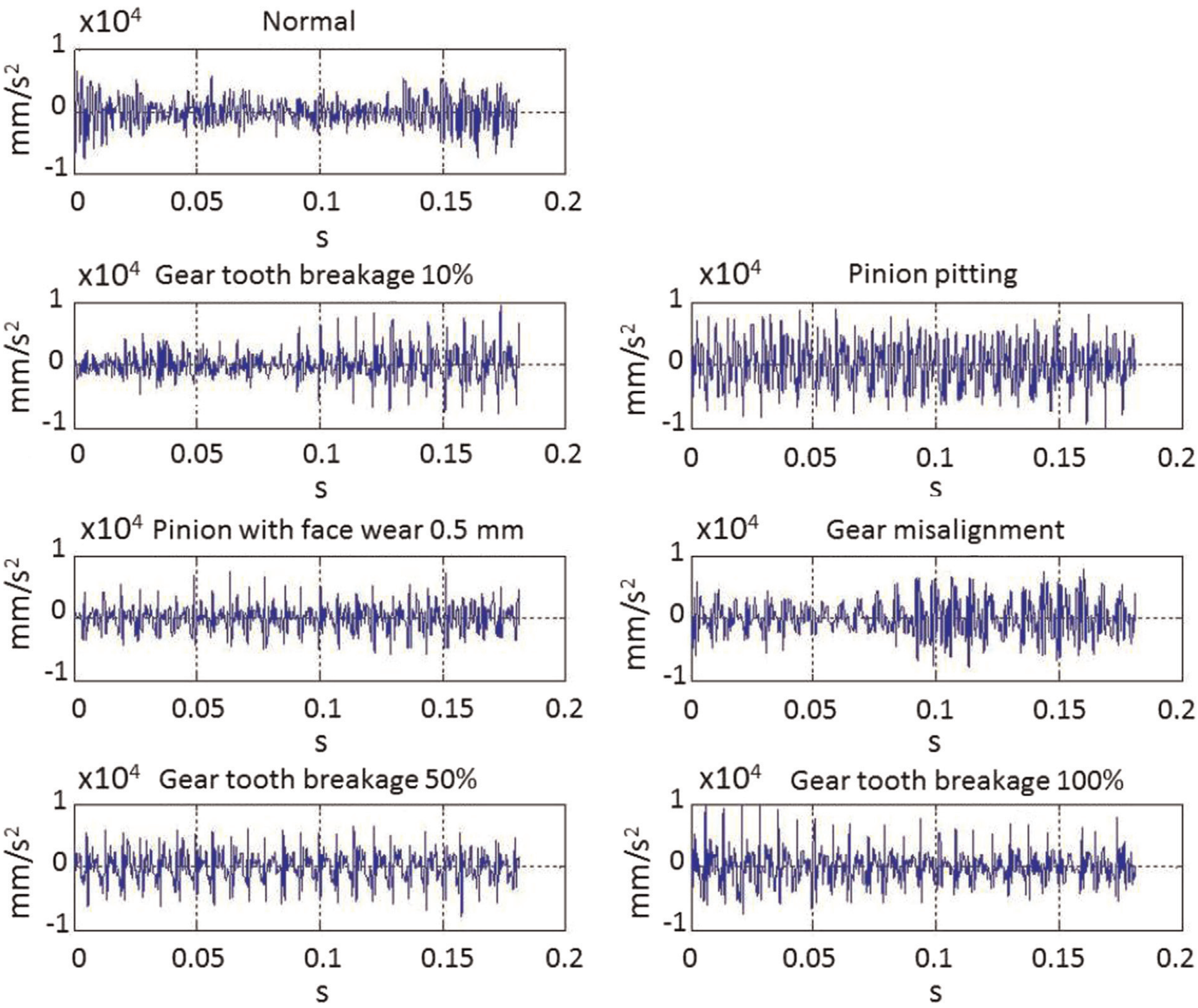


Fig. 8. A sample of time domain signals for the simulated faults in the gearbox.

$$FC = \frac{\sum_{k=1}^{K} f_k s(k)}{\sum_{k=1}^{K} s(k)} \tag{14}$$

Another condition parameter on frequency domain is the RMS value on a frequency band. In all of the experiments, maximum frequencies around 3600 Hz have been reached, then, in addition to the previous frequency parameters, the RMS value was calculated on twenty frequency bands, over the frequency range of 3660 Hz, each band size of 183 Hz. The rationale for using frequency bands is because a fault can generate clear changes on the vibration amplitude in a band where, usually, this amplitude is non-significant in case of no faults.

Then, 39 condition parameters for each signal have been calculated and stored in a data matrix of 1050 rows (samples) and 39 columns (attributes).

3.2. Condition parameters from wavelet packets analysis

The time raw signals were used as input data to the wavelet packet analysis as described in Section 2.1, in order to obtain the energy of each wavelet coefficient. These energies are defined as features in the diagnosis problem. Five mother wavelets are involved in the analysis: Daubechies (db7), Symlet (sym), Coifier (coif4), Biorthogonal (bior6.8) and Reverse Biorthogonal (rbior6.8), these wavelet functions are used to perform the scaling filter in Eq. (2). The rationale for using several wavelets is to evaluate the most suitable wavelet for our applications.

