Fault diagnosis of spur gearbox based on random forest and wavelet packet decomposition

Article *in* Frontiers of Mechanical Engineering · September 2015

DOI: 10.1007/s11465-015-0348-8

7 authors, including:

Diego Román Cabrera
Universidad Politécnica Salesiana (UPS)
72 PUBLICATIONS **1,395** CITATIONS
SEE PROFILE

René Vinicio Sánchez
Universidad Politécnica Salesiana (UPS)
66 PUBLICATIONS **1,497** CITATIONS
SEE PROFILE

Grover V. Zurita
Universidad Privada Boliviana
44 PUBLICATIONS **1,009** CITATIONS
SEE PROFILE

Mariela Cerrada
Universidad Politécnica Salesiana (UPS)
141 PUBLICATIONS **1,857** CITATIONS
SEE PROFILE

Diego CABRERA, Fernando SANCHO, René-Vinicio SÁNCHEZ, Grover ZURITA, Mariela CERRADA, Chuan LI, Rafael E. VÁSQUEZ

Fault diagnosis of spur gearbox based on random forest and wavelet packet decomposition

Abstract This paper addresses the development of a random forest classifier for the multi-class fault diagnosis in spur gearboxes. The vibration signal's condition parameters are first extracted by applying the wavelet packet decomposition with multiple mother wavelets, and the coefficients' energy content for terminal nodes is used as the input feature for the classification problem. Then, a study through the parameters' space to find the best values for the number of trees and the number of random features is performed. In this way, the best set of mother wavelets for the application is identified and the best features are selected through the internal ranking of the random forest classifier. The results show that the proposed method reached 98.68% in classification accuracy, and high efficiency and robustness in the models.

Diego CABRERA (✉), René-Vinicio SÁNCHEZ, Grover ZURITA, Mariela CERRADA, Chuan LI
Departamento de Ingeniería Mecánica, Universidad Politécnica Salesiana, Cuenca, Ecuador
E-mail: dcabrera@ups.edu.ec

Diego CABRERA, Fernando SANCHO
Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Sevilla, España

Grover ZURITA
Departamento de Ingeniería Electro-Mecánica, Universidad Privada Boliviana, Cochabamba, Bolivia

Mariela CERRADA
Departamento de Sistemas de Control, Universidad de Los Andes, Mérida, Venezuela

Chuan LI
Research Center of System Health Maintenance, Chongqing Technology and Business University, Chongqing 400067, China

Rafael E. VÁSQUEZ
Facultad de Ingeniería Mecánica, Universidad Pontificia Bolivariana, Medellín, Colombia

Keywords fault diagnosis, spur gearbox, wavelet packet decomposition, random forest

1 Introduction

Power transmission mechanisms are fundamental components used in several types of machines, and spur gearboxes are one of the most used types in low and medium speed applications. As a fundamental tool in condition-based maintenance, fault diagnosis of power transmission devices has become an interesting tool in order to avoid unwanted downtime in industrial processes. Specialized literature shows that several works have been developed in dynamic analysis and fault diagnosis for gearboxes. Walha et al. [1] examined the effect of backlash in a two-stage gear system using standard methods for nonlinear systems. Abbes et al. [2] performed the dynamic analysis of gearboxes by using substructures, considering the static error as a vibratory excitation. Tian et al. [3] identified effective and sensitive health indicators in order to monitor the crack propagation in a one-stage gearbox using the dynamic model and the discrete wavelet transform. In order to avoid devastating consequences following to unexpected failures, several works addressing detection and diagnosis methods have been developed. Within such methods, one can mention: Lubricant monitoring using chemical analysis [4], analysis of motor current signals using spectral analysis based on the fast Fourier transform [5], and analysis of vibration signals [6], among others.

One of the major drawbacks of all the mentioned methods is that they require the presence of a human expert to obtain results, which can produce biased conclusions that rely on the expert's experience. Therefore, there is a need to automate the fault detection and diagnosis processes using more robust methods such as those coming from data mining and machine learning. Rafiee et al. [7] and Sanchez et al. [8] showed the use of multilayer perceptron neural networks to recognize fault conditions in gearboxes. Other methods, like adaptive neural networks, able to change its topology with the inclusion of new samples were introduced by Barakat et al. [9], while Yang et al. [10] introduced the use of clustering techniques based on self-organized Kohonen neural networks.

The problem of techniques based on neural networks (NN) is that they are highly dependent on the number of features; therefore, if the problem is highly dimensional, it is likely that the NN will not converge. Additionally, falling into local minima is possible due to an iterative-deterministic search carried out by the training algorithm, preventing the convergence to the global optimum. New techniques have been developed in order to overcome such problems: For instance, support vector machines (SVMs) to identify the presence or absence of faults [11]; SVMs sets with multi-classification least squares support vector machines in fault diagnosis [12]; or Bayesian network [13], which in comparison with back-propagation neural network and probabilistic neural network have shown better results both in the error rate and computation time.

Breiman et al. [14] introduced the random forest (RF) model, based on the grouping of trees for classification and regression (CART) [15]. The method is based on the use of a large number of CARTs, called weak learners, which are trained using the technique of bagging (random sampling with replacement). Subsequently, the results of classification of these trees are weighted to provide a single response. The split feature in each node of the weak learner is selected from a random population constructed from the initial population of features, following the Gini criterion as a measure of information gain. Both, measurement of random population of features and size of the forest (number of weak learners trees), are parameters selected by the user.

These two parameters change the correlation among the trees, keeping enough diversity in the forest while avoiding over-training [16]. The resulting RF model is inherently multi-class and the decision-making process is done by counting each tree in the forest. RF offers a performance metric called out of bag error (oob-error) calculated as the average of the rate of error in each weak learner, validating with untrained samples remaining from the bagging process. From the previous properties, the characteristics of RF over other methods are summarized into an ability to perform classification with few training samples and lots of attributes, a reduction in bias for better generalization and a reduction in variance, which provides a robust RF classification model.

