


Gearboxes fault detection under operation varying condition based on MODWPT, Ant colony optimization algorithm and Random Forest classifier

Advances in Mechanical Engineering
2021, Vol. 13(8) 1–12
© The Author(s) 2021
DOI: 10.1177/16878140211043004
journals.sagepub.com/home/ade


Boualem Ikhlef, Chemseddine Rahmoune , Bettahar Toufik and Djamel Benazzouz

Abstract

Gearboxes are massively utilized in nowadays industries due to their huge importance in power transmission; hence, their defects can heavily affect the machines performance. Therefore, many researchers are working on gearboxes fault detection and classification. However, most of the works are carried out under constant speed conditions, while gears usually operate under varying speed and torque conditions, making the task more challenging. In this paper, we propose a new method for gearboxes condition monitoring that is efficiently able to reveal the fault from the vibration signatures under varying operating condition. First, the vibration signal is processed with the Maximal Overlap Discrete Wavelet Packet Transform (MODWPT) to extract the AM-FM modes. Next, time domain features are calculated from each mode. Then the features set are reduced using the Ant colony optimization algorithm (ACO) by removing the redundant and unimportant parameters that may mislead the classification. Finally, an ensemble learning algorithm Random Forest (RF) is used to train a model able to classify the fault based on the selected features. The innovative aspect about this method is that, unlike other existing methods, ACO is able to optimize not only the features but also the parameters of the classifier in order to obtain the highest classification accuracy. The proposed method was tested on varying operating condition real dataset consisting of six different gearboxes. In the aim to prove the performance of our method, it had been compared to other conventional methods. The obtained results indicate its robustness, and its accuracy stability to handle the varying operating condition issue in gearboxes fault detection and classification with high efficiency.

Keywords

Rotary machines, gearboxes, fault detection, feature extraction, selection, optimization, classification

Date received: 13 April 2021; accepted: 10 August 2021

Handling Editor: James Baldwin

Introduction

Gearboxes defects and degradation have a huge impact on the efficiency of rotating machines. These faults can be generated from various factors such as improper assembly, poor lubrication, corrosion, and overload¹ and they can lead to substantial economic losses and serious risks and dangers on the working staff.

For this reason, gears fault diagnosis has always been a research subject where many techniques and

Solid Mechanics and Systems Laboratory (LMSS), University M'hamed Bougara, Boumerdes, Algeria

Corresponding author:

Boualem Ikhlef, Solid Mechanics and Systems Laboratory (LMSS), University M'hamed Bougara, rue de l'indépendance, Boumerdes 35000, Algeria.
Email: b.ikhlef@univ-boumerdes.dz



approaches were proposed to address this issue in order to increase the efficiency and reliability of the addressed system.

It has been noticed that vibration signal has a wide utilization and shows dominance over temperature, current and acoustic emissions for fault diagnosis.² This is due to the simplicity of its acquisition and the importance of the information it provides concerning the source and the gravity of the fault.³

Once the fault occurs, impulses periodically appear in the vibration signal whenever the damaged part of gear comes in contact, this repetition is known as Fault Characteristic frequency (FCF).

In gear fault detection, FCF extraction is applied based on many signal processing techniques, for example in Yang et al.,⁴ Yang and Chen¹ propose a precise gearboxes diagnosis method based on multi-feature and BP-AdaBoost, used Radial Basis Particle Filter as an extracted signal denoising technique, to pretreat it for further diagnostic classification,⁵ introduced an effective fault component separation method that integrates ensemble empirical mode decomposition (EEMD; an adaptive signal decomposition method in time-frequency domain) with independent component analysis (ICA; a blind source separation technique).

The efficiency of the previous techniques in fault detection is proved at constant speed and torque, but gears generally operate in severe environments under non-stationary working conditions, therefore the FCF changes, which makes the applicability of such methods impossible and non-accurate.

In recent years, more studies were founded to diagnose gears in different variable operating condition to approach the real particle case.

Many research groups exploit a phase reference signal obtained from an encoder or a tachometer to remove speed variation effect.⁶ Hun et al.⁷ introduced a deep belief network (DBN) algorithm for gear fault diagnosis based on wavelet packet energy entropy (WPEE) and multi-scale permutation entropy (MPE). Zheng et al.⁸ worked on the extraction of gearbox fault feature of wind turbine under variable speed condition using improved adaptive variational mode decomposition (VMD). In the research work published by Gougam et al.⁹ the decomposition of the initial signal into different modes is done using EWT. Then the most significant modes are selected to reconstruct a new signal relying on Kurtosis. In the next step, time domain features are extracted; and the Fuzzy Logic System (FLS) is utilized for the classification of bearings faults. This technique showed a high performance in fault classification. However, for EWT decomposition, an inappropriate selection of the modes number may lead to disagreeable decomposition results. Moreover, the linearity of the wavelet filtering bandwidth makes it non-adaptive for all the cases.¹⁰ Furthermore, many

condition monitoring and fault diagnosis works preferred DWT as a signal decomposition tool,^{11,12} where the analyzed data is decomposed with a band-pass filter in time and frequency domains into a collection of signals with a particular frequency band.¹³ Unfortunately, the dyadic step in the down-sampling process seems to be the main limit of DWT. This issue is addressed using the Maximal overlap discrete wavelet transform (MODWT) as an optimized version of DWT to overcome the down-sampling process.^{14,15} Yet, same as DWT, this technique still endures the poor frequency resolution. A better resolution is ensured using Maximal overlap discrete wavelet packet transform (MODWPT). In this method, the decomposition of the complicated signal into several single components and the property of circular shift equivariance are insured for gears condition monitoring in various working condition.¹⁶ Uniform frequency bandwidths are provided using MODWPT, this helps to overcome the time-variant transformation and also allows the reconstruction of the original signal and maintain the necessary information.