The coefficients are collected from the level six for each mother wavelet, then 2^6 coefficients are obtained. The energy operator is applied to each coefficient [2], and 320 energy coefficients are collected as it is illustrated in Fig. 10. In this figure, only the right side of the binary tree is presented, the details of this decomposition are shown in Fig. 2, Section 2.1.

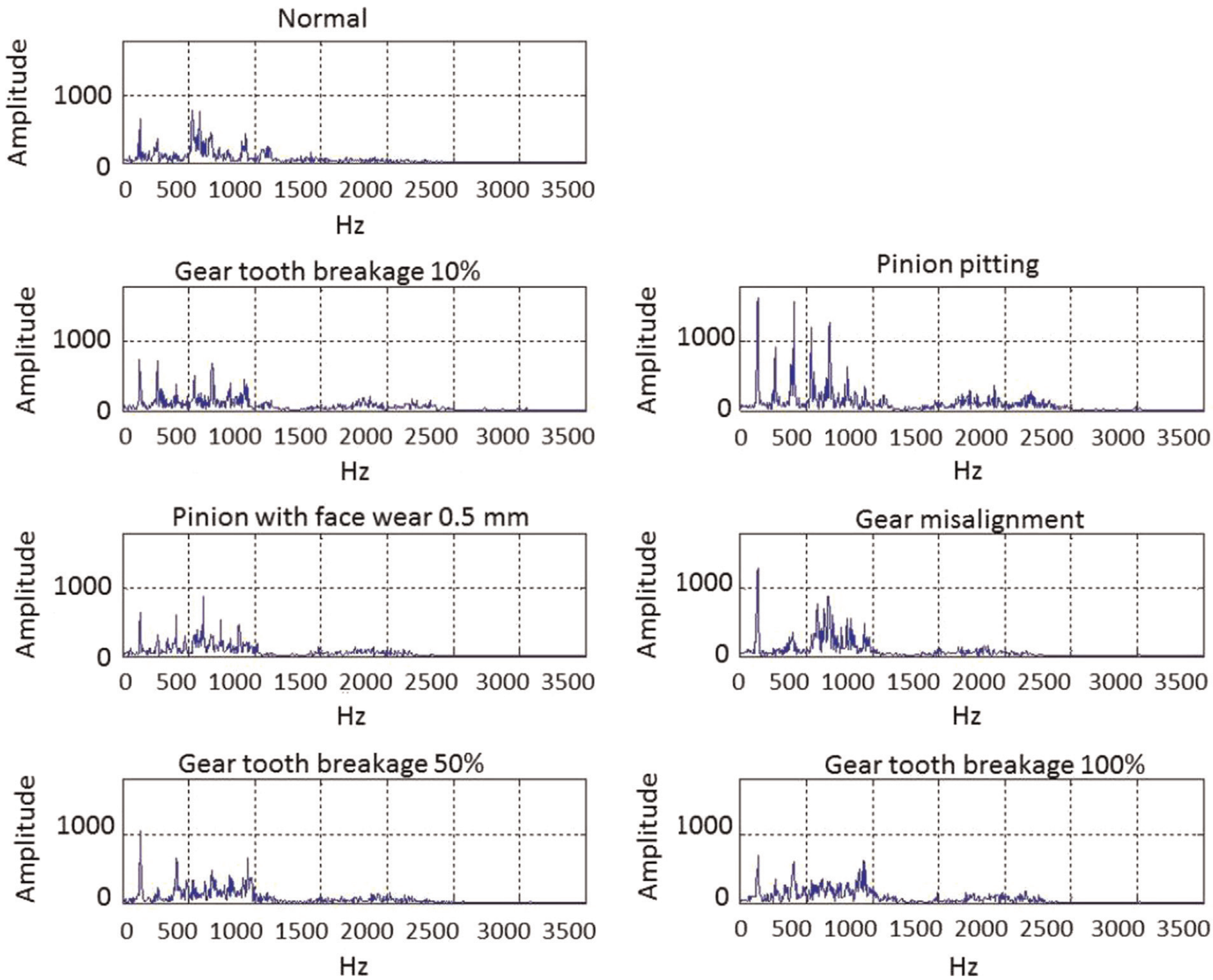


Fig. 9. A sample of frequency domain signals for the simulated faults in the gearbox.

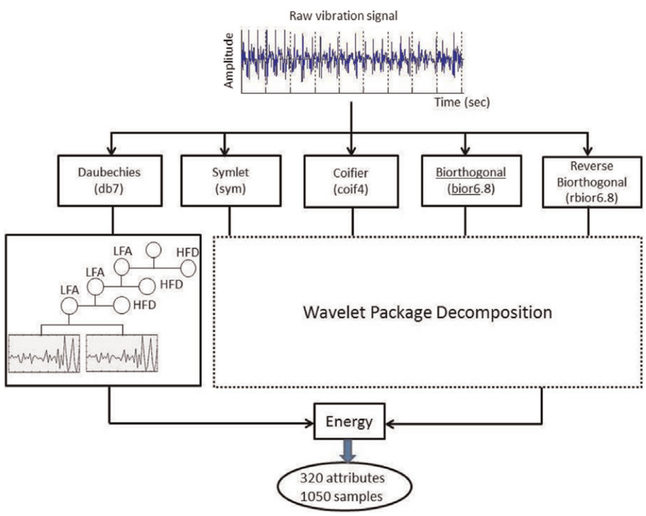


Fig. 10. The wavelets packet family analysis.