This model has already been considered for fault diagnosis earlier. Han et al. [17] proposed to use it in fault diagnosis for bearings in electric motors with features obtained from the standard deviation of wavelet coefficients. Yang et al. [18] proposed features extraction from time and frequency domains, and parameters optimization through the use of genetic algorithms. Karabadji et al. [19] made a comparative study of different models based on trees using genetic search, showing that RF is among the classifiers that provide better performance.

This paper addresses the development of a RF classifier for the multi-class fault diagnosis in spur gearboxes. The rest of the paper is structured as follows: Section 2 presents the background of decision trees and RF model; Section 3 shows the methodology used with RF for faults diagnosis and the configuration of the experimental system used to obtain signals of vibration with the extraction and selection of features for the classification; in Section 4, the results obtained with the methodology proposed by performing an analysis of the features selected by the model and the best parameters are shown; finally, some conclusions and future work are presented.

2 Decision trees and random forest model

An instance of the problem can be defined as a vector of features:

$$\boldsymbol{v} = (x_1,\ x_2,\dots,\ x_d)\ \in\ \boldsymbol{R}^d,$$

where each element x_i represents a feature of the instance, and d is the problem's dimension. Hence, the feature space has d dimensions. Since d could be very large (even infinite), which would make the problem unsolvable, it is necessary to use a function to select a reduced set of "best" features, $\boldsymbol{\phi}(\boldsymbol{v}) = (x_{\phi_1},\ x_{\phi_2},\dots,\ x_{\phi_d}) \in \boldsymbol{R}^d$, where each element ϕ_i belongs to $[1,d]$.

In a binary tree-based model, each node j has a decision function, $h(\boldsymbol{v},\boldsymbol{\theta}_j) : \boldsymbol{R}^d \times P \longrightarrow \{0,1\}$, that receives the vector of features and splits parameters, $\boldsymbol{\theta}_j \in \boldsymbol{P}$, associated to the node j, and decides if the vector goes down right or left in the tree; $\boldsymbol{P}$ is the space of parameters.

2.1 Training the model

Training a tree-based classifier [15] is similar to building the tree in a recursive manner. For each node j, a set of split parameters is selected resolving the optimization problem:

$\boldsymbol{\theta}_j = \underset{\boldsymbol{\theta} \in \boldsymbol{P}}{\arg\max}\, I(\boldsymbol{S}_j,\boldsymbol{\theta})$, where I is a fitness function, some gain information criterion, and $\boldsymbol{S}_j$ is a subset of training set belonging to node j. The tree is built to achieve the maximum purity in set $\boldsymbol{S}_j$, and some stop criterion can be used to finish the construction, for example a defined depth.

An RF classifier [14] is formed from a group of decision trees trained with injected randomness following two processes (see Fig. 1):

1) Bagging. The training set $\boldsymbol{S}$ is sub-sampled with replacement to build T new subsets in order to train T trees.

2) Randomized node. In each node of each tree the

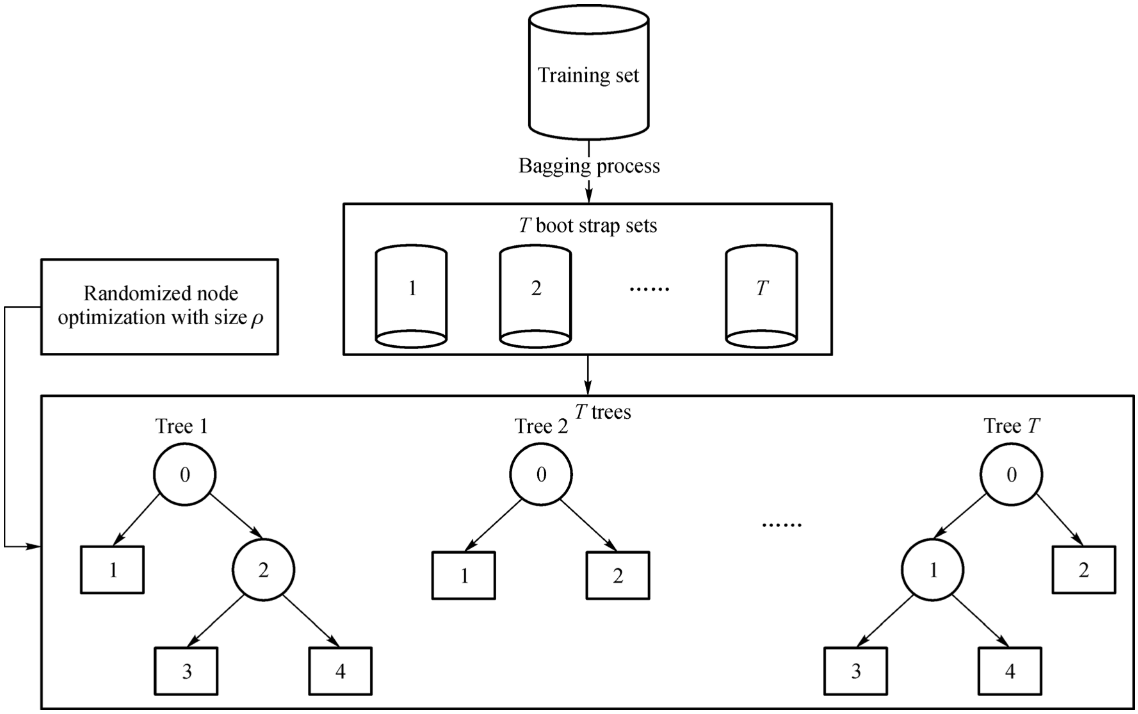


Fig. 1 Training process for the RF classifier

optimization problem is modified to: $\theta_j = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{P_j}} I(\boldsymbol{S}_j, \boldsymbol{\theta})$, where $\boldsymbol{P}_j$ is a randomly selected subset of $\boldsymbol{P}$, $|\boldsymbol{P}_j| \ll |\boldsymbol{P}|$. We will use $|\boldsymbol{P}_j|$, rho and ρ indistinctively.