Nevertheless, for the purpose of reducing the number of FLS entries, numerous potentially important pieces of information are obscured if they belong to the excluded modes. To overcome this matter, optimization is considered instead of eliminating entire modes based directly on a single parameter. In some studies, optimization is found to be used as a main step in condition monitoring considering the improvement it imports to the classification performance.

In this paper, a new method is proposed to diagnose and monitor gears in various operating conditions. The innovative aspect about this method is that, unlike other existing methods, ACO is able to optimize not only the features but also the parameters of the classifier in order to obtain the highest classification accuracy.

The flowchart of the proposed method is shown in Figure 1. It begins with processing the vibration signal using MODWPT to extract the different AM-FM modes. Then time domain features are extracted from those modes. After that, the optimization of the features set is done by eliminating the unimportant parameters by Ant colony Algorithm. Finally a model is trained to detect the fault by the supervised learning method "random forest." This work is done on vibration data and it shows a high performance.

Experimental description

In this section, a speed reducer with a gear ratio of 25/56 is considered as a test bench (Figure 2). A nominal speed of 3600 r/min electric dc motor is considered as a source of motion between the two shafts and different

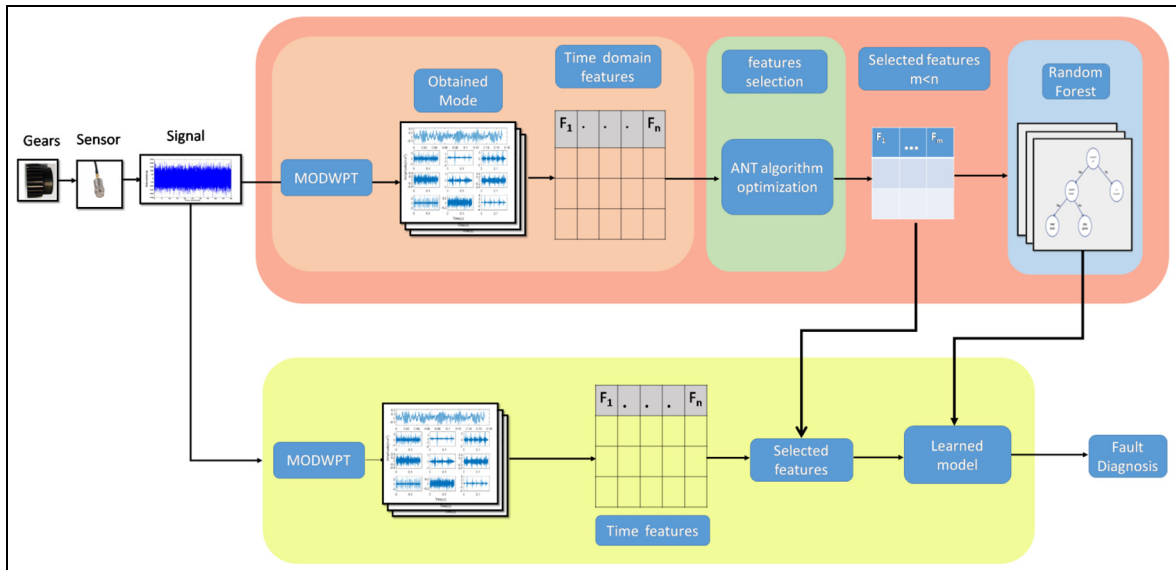


Figure 1. Flowchart of proposed method.