Please cite this article as: M. Cerrada, et al., Fault diagnosis in spur gears based on genetic algorithm and random forest, Mech. Syst. Signal Process. (2015), http://dx.doi.org/10.1016/j.ymssp.2015.08.030

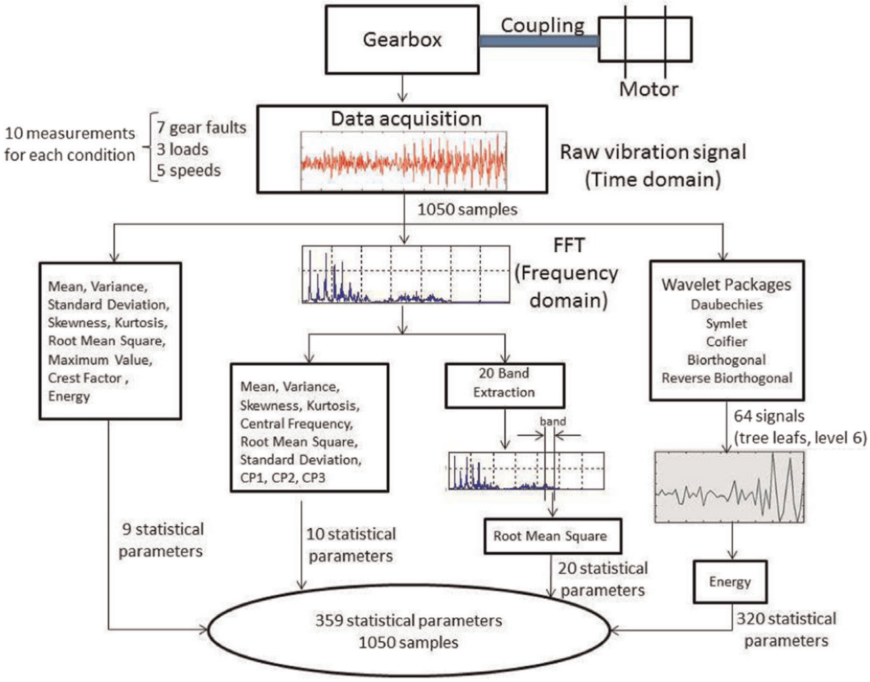


Fig. 11. Feature extraction procedure.

Wavelet Toolbox that is provided by MatLabR2013a® has been used to perform the wavelet packet analysis. WT is into the signal processing tools called time–frequency representation, then, the calculated energies in this analysis are condition parameters on time–frequency domain.

Finally, 359 attributes are the total number of condition parameters on time, frequency and time–frequency domain that are used in our approach. This data set is stored into a matrix with 1050 rows (sampled signals) and 359 columns (attributes), and only 2.72 MB of memory is needed. This data matrix will be used in the attribute selection process as described in Section 4. Fig. 11 illustrates, in detail, the feature extraction procedure that is explained above, from the raw signals to the data set that will be used in our approach according to Fig. 1.

4. GA for the attribute selection in classifiers based on RF

In this section, the development of our approach for attribute selection, also called feature selection, is presented. Once the data set is obtained in the previous feature extraction process, the training phase is performed by using the libraries for GA and RF that are available in Matlab®. These libraries have options for customizing our own applications, according to our design.

Let A and X be the set of attributes and the data set, respectively. The set A is $A=\{a_1, a_2, \dots, a_n\}$ where a_j is an attribute and the vector x_i in the data set is $x_i=\{x_{i1}, x_{i2}, \dots, x_{in}\}$. If the attribute a_j is in the data set, then the input x_{ij} is the value of this attribute, otherwise this input is not defined for the data set. Let $A_s \subset A$ and $f(X_{A_s})$, respectively, be a subset of attributes and the function that is associated to the classification performance by using a classification model, for the data set X with all attributes in A_s. The attribute selection is stated as the optimization problem in the following equation:

$$\min_{X_{A_s}} f(X_{A_s}) \tag{15}$$

In this work, the GA was designed for selecting the best subset of attributes A_s that points to the best RF classifier, by minimizing the classification score that is calculated by $f(X_{A_s})$. In this case, $f(X_{A_s})$ is the objective function for the GA and it is defined on performance metrics for classification, as it will be presented later in this section.