2.2 Testing the model

After building the group of trees, new samples can be classified. For that, the sample is classified for every tree, and from the set of results we obtain a probability distribution by averaging different results (Fig. 2).

3 Application of random forest for spur gearbox fault diagnosis

From vibratory signals, a set of features based on wavelet packet decomposition (WPD) is obtained. RF is then used in feature ranker mode to select the most important features; the parameters of the model are optimized with a greedy algorithm.

3.1 Measurement set up

The measurements were carried out at the Mechanical Engineering Division (Vibration Laboratory) from the Salesian Polytechnic University of Cuenca (Ecuador). Figure 3 shows the measurement set up. The vibration signatures are obtained from a uni-axial accelerometer. Seven classes are used to evaluate the classifier: Normal operation (Class 1), and the following 6 types for failures:

- 10% broken tooth (Class 2);
- Pinion pitting (Class 3);
- Wear of 0.5 mm in the face of the pinion (Class 4);
- Misaligned gear (Class 5);
- 50% broken tooth (Class 6);
- 100% broken tooth (Class 7).

3.2 Extraction of features

Preprocessing raw data in order to obtain the features is done by using WPD: Firstly, the raw signal is decomposed using a mother wavelet in approximation signal and detail signal. After that, approximation and detail signals are decomposed again. This process continues until a level 6 of decomposes signals, as shown in Fig. 4(a).

Then, for each signal obtained in level 6, the energy is calculated building a vector of 64 real values; this process is repeated using 5 different mother wavelets: Daubechies 7 (db7), symlet 3 (sym3), coiflet 4 (coif4), biorthogonal 6.8 (bior6.8) and reverse biorthogonal 6.8 (rbior6.8). All these vectors of energy are put together resulting in a features vector with 320 real values, as shown in Fig. 4(b). The purpose of using these families of wavelets is to verify that the diagnostic information specifically supplied by these wavelets is sufficient to achieve an acceptable ranking in the classification task.