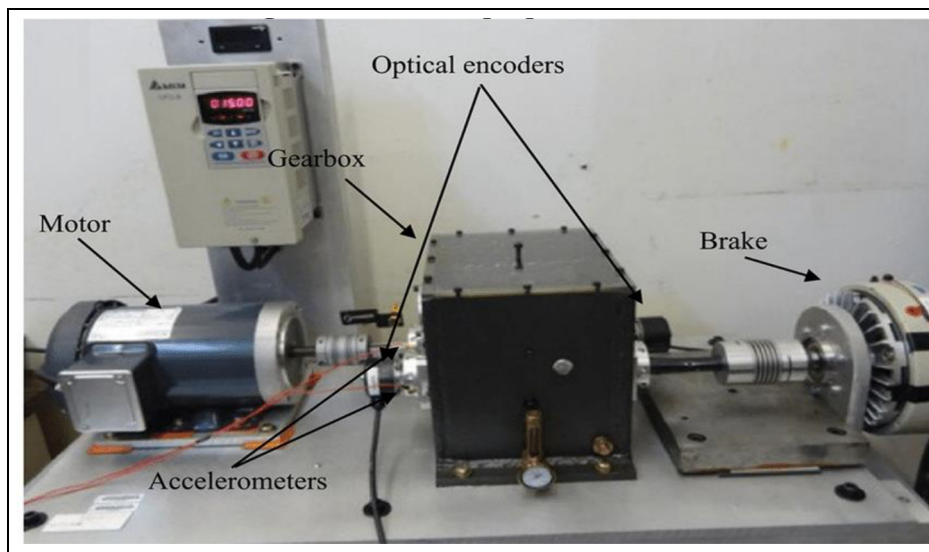


Figure 2. The gearbox test bed.

resistive torques are generated by a magnetic power brake that is coupled to the output shaft.^{17,18}

The efficiency of the suggested method is tested using six pinions with different health states. The first one is a faultless pinion, and it is referred as good (G), while the rest have various types of defects, such as a tooth root crack (TRC), a chipped tooth in length (CTL), a chipped tooth in width (CTW), a missing tooth (MT), and general surface wear (GSW) (Figure 3).^{17,18}

Three pinions are installed simultaneously, on the input shaft of the gearbox. With a simple axial movement of the wheel of the output shaft, the engagement of each of them is achieved (Figure 4). Two

accelerometers (sensitivity: see Table 1 above 100 mV/g) are radially installed to record vibration signals, in horizontal and vertical positions on the bearing case of the output shaft. The accelerometer channels time sampling frequency is equal to 125 kHz, the sampling frequency of the anti-aliasing filter is 27 kHz, and the acquisition duration is equal to 30 s.

The accelerometer signals have been collected for several operating conditions under different loads and different rotation speeds (see Table 2). Figure 5 shows the acceleration vibration signals recorded from pinions with different gear state for an operating speed equal to 900 r/min with 11-Nm load. From Figure 5, it can be

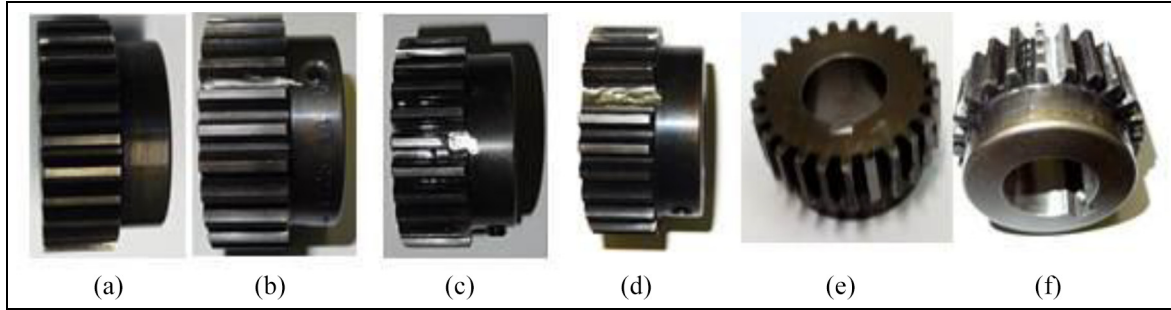


Figure 3. Six pinion states: good (a), tooth root crack (b), chipped tooth in length (c), chipped tooth in width (d), MT (e), and GSW (f).

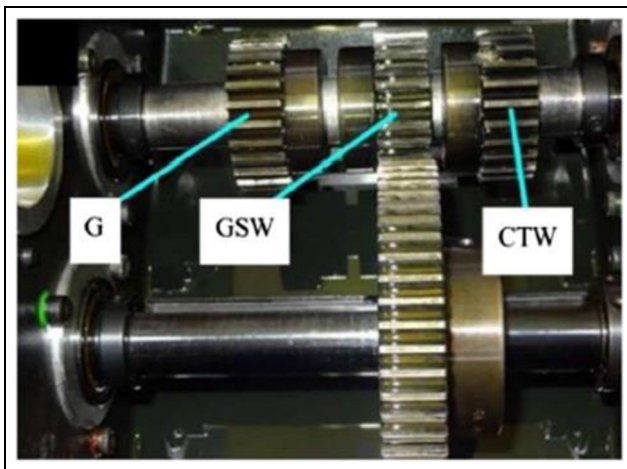


Figure 4. Single stage gear pinions location in the gearbox.

clearly see that the five different defects do not show a significant signature in the vibration signal. No significant increase in the energy of the temporal signal is noticed.

Signal pre-processing and feature extraction

Maximal overlap discrete wavelet packet transform

Let $X = [X_0, X_1, \dots, X_{N-1}]$ be a column vector of sampled sequences of a continuous-time data x , and N is a power of 2. The even-length scaling (low-pass) filter $\{g_l: l = 0, \dots, L-1\}$ and the wavelet (high-pass) filter $\{h_l: l = 0, \dots, L-1\}$ are used to obtain the DWT of the sampled vector.^{14,15} These even-length filters satisfy the following equation:

$$\sum_{l=0}^{l-1} g_l^2 = \sum_{l=0}^{l-1} g_l g_{l+2n} = \sum_{l=-\infty}^{\infty} g_l g_{l+2n} = 0 \quad (1)$$

Table 1. Geometrical parameters of the gear system.