The process of attribute selection was carried out as shown in Fig. 1. Firstly, the features extraction process is performed according to the procedure in Section 3. The result of this step is a data matrix with M samples and n available attributes that will be used in the attribute selection process. Next, the attribute selection process runs in an iterative way by executing one cycle of the GA and one training phase of RF. One cycle of the GA corresponds with executing the processes in Fig. 5, the output of this cycle is an attribute matrix (N best individuals) as indicated in Section 2.3. For each individual, the data matrix is built up, and the output here is a set of N data matrix each one with 1050 samples, and the corresponding attribute candidates. In training phase, one RF-based classifier is adjusted for each data matrix and each performance is sent to the GA. Once the best subset of attributes is selected by accomplishing the stop criteria of the GA, a retraining phase for the classifier is performed with the best individual, and the diagnoser is obtained. In this work, the Gini index is used as impurity function in training and retraining phases for the RF. Finally, the RF-based classifier is tested with a new data matrix for the selected attributes.

The GA was designed with the following features:

1. *Individual encoding*: Let A be the set of n attributes $a_i, i = 1, \ldots, n$. Every possible solution S^j is encoded as a binary vector of size n, where the entry $b_i{=}1$ states for the attribute a_i that is defined for this individual, $b_i{=}0$ if the attribute a_i is not defined, according to Fig. 6.
2. *Initial population*: Given the binary representation of the individual S^j, the population is encoded as a binary matrix M where the rows $j = 1, .., N$ are the randomly selected individuals, and the columns $i{=}1,\ldots,n$ are the available attributes. Then, the entry $m_{ji} = 1$ in M states for the selected attributes a_i in the individual S^j. A population P with 100 individuals was generated with random selection of 0 or 1 for each entry m_{ji}.
3. *Fitness function*: The performance of a classifier is based on specific metrics such as the F_β-score, which is a weighted average of the precision and recall [33]. These parameters are obtained from the confusion matrix, which reports the results according to the true positives, true negatives, false positives and false negatives, in classification results. The value of $F_\beta \rightarrow 1$ is a good measure of the classifier performance. On the other hand, the OOB-error that is mentioned in Section 2.2 is usually used as a metric of the RF performance. In this work, the fitness function takes into account the F_1-score, ($\beta = 1$), and the OOB-error as defined in Eq. (16). The objective of the GA is to minimize this function

$$f = OOB_{error} + (1 - F_1) \tag{16}$$

In order to determine the value of F_1 in Eq. (16), the confusion matrix is built up for each tree of the RF. The following procedure is proposed:

(a) Define the test set TS as proposed in Eq. (17), where OOB_b is the set of out of bag samples for the tree T_b, in one cycle of the GA, and ts_k is one test sample:

$$TS = \bigcup_b OOB_b = \{ts_1, \ldots, ts_n\} \tag{17}$$

(b) Define the tree subset $T_{ts_k} \subset \{T_b\}_1^B$, for each sample $ts_k \in TS$, as described in the following equation:

$$T_{ts_k} = \{T_b | ts_k \notin OOB_b\} \tag{18}$$

(c) Select the class $\hat{C}_{rf}^B(ts_k)$ according to Eq. (9) by using the set T_{ts_k}, for each sample $ts_k \in TS$.
(d) Built up the confusion matrix once whole the set TS has been processed.
(e) Compute F_1 for the current GA cycle.

Eq. (18) states that the sample ts_k is not in the OOB samples of the trees that will be used for estimating its class, and the confusion matrix is accepted as a reliable result for calculating the F_1-score during the selection procedure.

4. *Parent selection*: A fitness scaling based on the rank was used, in order to have a good representation of the fitness value over a suitable range [26]. For minimization problems, an individual with lower value of raw fitness is assigned to a higher value of scaling fitness, consequently, it has a higher probability to be chosen as a parent by the selection function. The *rank* is a function *rank*: $f \rightarrow N$, which sorts individuals from 1 to N, based on the best fitness value. Once the individuals are ranked, their new fitness f_{ranked} are recalculated, according to the following equation:

$$f_{ranked} = p - 2\frac{(r-1)(p-1)}{N-1} \tag{19}$$

where r is the rank of the individual, p is the desired selective pressure and N is the size of the population. The selective pressure is the probability to selecting the best individual regarding the average selection probability of the remaining individuals. The rank-based fitness scaling allows a selective pressure between [1.0–2.0].

In this work, the Uniform Stochastic Selection method was applied. It operates in the following way:

(a) Determine the cumulative probabilities q_i based on the value f_{ranked} of each individual: as follows:

$$q_0 \leftarrow 0$$

for $i = 1, \ldots, N$,

$$q_i \leftarrow q_{i-1} + f_{ranked}^{S^i} / \sum_{j=1}^{N} f_{ranked}^{S^j}$$

end

Table 3
Sets of initial population.

Number of attributes	Condition parameters
256	From wavelet coefficients (db7+sym3+coif4+ bior6.8)
295	9 from time domain + 10 from frequency domain + 20 from frequency bands + 256 from wavelet coefficients (db7+sym3+coif4+ bior6.8)
320	from wavelet coefficients (db7+sym3+coif4+ bior6.8+rbior6.8)
359	9 from time domain + 10 from frequency domain + 20 from frequency bands + 320 from wavelet coefficients (db7+sym3+coif4+ bior6.8+rbior6.8)

(b) Select K parents, in the following manner:

For $i=1,\ldots,K$,

$r \leftarrow random(0, q_N)$

$parent^i \leftarrow S^i \quad \text{if } q_{i-1} < r < q_i$

end

5. *Genetics operators*: Crossover and mutation operators were applied. The crossover fraction was set on 80%, that is, the number of children that will be obtained from crossover. The crossover-point selection was defined by applying a random scattered selection, as follows:

(a) Select the father 1.
(b) Select the father 2.
(c) Generate a random binary vector v of N bits, this is $v_i=0$ or $v_i=1$.
(d) If $v_i=1$,

(i) Gene i of the parent 1 is preserved, otherwise.
(ii) Replace the gen i of the parent 1 with the gene i of the parent 2.