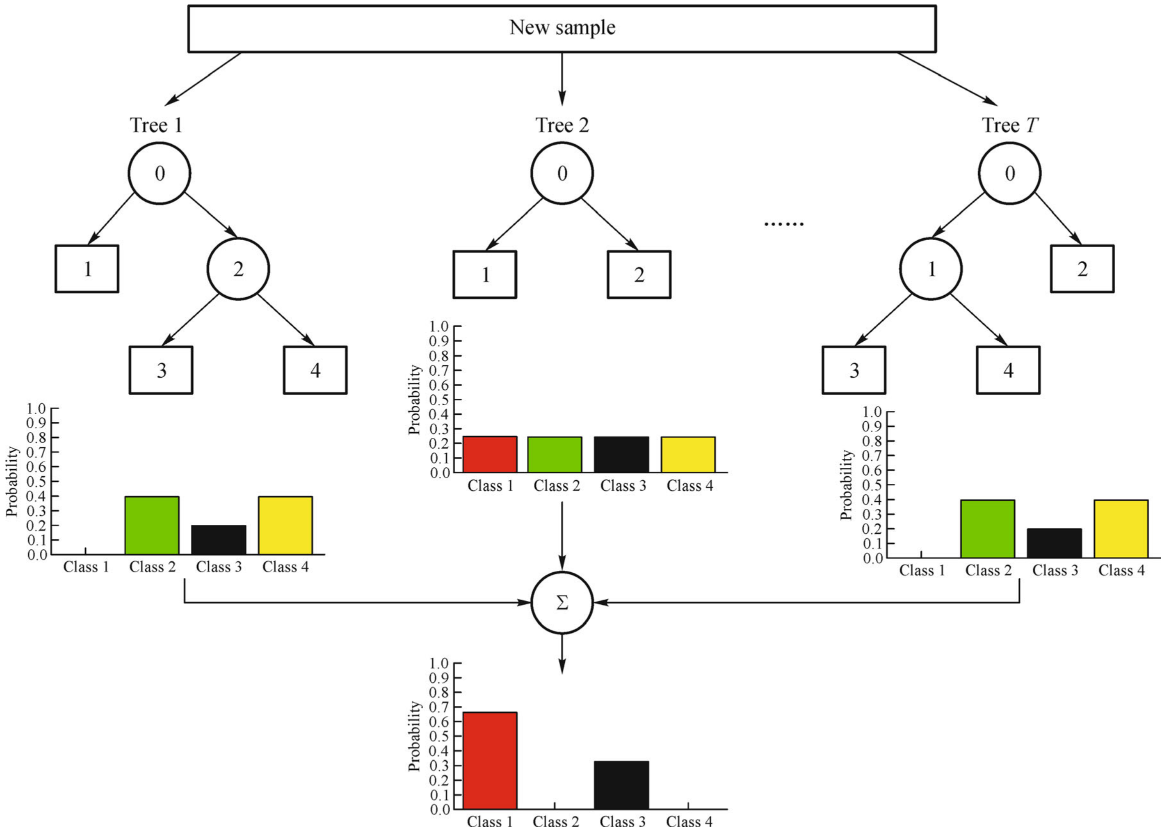


Fig. 2 Test process for the RF classifier

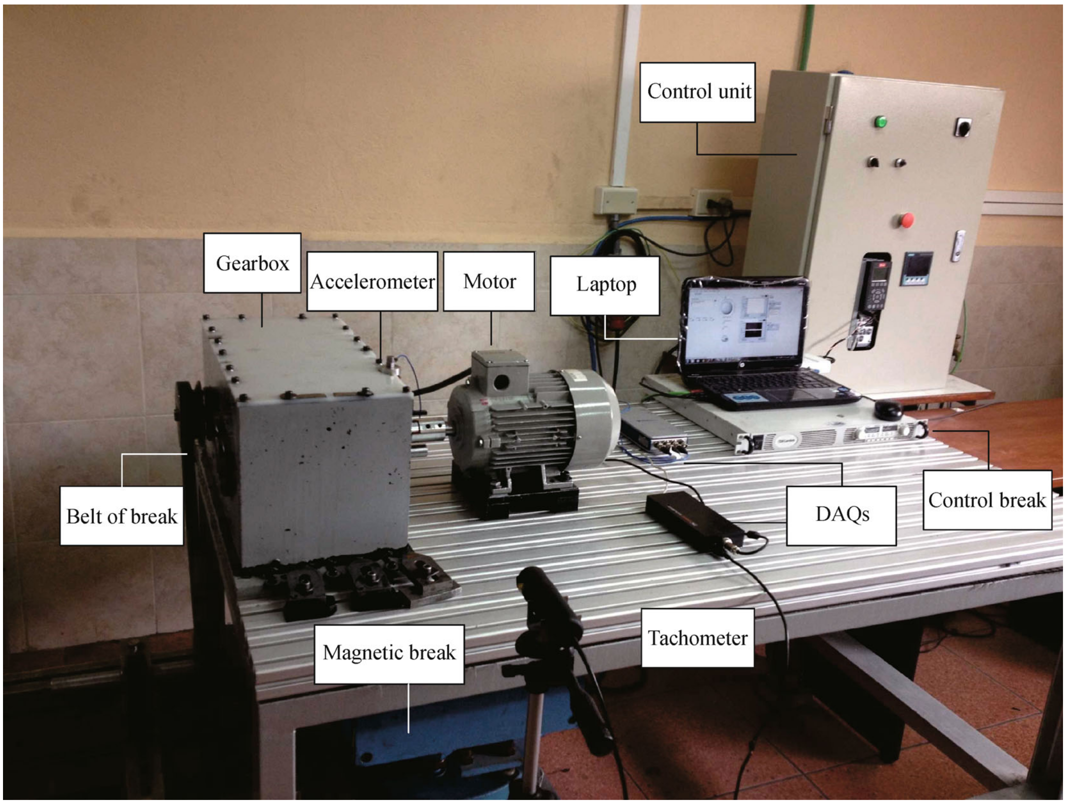


Fig. 3 Configuration of system for failure's simulation

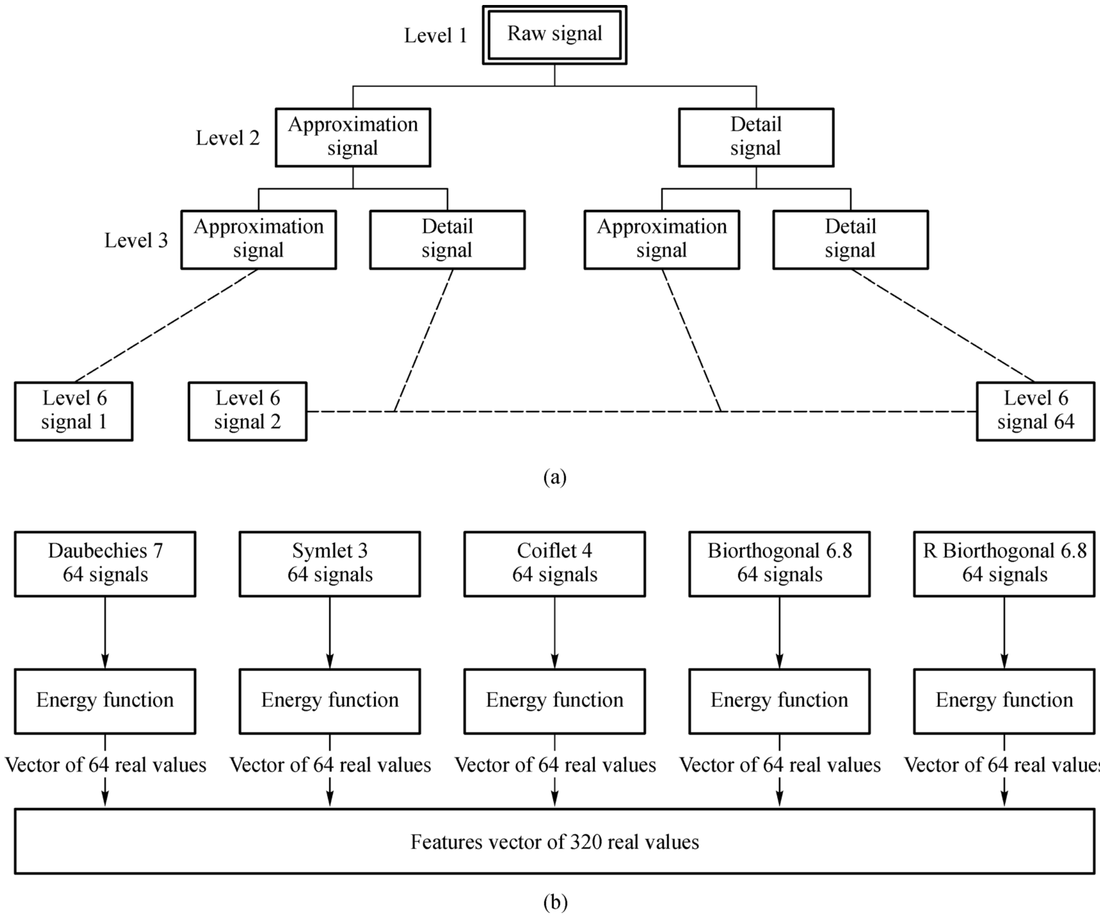


Fig. 4 Feature extraction process. (a) WPD; (b) energy extraction and features vector building