Parameter	Pinion	Gear
Number of teeth	25	56
Pressure angle (°)		14.5
Face width (mm)		19.05
Diametric pitch		12
Material		1140 carbon steel
Outside diameter (mm)	57.2	122.7582
Pitch diameter (mm)	52.9082	101.6

Table 2. Operating conditions.

Gear state	Speed (r/min)	Load (Nm)
G	900, 1200, 1500, 1800, 2400	0, 8, 11
MT	900, 1200, 1500, 1800, 2400	0, 8, 11
TRC	900, 1200, 1500, 1800, 2400	0, 8, 11
GSW	900, 1200, 1500, 1800, 2400	0, 8, 11
CTW	900, 1200, 1500, 1800, 2400	0, 8, 11
CTL	900, 1200, 1500, 1800, 2400	0, 8, 11

The two filters are related for being quadrature mirror of all nonzero integers n as in:

$$h_l = (-1)^l g_{L-1-l} \quad (2)$$

Or

$$g_l = (1 -)^{l+1} h_{L-1-l} \quad (3)$$

The j th level wavelet and scaling coefficients for $t = 0, \dots, N-1$ are given by:

$$V_{j,t} = \sum_{l=0}^{l-1} g_l V_{j-1, (2t+1-l) \bmod N_{j-1}} \quad (t = 0, \dots, N_{j-1}) \quad (4)$$

$$W_{j,t} = \sum_{l=0}^{l-1} h_l V_{j-1, (2t+1-l) \bmod N_{j-1}} \quad (t = 0, \dots, N_{j-1}) \quad (5)$$

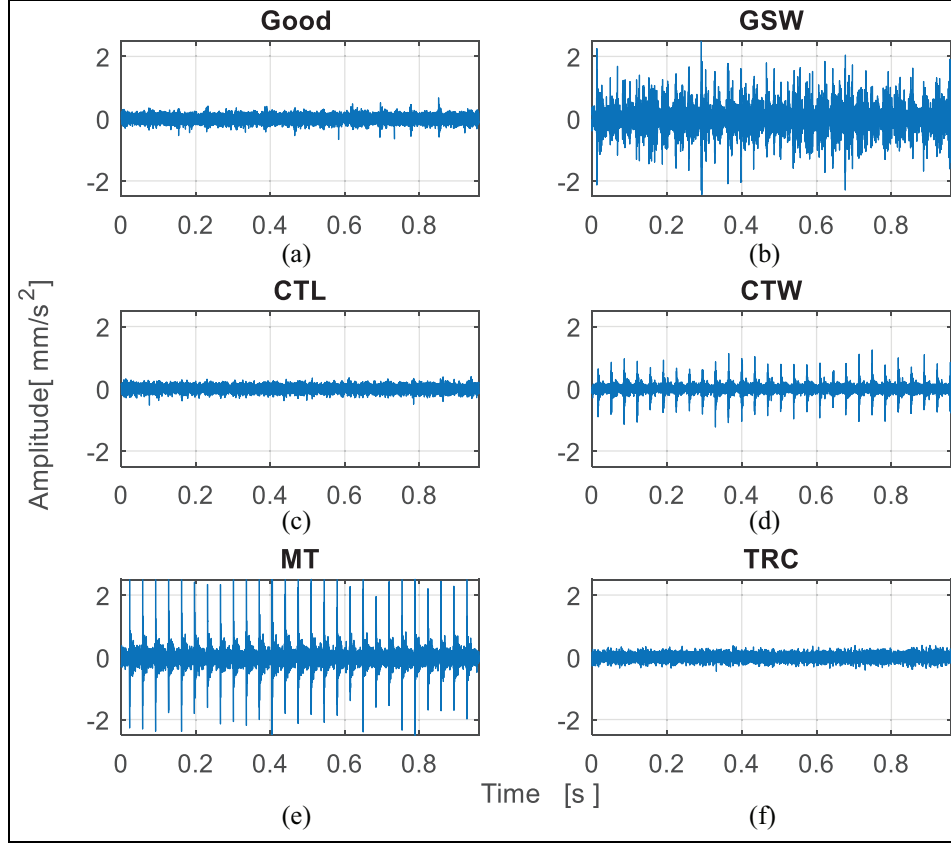


Figure 5. Pinion vibration signals of gearbox with different faults (speed: 900 r/min, load: 1 l): G (a), TRC (b), CTL (c), CTW(d), MT (e), and GSW (f).

where MOD stands for the modulus after division.

MODWT is an enhanced version of the DWT. Differently from DWT, MODWT is well defined for any sample size N at every level j .