The 20% of the remaining population will be obtained by mutation, with a mutation rate of 0.05 for each gene of the selected father.

6. *Parents replacement*: In this case, a direct replacement mechanism was used. The 10% of the current population was selected by elitism and it will be part of the next generation, the remaining individuals are replaced according to the fraction of children obtained by crossover and mutation.
7. *Stop criteria*: Two stop criteria were defined, the first one by reaching the maximum number of generations (set at 200 generations), and the second one by obtaining a minimum of change in the fitness value between two successive generations (set on 10^{-6}). The GA was stopped by maximum number of generations in all the experiments.

5. Results and analysis

Four sets of initial population have been analysed, as shown in Table 3, in order to evaluate the performance of the attribute selection process. Each initial set is composed of several condition parameters that have been extracted from the vibration signal as described in Section 3. This selection aims to analyse the effect of extending the best empirical selection of the first initial set, by incorporating time and frequency parameters. This first initial set is composed of 256 energy parameters from four wavelet families, and it has been obtained in an incremental, exhaustive search, on the wavelets families, by using RF classifier without GA-based feature selection.

Each RF in the attribute selection process has been trained with 800 trees at maximum depth and a number of random variables $\rho=\sqrt{Number\ of\ attributes}$, that are selected from the attributes in the corresponding data matrix. In the retraining phase, the RF uses 1000 trees in the same conditions and the obtained results are described in Table 4. This table indicates the set size before and after the selection process, and the performance metrics that have been calculated from a micro-average approach as presented in [33]. By using this micro-average approach, it is possible to obtain the same values for all three metrics, in case of good classification. The best result is with 122 attributes from the initial set with 295

Table 4
Attribute selection by using GA.

Initial number	Final number	Precision	Sensibility	F_1−score
256	115	0.9762	0.9762	0.9762
295	122	0.9781	0.9781	0.9781
320	141	0.9733	0.9733	0.9733
359	175	0.9705	0.9705	0.9705

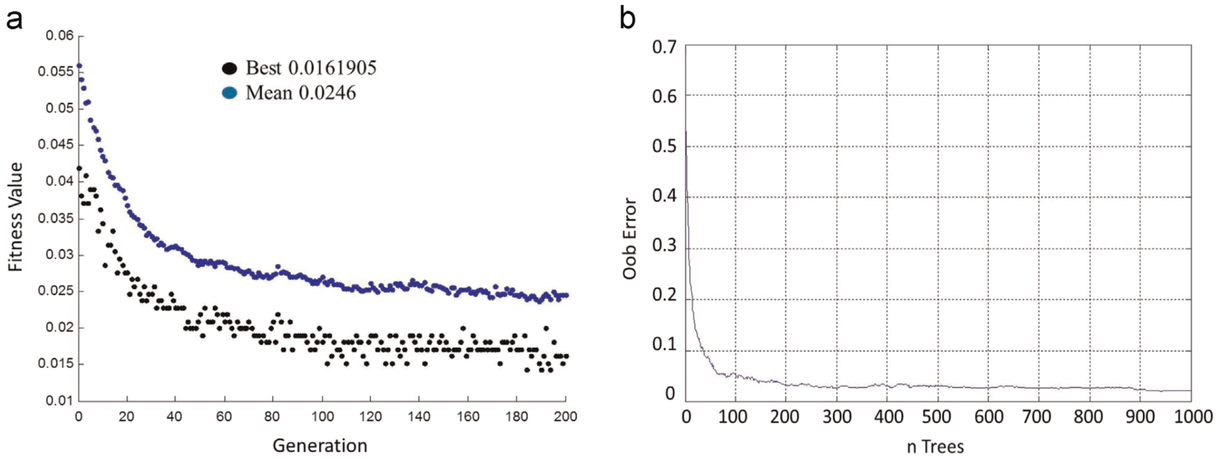


Fig. 12. Optimization process for attribute selection by using GA

attributes from the wavelets coefficients, time and frequency statistical parameters. Moreover, higher quality results are not reached by extending the attributes set, even with the inclusion of the Reverse Biorthogonal wavelet family.

The optimization process in the best case with 122 attributes is illustrated in Fig. 12, where part (a) is the fitness value in each generation during the selection process and part (b) is the OOB-error in the training process by using different number of trees (n Trees). In part (a), the fitness value for the best individual is around 0.0161905 and the mean value in the last generation is around 0.0246. Table 5 shows the best individual that is obtained in this experiment and Table 6 shows the corresponding confusion matrix. The worst case in the classification is the diagnosis of the fault 3 (pinion pitting) and the incipient fault 4 (pinion with face wear). This case was expected because both of associated spectra have similar characteristics as mentioned in Section 3.