Table 1 The best selected wavelets and the best model parameters

Mother wavelets	oob-error	Feature's number	Tree's number
db7	0.0590	12	1901
db7 + sym3	0.0419	8	1713
db7 + sym3 + coif4	0.0410	7	1671
db7 + sym3 + coif4 + bior6.8	**0.0390**	**17**	**727**
db7 + sym3 + coif4 + bior6.8 + rbior6.8	0.0438	10	1191

3.3 Features selection

The features selection is performed in two stages:

1) In the first stage, the best set of wavelet mother is selected by incrementally adding each subset of features to the data, and training a RF model modifying iteratively the number of trees and the number of random features selected in each node. The oob-error is used as a performance measure to choose the best set of wavelets and the best parameters. Table 1 shows a summary of results of this process. It is identified that the combination db7 + sym3 + coif4 + bior6.8 presents the minimum oob-error with a total of 17 random variables (rho) and 727 planted trees (see Fig. 5 for a comparison of different parameters' sets).

2) The second stage uses the ranking assigned by RF to each feature in order to construct the diagram of importance. Figure 6 shows the importance of features diagram before making the selection of features (Fig. 6(a)), and after the selection where 185 features are selected from the original set of 256 (Fig. 6(b)). Features ranked lower than 0.2 are removed.

3.4 Validation process

The database is divided into a 75%–25% for the training-

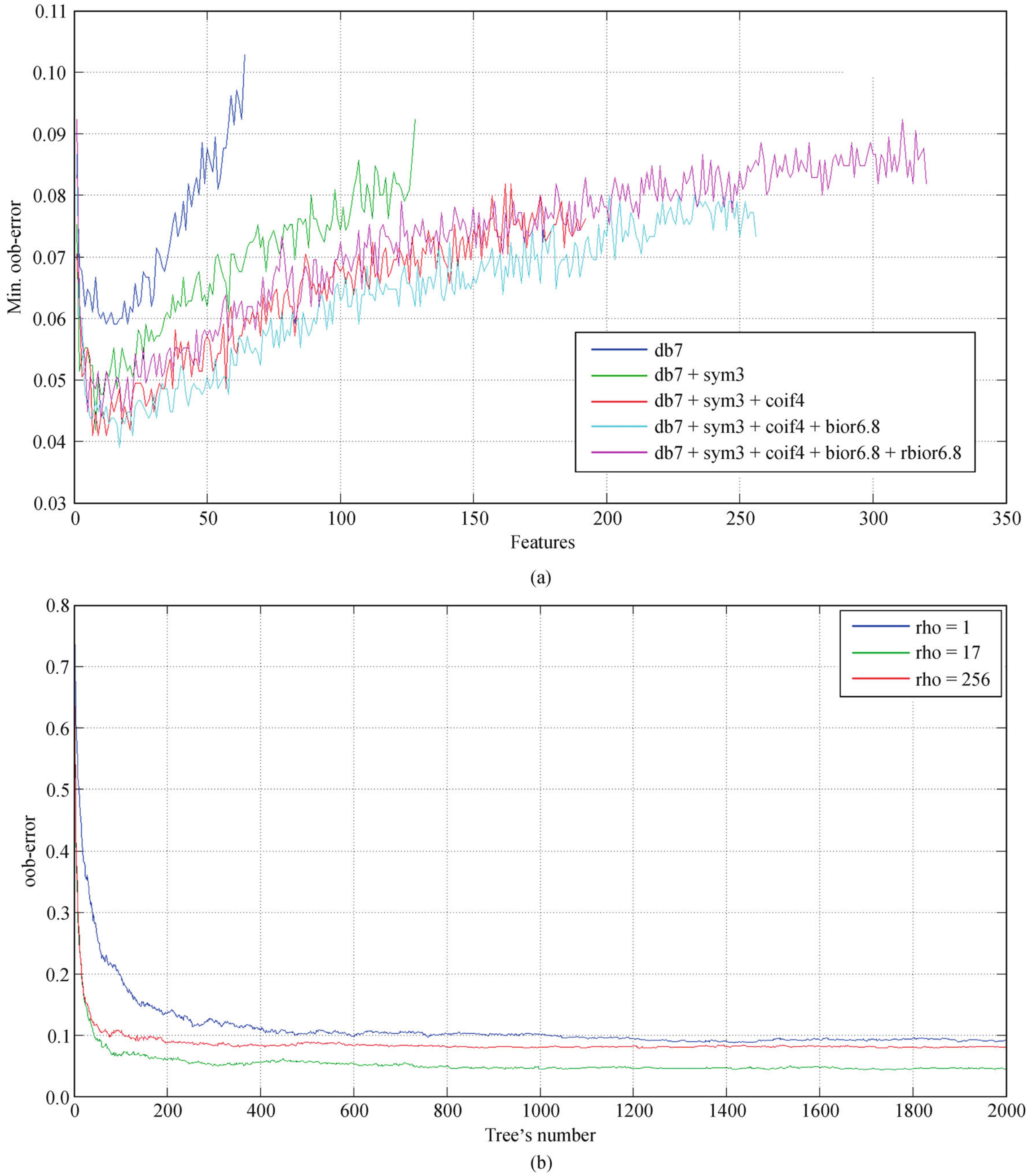


Fig. 5 Curves of training: (a) oob-error for each set of wavelets with variable number of random features; (b) oob-error for the set of the best wavelets for maximum randomness (1), maximum correlation (256) and perfect randomness (17) with a variable number of trees

test sets. Table 2 shows the confusion matrix resulting from the process of validation. The largest error can be found in pinion pitting fault which shows a big confusion with a 10% broken tooth. It is important to note that in Classes 3 and 4 the classifier makes mistakes with a specific fault, allowing focus on improving the accuracy without changing the classifier completely.