Energy conservation is ensured by scaling the defining filters as in:

$$\tilde{g}_l = \frac{g_l}{\sqrt{2}} \quad (6)$$

$$\tilde{h}_l = \frac{h_l}{\sqrt{2}} \quad (7)$$

Thus, equation (1) becomes

$$\sum_{l=0}^{l-1} \tilde{g}_L^2 = \frac{1}{2}, \quad \sum_{l=0}^{l-1} \tilde{g}_L \tilde{g}_{L+2n} \quad (8)$$

The expression of the quadrature mirror filters becomes as follows:

$$\tilde{h}_l = (-1)^l \tilde{g}_{L-1-l} \quad (9)$$

$$\tilde{g}_l = (-1)^l \tilde{h}_{L-1-l} \quad (10)$$

To avoid the down-sampling problem, MODWT uses new filters by insuring $2^{j-1} - 1$ zeros between the elements of $\{\tilde{g}_l\}$ and $\{\tilde{h}_l\}$. The scaling coefficients $\{V_{j,t}^M\}$ are produced by the pyramid algorithm of MODWT and the MODWT wavelet coefficients $\{M_{j,t}^M\}$ as shown in

$$V_{j,t} = \sum_{l=0}^{l-1} \tilde{g}_l V_{j-1, (2t+1-l) \bmod N} \quad (t = 0, \dots, N-1) \quad (11)$$

$$W_{j,t} = \sum_{l=0}^{l-1} \tilde{h}_l V_{j-1, (2t+1-l) \bmod N} \quad (t = 0, \dots, N-1) \quad (12)$$

MODWPT is a further developed method adopted to ensure a perfect resolution at high frequencies, and $W_{j,n} = \{W_{j,n,t}, t = 0, \dots, N-1\}$ is the sequence of MODWPT coefficients at level j and frequency-index n . $\{W_{j,n,t}\}$, is produced using the following equation:

$$W_{j,n,t} = \sum_{l=0}^{l-1} \tilde{f}_{n,l} W_{j-1, [\frac{t}{2}], (t-2^{j-1}) \bmod N} \quad (t = 0, \dots, N_{j-1}) \quad (13)$$

Table 3. Table of extracted features.

Name	Formula
Root mean square	$\sqrt{\frac{1}{n} \sum_{i=1}^n x_i ^2}$
Crest factor	$C = \frac{\ x\ _{\infty}}{\sqrt{\frac{1}{n} \sum_{i=1}^n x_i ^2}}$
Peak to peak	$Max(x) - Min(x)$
Skewness	$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$
Kurtosis	$\frac{E(x-\mu)^4}{\sigma^4}$
Entropy	$-\sum_i p_i \log_2(p_i)$
Mean	$\frac{1}{N} \sum_{i=1}^N A_i$
Std	$\sqrt{\frac{1}{N-1} \sum_{i=1}^N A_i - \mu ^2}$
Var	$\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^3$
Root sum square	$X_{rss} = \sqrt{\sum_{n=1}^N X_n ^2}$
Max	$Max x_i $
Min	$Min x_i $

$$V_{j,t} = \sum_{l=0}^{l-1} \tilde{g}_l V_{j-1, (2t+1-l) \bmod N_{j-1}} \quad (t = 0, \dots, N_{j-1}) \quad (14)$$

Where $\tilde{f}_{n,l} = \tilde{g}_l$ when $n \bmod 4 = 0$ or 3 , while $\tilde{f}_{n,l} = \tilde{h}_l$ when $n \bmod 4 = 1$ or 2 .

Time domain features

In this study, after the acquisition of vibration signals from the test gear, the signal was decomposed to sixteen modes using MODWPT. For each mode, twelve statistical features from the time domain are extracted as fault signatures. Hence, a total of 192 features are extracted. These features mathematical formulas are listed in Table 3.

Features selection and classification

Many irrelevant or redundant attributes can be found in the extracted feature dataset. A satisfying precision for gear faults prediction or classification can be ensured using feature-selection process by removing irrelevant, redundant, or noisy features and selecting those that contains maximum useful information. Additionally, this leads to improved learning accuracy and fault classification process.¹⁹

In this article, the proposed method is based on the use of the ACO whose objective function is based on the RF classification algorithm.

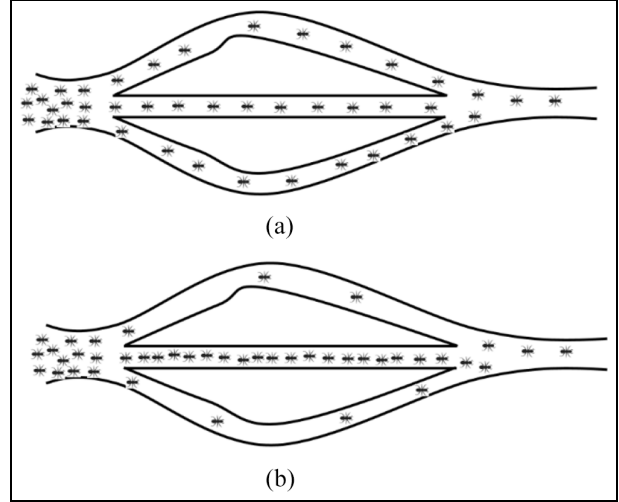


Figure 6. Basic explanation of the Ant colony algorithm (a), Before finding the shortest way (b), After finding the shortest way.