In order to compare the diagnosis results with and without the selection process by using GA, Table 7 presents the precision, sensibility and F_1-score that have been obtained with the following considerations: (i) 39 attributes from time and frequency domain, without GA; (ii) 256 attributes from wavelet coefficients, without GA; (iii) 115 selected attributes from 256 wavelet coefficients and (iv) 122 selected attributes from time, frequency and wavelet coefficients, by using GA (the best result). The performance of the RF-based classifier in case (ii) is better than the performance in case (i). Moreover, the classification result by using 256 attributes from wavelet coefficients in case (ii) is improved by the selection process with only 115 attributes. Finally, this last result can be improved again by the selection process from the initial set with 295 attributes including time and frequency parameters. This previous analysis justify the inclusion of wavelets for diagnostic information, the F_1-score was improved from 0.9466 to 0.9781 by using the GA-based attributes selection process.

Additionally, other alternatives of attribute selection have been evaluated for comparing with the best result in the proposed approach. The rationale of this experiment is to explore the capability of the GA in selecting a similar set of attributes by using statistical techniques that exploit the meaningful data.

Two approaches are studied: (i) hierarchical attribute clustering [22], based on the p-values matrix for the hypothesis testing of non-correlation, and (ii) attribute selection based on the importance variables by using the RF algorithm [13]. The importance variables estimation is a very useful tool in RF, that leads to rank each attribute according to its good performance in classification, by following a random search process. Particularly, this comparison is regarding the best result, that is, the best 122 attributes. Then, for both approaches, the data matrix with 1050 samples (rows) and 295 attributes (columns) has been used (see the set of experiments in Section 4).

For the hierarchical attribute clustering, the p-values matrix was calculated from the correlation coefficients matrix of the data matrix. As a result, a new square symmetric matrix P is obtained where each entry is a value that is related with a non-correlation measure between two attributes. This P matrix was used as a measure of similarity for building the hierarchical groups of attributes, the threshold value $p < 0.05$ indicates a significant correlation between the attributes. Regarding this threshold, the hierarchical decomposition was developed until reaching 122 clusters from the initial group of 295 attributes.

Please cite this article as: M. Cerrada, et al., Fault diagnosis in spur gears based on genetic algorithm and random forest, Mech. Syst. Signal Process. (2015), http://dx.doi.org/10.1016/j.ymssp.2015.08.030

Table 5
Best individual by using GA.

Parameter domain	Total number	Description
Frequency bands	7	–183 Hz (1st band) 184–367 Hz (2nd band) 733–915 Hz (5th band) 1099–1281 Hz (7th band) 1465–1647 Hz (9th band) 2380–2562 Hz (14th band) 3478–3660 Hz (20th band)
Frequency	5	Kurtosis Central frequency RMS CP1 CP3
Time	7	Mean Variance Kurtosis RMS Maximum value Crest Factor Energy
Wavelets	103	30 energy parameters from db7 25 energy parameters from sym3 22 energy parameters from coif4 26 energy parameters from bior6.8

Table 6
Confusion matrix for the best set of attributes.

Fault label	1	2	3	4	5	6	7
1	149	1	0	0	0	0	0
2	1	144	3	0	0	1	1
3	0	2	145	3	0	0	0
4	0	0	5	145	0	0	0
5	0	0	1	0	149	0	0
6	0	0	1	1	0	146	2
7	0	1	0	0	2	1	146

Table 7
Comparative values on different approaches by using RF and GA.

Number of attributes	Diagnosis algorithm	Precision	Sensibility	F_1−score
(i) 39	RF	0.9466	0.9466	0.9466
(ii) 256	RF	0.9695	0.9700	0.9699
(iii) 115	RF-GA	0.9762	0.9762	0.9762
(iv) 122	RF-GA	0.9781	0.9781	0.9781

In case of the variable importance estimation by using RF, the same data matrix was used to train the RF, and the best 122 ranked variables were selected. This procedure is illustrated in Fig. 13.

The comparative analysis was performed based on the following premise: each group of attributes from the hierarchical clustering contains at least one attribute that was found by one selection technique (GA or RF). This premise suggests that if an attribute's cluster has more than one selected attribute, the selection technique is not able to discriminate the attributes that are correlated according to the selected threshold value ($p < 0.05$). However, the attribute has an importance level according to the selection criteria in GA or RF. Otherwise, if no attribute was selected from an attribute cluster by the selection technique, such a technique did not assign any importance to the attributes that are not correlated with other clusters.