The general performance is shown in Table 3. The accuracy analysis leads to the conclusion that the classifier has a great success separating different faults. Through sensibility analysis we can focus specifically on the ability of the classifier to identify samples belonging to a group. In our case, F score weights the relationship between accuracy and sensibility metrics. We have assigned the same importance to both values, showing a more realistic metric without bias introduced for percentages of both, true positives and true negatives.

Another measure of assessment is the area under the

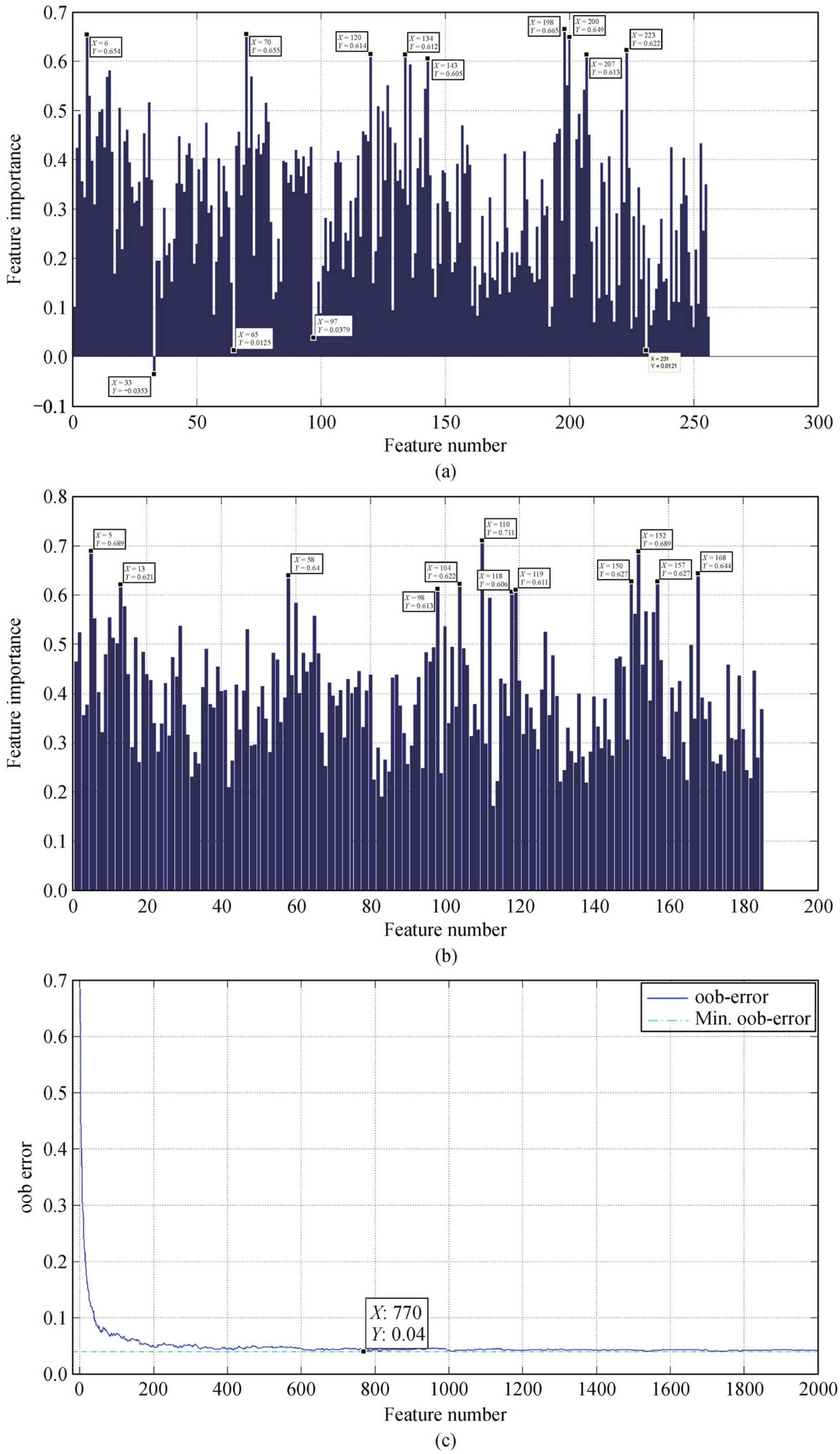


Fig. 6 Selection of features: (a) Importance of features without selection; (b) importance of features with selection; (c) curve of further training to the selection of features

Table 2 Confusion matrix

Class	1	2	3	4	5	6	7
1	37	0	0	0	0	0	0
2	0	37	0	0	0	0	0
3	0	3	34	1	0	0	0
4	0	0	2	35	1	0	0
5	0	0	0	0	38	0	0
6	0	0	0	0	0	36	1
7	0	0	0	0	0	0	37

Table 3 Classifier's performance measures

Metric	Value/%
Accuracy	96.950
Sensibility	97.000
F score	96.975

curve (AUC) of each class against each of the remaining classes, as shown in Table 4, and obtained from the relative operating characteristic curve. From this measure we obtain the degree of classification percentage between each of the classes. Figure 7 shows the separation of data and the importance of each coordinate after the application of the multidimensional scaling (MDS) technique to the proximity matrix.

4 Analysis and results

Figure 5(a) shows that the set of features with the best group of mother wavelets gets always an error lower than the other groups of wavelets, regardless of the parameters of the RF model. Among the possible sets of random features, it is observed that with a specific set of 17 of them we get the minimum oob-error regardless of the number of trees.