Ant colony algorithm (ACO)

In the field of swarm intelligence, Ant colony is considered as a typical algorithm for optimization. The “swarm” is a group that can indirectly communicate by changing the local environment, and be able to solve the distribution problems in cooperation. We mean by “intelligence” an agent that shows intelligent behavior compared to the rest of the group through cooperation. The so-called “ants” is considered the basic unit of the swarm and hence algorithm.²⁰

In ant’s community, the used medium to communicate between individuals is pheromone. In the feeding process, and in order to mark the path, the moving ant lays some pheromone on the ground. The basic principle is that other ants will follow the laid trail already marked by a previously ant’s pheromone, the encountered trail will be then forced and increased by its own pheromone. The probability of choosing the same path increases according to the number of ants releasing their pheromones in it, the process can be expressed as a positive feedback loop based on the preceding steps.²¹ When reaching the food source, ants follow the same route to their nest. After finishing their tasks, ants pheromones evaporate gradually over time, this is called pheromone evaporation mechanism. The illustration of the based positive feedback loop identification of the shortest path is shown in Figure 6. The stronger pheromone trail left by the preceding ants will be chosen as the shortest path compared to the other one and more ants will reinforce it as explained in Figure 6(b).

The mathematical description of ACO

ACO was firstly introduced by Dorigo et al.²² in the early 1990s. It was considered a new nature-inspired metaheuristic method intended to solve TSP problems.

The TSP problem can be described as the problem of finding a minimal length closed tour that visits each town once. ACO usage in TSP solution will be described in what follows.

Pheromone updates must be considered after each round: this includes the evaporated quantity per unit length and laid on the edge by the ant. Ants will ignore the poor path selected before thanks to pheromone evaporation mechanism. Thus, early local optimization will be avoided.

ρ is an evaporation coefficient. When the evaporation rate is set to 1, there is no pheromone evaporation, and not easy to get convergence. But setting ρ too low is prone to get a local best answer.

The intensity of pheromone on path- ij at time $t + 1$ is given by equation (15)²³:

$$\tau_{ij}(t + 1) = \rho \cdot \tau(t) + \Delta\tau_{ij}(t, t + 1) \quad (15)$$

The transition probability for the k -th ant from town i to town j as equation (16)²¹:

$$\rho_{ji}^k = \frac{\sum_{i=1}^n [\tau]^\alpha [n]^\beta}{[\tau]^\alpha [n]^\beta} \quad (16)$$

The trail update pattern determines the three categories that the field of the ant system (AS) can be divided to: ant-cycle, ant-quantity, and ant-density algorithms, their formulas are given by equations (17) and (19).²³ The ant-cycle model shows that each ant lays its trail at the end of the tour, but the other two model sup-dated the trail after each step, which explains the wide use of the Ant-cycle algorithm and the elimination of the two other models.

Where $\Delta\tau_{ij}^k$ is the quantity per unit of length of pheromone laid on edge (i, j) by the k -th ant between time t and $t + 1$; Q is a constant; d_{ij} is the Euclidean distance between i and j ; L^k is the tour length of the k th ant.

ANT-quantity:

$$\Delta\tau_{ij}^k(t, t + 1) = \begin{cases} \frac{Q_s}{L^k} \\ 0 \end{cases} \quad (17)$$

ANT-density:

$$\Delta\tau_{ij}^k(t, t + 1) = \begin{cases} \frac{Q_2}{0} \\ 0 \end{cases} \quad (18)$$

ANT-cycle:

$$\Delta\tau_{ij}^k(t, t + n) = \begin{cases} \frac{Q_s}{L^k} \\ 0 \end{cases} \quad (19)$$

In cluster analysis, the clustering center value equals to the average of objects belonging to the class. Numerous methods be used to calculate the distance between the sample and the clustering center, such as Euclidean

distance, binary angle cosine law, cosine angle distance method. The Euclidean distance is chosen to be used in this paper. The physical meaning of Euclidean distance is expressed as m -dimensional real space distance of two points, which can be expressed as in equation (20).

$$\rho = \sqrt{\sum_{i=1}^m (x_i - y_i)^2} \quad (20)$$

Random Forest

Random Forest is considered as a type of integrated tree classifier.^{24,25}

Figure 7 shows the Random Forest flow chart and its detailed algorithm is described as follows.

RF is a popular supervised and ensemble learning method based on N decision tree constructed using bagging (boots trap aggregating) where each tree uses a random sample of the data and each node of the tree is split depending on the best variable in the input subset of features which is determined by Gini Index as an attribute selection measure²⁶ after calculating the impurity of attributes with respect to the classes by equation (21).

$$Gini\ Index = 1 - \sum_{i=1}^n (P_i)^2 \quad (21)$$

Where p_i denote an element's probability to be classified for a distinct class.

The following lines summarizes the RF steps. The RF takes into consideration the prediction of every decision tree then outputs the best result after voting. Figure 7 summarizes the RF algorithm.