By construction, the hierarchical clustering proposes clusters of attributes with some correlation level. After analysing all the 122 clusters, the GA has selected 56 non-correlated attributes and RF has selected 46 non-correlated attributes. Moreover, GA has not selected 56 non-correlated attributes and RF has not selected 66 non-correlated attributes. Table 8 illustrates a part of the previous analysis, where the numbers between brackets denote the attribute identifier. The first row shows that GA has selected one attribute against the three selected attributes by RF, the second row shows that no attribute has been selected by any method. In the third row, RF has selected one attribute against the three attributes by using GA, the

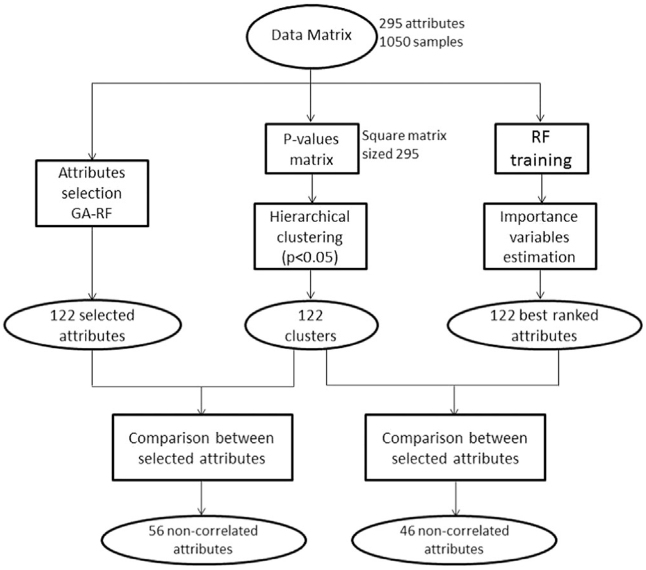


Fig. 13. Comparison scheme in selecting non-correlated attributes.

Table 8
Comparative analysis in feature selection by using HC, GA and RF.

Cluster label	Cluster by HC	Selected attribute by GA	Selected attribute by RF
117	[27, 91, 155, 219, 283]	155	[27, 91, 155]
113	[201, 202, 265, 266]	[]	[]
110	[32, 160, 224, 288]	[32, 160, 224]	160
107	[192, 256, 320]	256	[]
89	8	[]	8

last two rows show the case where one technique has selected one attribute but the other one do not. Then, the selection process by using GA is adequate for discriminating the significant non-correlated attributes with regard to the selection by RF, taking into account the *p*-value for non-correlation test.

6. Conclusions

In this work a two step approach for designing a fault diagnosis system for spur gears is proposed. In the first step, genetic algorithm for attribute selection is used, in a supervised environment with Random Forest based diagnoser, in the second step, the fault diagnoser is trained again by using the selected attributes. This approach aims to discover which is the adequate subset of features regarding the classical time and frequency parameters, and the energy from wavelet coefficients that leads to a better performance for fault diagnosis when a RF classifier is used.

The experimental results show interesting elements for fault diagnosis in spur gears with the proposed approach:

(i) The use of only time and frequency condition parameters, or only energy parameters from wavelets are not enough for obtaining a good precision in classification. This can be seen in Table 7 where the use of 39 attributes from time and frequency domain reaches the precision value around 0.9466.
(ii) Energy from the coefficients of wavelet packages, as condition parameter, improves the classification performance regarding the use of time and frequency parameters. This can be seen in Table 7 where the use of 256 attributes from wavelets reaches a precision value in 0.9695. In this sense, the precision has been improved around 2.36%, regarding the value in (i).
(iii) A suitable feature selection from time, frequency and wavelet parameters can lead to an adequate decision model for fault diagnosis based on the RF technique. This can be seen in Table 4 where the use of the initial set with 295 attributes reaches the best precision value in 0.9781. This result improves around 3.22% the result in (i) and 0.87% the result in (ii). This last case uses around 33% of the original set size, that is 122 attributes regarding 359 available attributes, and still get an increase precision in the diagnosis. With a reduced set of attributes, the computational requirements for running the classification model can be improved. This is an important characteristic for implementing fault diagnosers in real industrial environments.

Please cite this article as: M. Cerrada, et al., Fault diagnosis in spur gears based on genetic algorithm and random forest, Mech. Syst. Signal Process. (2015), http://dx.doi.org/10.1016/j.ymssp.2015.08.030

Future works aim to propose a multi-stage procedure by using the same approach, as a greedy algorithm. This leads to exploit the searching space from the current selected attributes, in order to discover new best solutions. Furthermore, a study of physical meaning of relevant condition parameters for characterizing failure modes in spur gears will be addressed.

Acknowledgements

The authors want to express a deep gratitude to The Ministry of Higher Education, Science, Technology and Innovation (SENESCYT) of the Republic of Ecuador and the Prometeo Project, the GIDTEC research group of the Universidad Politécnica Salesiana in Cuenca-Ecuador, and the Vibrational research group of the Universidad Nacional de Educación a Distancia in Madrid-Spain, for supporting the accomplishment of this research. The authors also want to express their thanks to the reviewers, the valuable comments and suggestions are very much appreciated.