The ranking obtained from the importance of features, as shown in Fig. 6(a), shows that the Feature 33, belonging to the wavelet db7, influences the model negatively. After applying the threshold to reduce the number of features according to their importance, an increment of 0.1% of the oob-error is observed (from 3.9% to 4%), a value that is irrelevant for the total classification.

The results shown in Tables 3 and 4 are justified by observing the eigen-values in Fig. 7(b), which shows a high degree of importance up to the sixth coordinate of the scale space; hence, the first 3 dimensions shown in are not enough for an adequate graphic representation of the separation between the data. However, it is easy to see that the classes most clearly separable are normal, tooth broken 100%, tooth broken 10% and gear misalignment. Nevertheless, the joint information provided by MDS and by the eigen-values helps to understand the distribution of data in a high-dimensional space as the one we are considering, so MDS can be seen as an additional and valuable tool in combination with classification models that work on high dimensional structures.

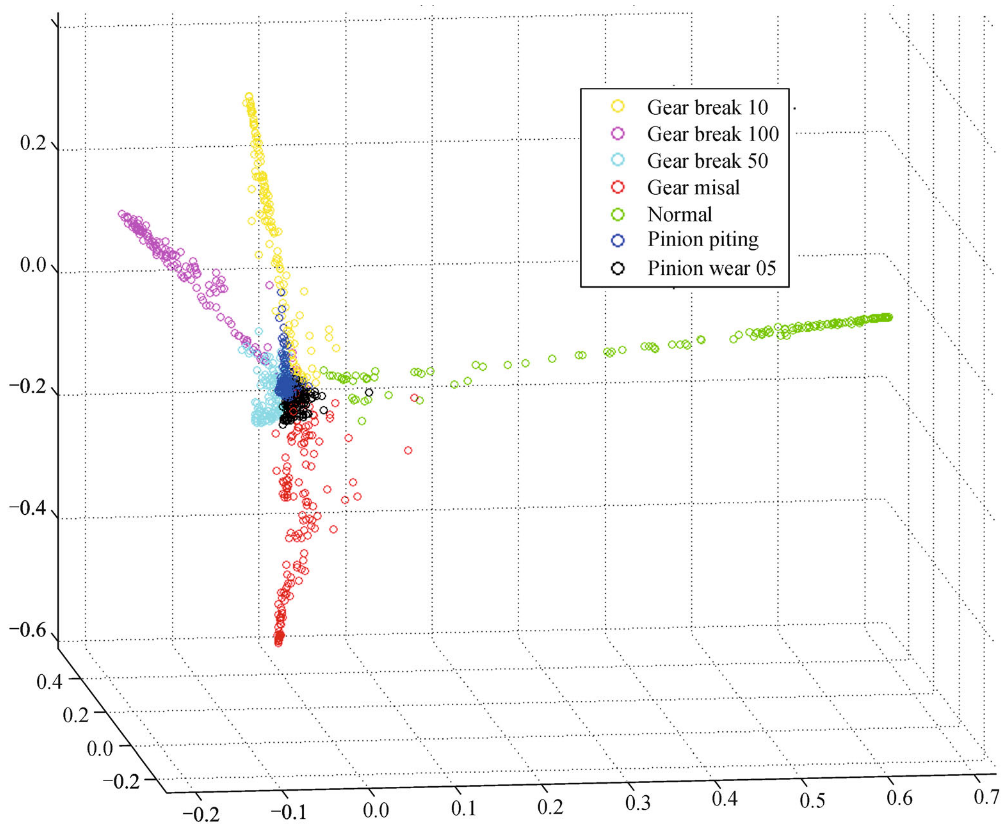
5 Conclusions and future work

The results obtained with this classifier show a lower error rate and over 96% of classification accuracy for the gear faults classes. It must be stressed out the efficiency in the use of wavelet energy coefficients as features predictors.

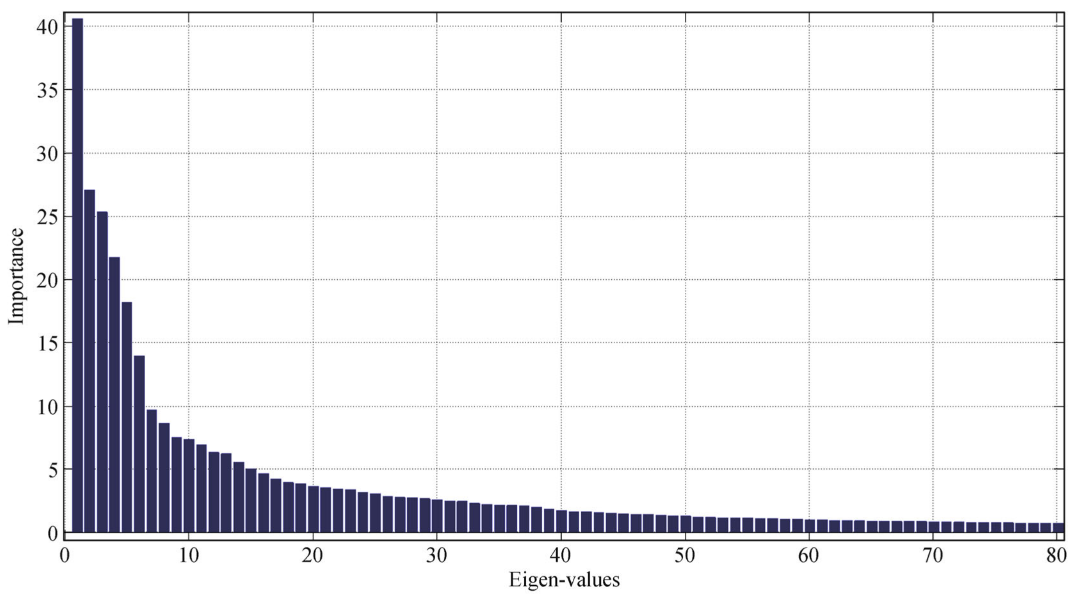
The lowest AUC is located between classes 2 (10% broken tooth) and 3 (pinion pitting), concluding that the energy of these two faults are most similar and can produce more errors for a correct classification. From a computational point of view, the training of the RF model does not present high complexity in time, and it is easy to do a

Table 4 AUC individual unit: %

Class	2	3	4	5	6	7
1	100	100	100	100	100	100
2		96.05	100	100	100	100
3			96.30	100	100	100
4				98.68	100	100
5					100	100
6						98.65



(a)



(b)

Fig. 7 Separation of samples. (a) Location of samples; (b) coordinates eigen-values

parallel implementation both for training stage and for validation stage, since each tree is independent of the others.