Results and discussion

Our method was firstly tested with twenty features; the obtained results are represented in Figures 8 and 9. It can be clearly noticed in the Confusion matrix that the classification accuracy is 99.1667%.

Several tests were conducted with different numbers. The classification accuracy was saved in each test aiming to determine the optimal number of the selected features. Figure 10 represents the classification accuracy variation according to the number of features. It can be concluded that the best classification accuracy equals to 99.89 and it is obtained when the features number of is 15.

In order to prove the efficiency of the proposed method, it was put in comparison with other techniques which are KNN, DT, and CNB. The number of features is fixed at 15 for all methods.

It worth mentioning that experiment in this paper is conducted under the condition of randomly select

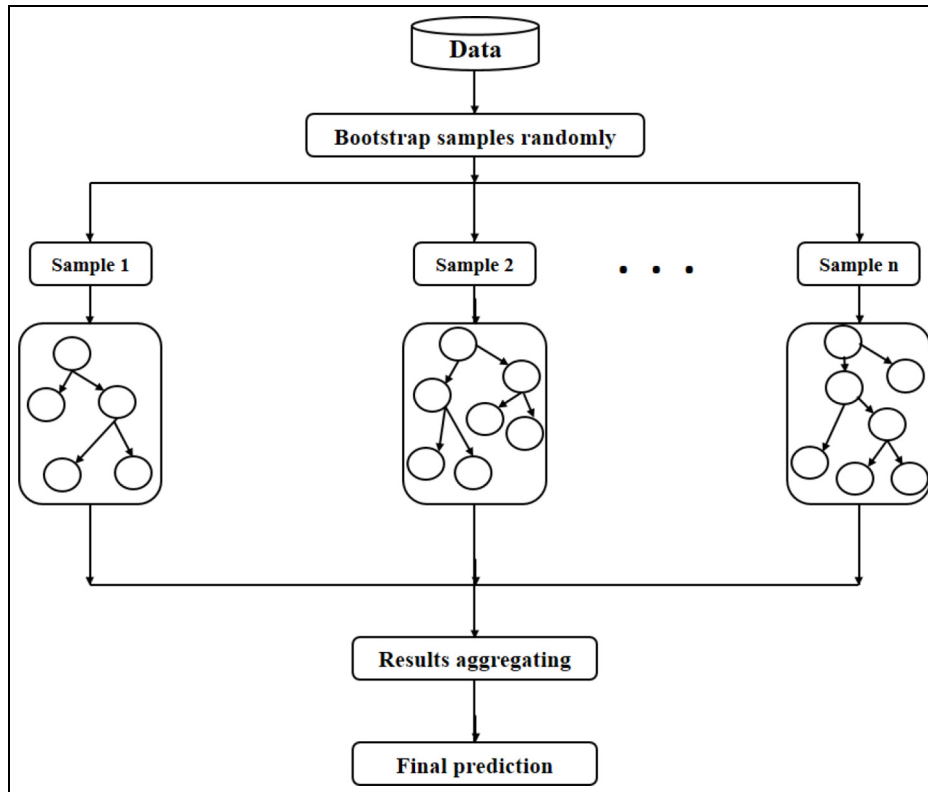


Figure 7. The flow chart of the Random Forest method.

Algorithm 1 Random Forest pseudocode.

Require: *Training_data*, *Testing_data*, *Training_labels*,
Testing-labels, N_{tree}
Ensure: Accuracy
For $i = 1 : N_{tree}$ **do**
Sample_i; bootstrap samples from *Training_data*
End for
for $i = 1 : N_{tree}$ **do**
 build unpruned classification tree,
end for
for $i = 1 : N_{tree}$ **do**
tree_i(*Testing_data*); predict using N_{tree} tree
End for
 Aggregate predictions of N_{tree} tree
prediction \leftarrow majority of votes
 $A \leftarrow$ size(*prediction*)
correct $\leftarrow 0$
for $i = 1 : A$ **do**
if(*prediction*(i) == *label*(i)) **then**
correct \leftarrow *correct* + 1
end if
end for
 Accuracy \leftarrow *correct*/ A

samples, where the average value of 15 experiments is taken, and the maximum and minimum values of classification accuracy are given. Furthermore, the standard

deviation of classification accuracy is considered to analyze the stability of the classification method.

The MODWPT extracted features is the considered input for each classification method. The minimum number of father nodes of DT is 5. The nearest neighbor number of KNN is $K = 5$. The classification results are shown in Table 4.

From the confusion matrix figure, we can see that the CNB has the lowest classification accuracy; we can show clearly that the CNB presents misclassification in the entire category. This misclassification is 36.84% of the testing samples are misclassified for the fourth category and around 25%, 31%, and 25% of the testing samples for the third, the fourth, and the sixth category respectively, see Figures 11 and 12 above.

KNN classification results indicate that 28 samples are misclassified, with an overall classification accuracy of 95%.

in the classification results using DT, 20% of testing samples are misclassified, among which 6.81% of category 2 are classified as category 3, 3.39% of category 3 are classified as category 2, 1.69% of category 3 are classified as category 5, 2.113% of category 4 are classified as category 1, 6, 2.39% of category 6 are classified as category 1, and 1.19% of category 6 are classified as category 4.