References

[1] R.B. Randall, Vibration-Based Condition Monitoring, John Wiley & Sons, Ltd., West Sussex, United Kingdom, 2011.
[2] R. Yan, R.X. Gao, X. Chen, Wavelets for fault diagnosis of rotary machines: a review with applications, J. Signal Process. 96 (2014) 1–15.
[3] D. Kanneg, A wavelet spectrum technique for machinery fault diagnosis, J. Signal Inf. Process. 2 (2011) 322–329.
[4] S. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way, Elsevier Academic Press, Amsterdam, 2009.
[5] G. Marichal, M. Artés, J.G. Prada, O. Casanova, Extraction of rules for faulty bearing classification by a neuro-fuzzy approach, J. Mech. Syst. Signal Process. 25 (2011) 2073–2082.
[6] H. Li, Y. Zhang, H. Zheng, Gear fault detection and diagnosis under speed-up condition based on order cepstrum and radial basis function neural network, J. Mech. Sci. Technol. 23 (2009) 2780–2789.
[7] K. Gryllias, I. Antoniadis, A support vector machine approach based on physical model training for rolling element bearing fault detection in industrial environments, J. Eng. Appl. Artif. Intell. 25 (2012) 326–344.
[8] M. Artés, L. Del Castillo, J. Pérez, Failure prevention and diagnosis in machine elements using cluster, in: Proceedings of the 10th International Congress on Sound and Vibration, Stockholm, Sweden, pp. 1197–1203.
[9] S. Hussain, H.A. Gabbar, Gearbox fault detection using real coded genetic algorithm and novel shock response spectrum features extraction, J. Nondestruct. Eval. 33 (2014) 111–123.
[10] B. Samanta, Gear fault detection using artificial neural networks and support vector machines with genetic algorithms, Mech. Syst. Signal Process. 18 (2004) 625–644.
[11] A.K. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mech. Syst. Signal Process. 20 (2006) 1483–1510.
[12] R. Timofeev, Classification and regression trees (CART). Theory and applications (Master's thesis), Center of Applied Statistics and Economics Humboldt University, Berlin, 2004.
[13] A. Verikas, A. Gelzinis, M. Bacauskiene, Mining data with random forests: a survey and results of new tests, Pattern Recognit. 44 (2011) 330–349.
[14] B.S. Yang, X. Di, T. Han, Random forests classifier for machine fault diagnosis, J. Mech. Sci. Technol. 22 (2008) 1716–1725.
[15] N. Karabadji, I. Khelf, H. Seridi, L. Laouar, Genetic optimization of decision tree choice for fault diagnosis in an industrial ventilator, in: T. Fakhfakh, W. Bartelmus, F. Chaari, R. Zimroz, M. Haddar (Eds.), Condition Monitoring of Machinery in Non-Stationary Operations, 2012, pp. 277–283.
[16] G. Zurita-Villarroel, A. Agren, Wavelet transform analysis of measurements of engine combustion noise, in: B. Challen (Ed.), Proceedings of the 1997 Noise and Vibration Conference, SAE Conference Proceedings, 1997.
[17] C. Gargour, M. Gabrea, V. Ramachandran, J. Lina, A short introduction to wavelets and their applications, IEEE Circuits Syst. Mag. 9 (2009) 57–68.
[18] I. Moumene, N. Ouelaa, Gears and bearings combined faults detection using Hilbert transform and wavelet multiresolution analysis, in: T. Fakhfakh, W. Bartelmus, F. Chaari, R. Zimroz, M. Haddar (Eds.), Condition Monitoring of Machinery in Non-Stationary Operations, 2012, pp. 319–328.
[19] S. Mitra, Digital Signal Processing: A Computer-Based Approach, McGraw-Hill, New York, 2011.
[20] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, New York, 2009.
[21] T. Mitchell, Machine Learning, McGraw-Hill, New York, 1997.
[22] I.H. Witten, E. Frank, Data Mining: Practical Machine Learning, Tools and Techniques, Morgan Kaufman, Boston, 2005.
[23] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[24] A. Ziegler, I.R. Knig, Mining data with random forests: current options for real-world applications, in: Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 4, 2013, pp. 55–63.
[25] M. Srinivas, L. Patnaik, Genetic algorithms: a survey, Computer 27 (1994) 17–26.
[26] F. Sadjadi, Comparison of fitness scaling functions in genetic algorithms with applications to optical processing, in: Proceedings of SPIE, Optical Information Systems II, vol. 5557, 2004, pp. 356–364.
[27] T. Blickle, L. Thiele, A Comparison of Selection Schemes used in Genetic Algorithms, Technical Report, Swiss Federal Institute of Technology (ETH) Zurich, Computer Engineering and Communications Networks Lab (TIK), 1995.
[28] K. Jebari, M. Madiafi, Selection methods for genetic algorithms, Int. J.Emerg. Sci. 3 (2013) 333–344.
[29] M. Srinivas, L. Patnaik, Adaptive probabilities of crossover and mutation in genetic algorithms, IEEE Trans. Syst. Man Cybern. 24 (1994) 656–667.
[30] F. Hilding, K. Ward, Automated operator selection on genetic algorithms, in: R. Khosla, R. Howlett, L. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Lecture Notes in Computer Science, vol. 3684, 2005, pp. 903–909.
[31] M. Lozano, F. Herrera, J. Cano, Replacement strategies to preserve useful diversity in steady-state genetic algorithms, Inf. Sci. 178 (2008) 4421–4433.
[32] Y. Lei, Z. He, Y. Zi, A new approach to intelligent fault diagnosis of rotating machinery, Expert Syst. Appl. 35 (2008) 1593–1600.
[33] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, Inf. Process. Manag. 45 (2009) 427–437.