The RF classifier has shown great ability to handle high dimensional samples as one of the most appropriate models for their implementation in learning complex patterns that require a very detailed classification.

A weakness of the RF classifier, like all other supervised classifiers used in fault diagnosis tasks, is the inability to classify patterns extracted from signals in the presence of more than one fault at the same time, despite the classifier has been trained with each fault pattern separately. Some information can be obtained from the fact that the underlying process follows a probabilistic model where the overall response is greater than the sum of its parts. This weakness can be addressed by training the classifier with the patterns obtained from mixed fault signals which will aim for a future work.

Other future work we are starting is about the use of RF classifier for fault diagnosis in different types of mechanical elements, as well as the inclusion of a greater number of incipient failures.

Acknowledgements The authors want to express a deep gratitude to the Ministry of Higher Education, Science, Technology and Innovation of the Republic of Ecuador (SENESCYT), for their support with this research work. Fernando Sancho wants to thank to the Prometeo Project from the same institution, and the Project TIC-6064 from Junta de Andalucía (Spain) for their support.

References

1. Walha L, Fakhfakh T, Haddar M. Backlash effect on dynamic analysis of a two-stage spur gear system. Journal of Failure Analysis and Prevention, 2006, 6(3): 60–68
2. Abbes M S, Fakhfakh T, Haddar M, et al. Effect of transmission error on the dynamic behaviour of gearbox housing. International Journal of Advanced Manufacturing Technology, 2007, 34(3–4): 211–218
3. Tian Z, Zuo M, Wu S. Crack propagation assessment for spur gears using model-based analysis and simulation. Journal of Intelligent Manufacturing, 2012, 23(2): 239–253
4. Ebersbach S, Peng Z. Fault diagnosis of gearbox based on monitoring of lubricants, wear debris, and vibration. In: Wang Q, Chung Y W, eds. Encyclopedia of Tribology. New York: Springer, 2013, 1059–1064
5. Rgeai M, Gu F, Ball A, et al. Gearbox fault detection using spectrum analysis of the drive motor current signal. In: Kiritsis D, Emmanouilidis C, Koronios A, et al., eds. Engineering Asset Lifecycle Management. London: Springer, 2010, 758–769
6. Hong L, Dhupia J S. A time domain approach to diagnose gearbox fault based on measured vibration signals. Journal of Sound and Vibration, 2014, 333(7): 2164–2180
7. Rafiee J, Arvani F, Harifi A, et al. Intelligent condition monitoring of a gearbox using artificial neural network. Mechanical Systems and Signal Processing, 2007, 21(4): 1746–1754
8. Sanchez R, Arpi A, Minchala L. Fault identification and classification of spur gearbox with feed forward back propagation artificial neural network. In: Proceedings of the 2012 Andean Region International Conference. Washington, D.C.: IEEE, 2012, 215
9. Barakat M, Lefebvre D, Khalil M, et al. Parameter selection algorithm with self-adaptive growing neural network classifier for diagnosis issues. International Journal of Machine Learning and Cybernetics, 2013, 4(3): 217–233
10. Yang B S, Han T, An J L. ART-KOHONEN neural network for fault diagnosis of rotating machinery. Mechanical Systems and Signal Processing, 2004, 18(3): 645–657
11. Jiang Z, Fu H, Li L. Support vector machine for mechanical faults classification. Journal of Zhejiang University SCIENCE A, 2005, 6 (5): 433–439
12. Jiao B, Xu Z. Multi-classification LSSVM application in fault diagnosis of wind power gearbox. In: Zhang T, ed. Mechanical Engineering and Technology. Berlin: Springer, 2012, 125: 277–283
13. Kang Y, Wang C, Chang Y. Gear fault diagnosis in time domains by using Bayesian networks. In: Melin P, Castillo O, Ramirez E, et al., eds. Analysis and Design of Intelligent Systems using Soft Computing Techniques. Berlin: Springer, 2007, 41: 618–627
14. Breiman L, Friedman J, Olshen R, et al. Classification and regression trees. The Wadsworth and Brooks-Cole statistics-probability series. Boca Raton: Chapman & Hall, 1984
15. Breiman L. Random forests. Machine Learning, 2001, 45(1): 5–32
16. Criminisi A, Shotton J. Classification forests. In: Criminisi A, Shotton J, eds. Decision Forests for Computer Vision and Medical Image Analysis. London: Springer, 2013, 25–45
17. Han X, Yang B S, Lee S J. Application of random forest algorithm in machine fault diagnosis. In: Mathew J, Kennedy J, Ma L, et al., eds. Engineering Asset Management. London: Springer, 2006, 779–784
18. Yang B S, Di X, Han T. Random forests classifier for machine fault diagnosis. Journal of Mechanical Science and Technology, 2008, 22 (9): 1716–1725
19. Karabadji N, Khelf I, Seridi H, et al. Genetic optimization of decision tree choice for fault diagnosis in an industrial ventilator. In: Fakhfakh T, Bartelmus W, Chaari F, et al., eds. Condition Monitoring of Machinery in Non-Stationary Operations. Berlin: Springer, 2012, 277–283