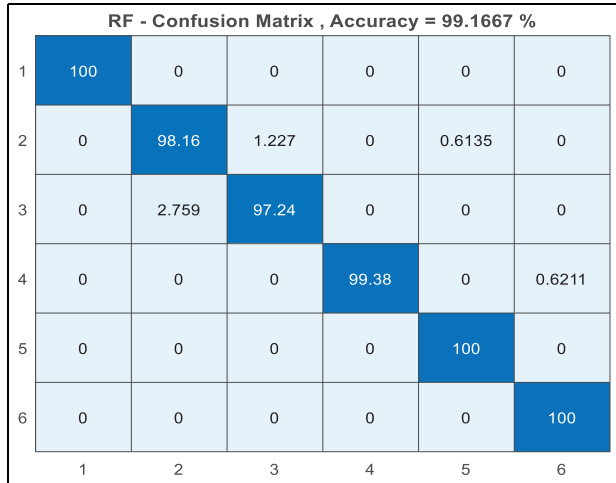


Figure 8. RF confusion matrix with 20 features.

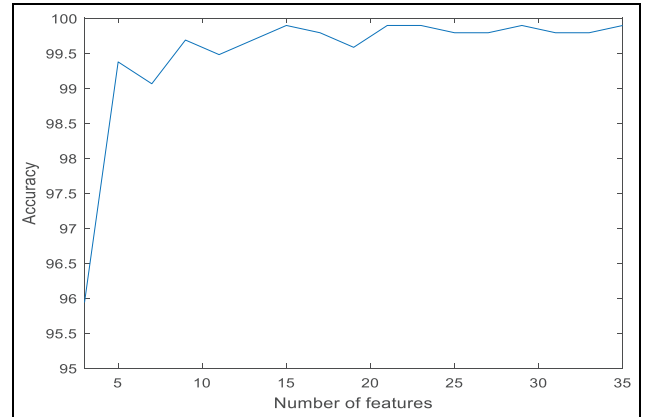


Figure 10. Accuracy classifications (RF) versus number of selected features.

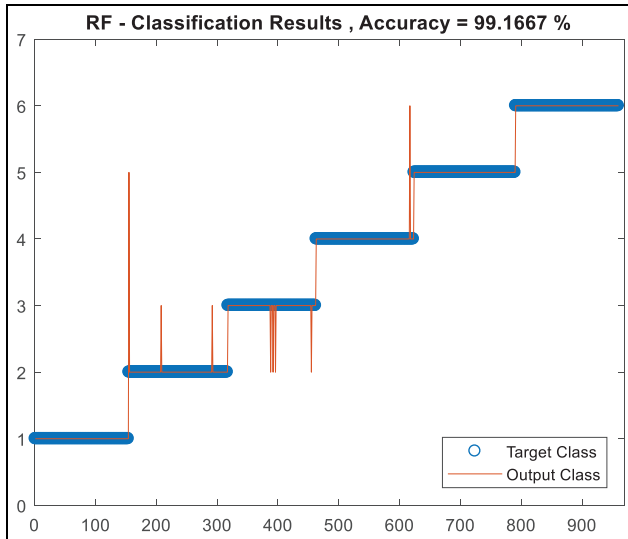


Figure 9. Plot of the classification results with 20 features.

RF results classification accuracy is 99.89% with only one misclassified samples.

Figure 13 represents the variation of the accuracy according to the number of features for different classifier that is KNN, DT, CNB, and RF. By comparing RF with other classifiers, we can see that RF always gives the best accuracy regardless of the number of features.

Conclusion

Rotating machines fault diagnosis is becoming increasingly important through time because of the fast

Table 4. Classification results of various methods.

	RF	KNN	DT	CNB
Max	99.8958	99.1667	98.4375	78.5417
Min	99.3750	86.7700	97.5000	73.1042
Moy	99.6597	96.5950	98.0486	76.4375
STD	0.1497	2.9857	0.2655	1.3640

KNN: *k*-nearest neighbor; DT: decision tree; RF: Random Forest.

development of industries. This is to help operators in troubleshooting by identifying and localizing the underlying problems. Especially that any delay or a misdiagnosis can put the human safety, the machine’s state and the industry’s economy is a real danger. In spite of the great interest this field has gained, only a little focus was devoted for machines diagnosing under variable operating condition which represents the practical case in most industries. In this paper, a new gearbox fault diagnosis method based on MODWPT-ANT and RF classifier is proposed to diagnose various faults of gearbox under variable operating condition. The effectiveness of the proposed method is validated by recognizing six fault types of gearbox. Compared with other existing classification methods, the obtained experimental results using RF classifier indicate that the proposed method provides an alternative way for gearbox health monitoring.

Under the premise of the same input, the RF classifier is always higher than that of DT, and the classification effect is better. By comparing with CNB and KNN, the proposed method has higher classification accuracy and can be better used for gearbox fault diagnosis, and the classification accuracy reaches 99.89%.

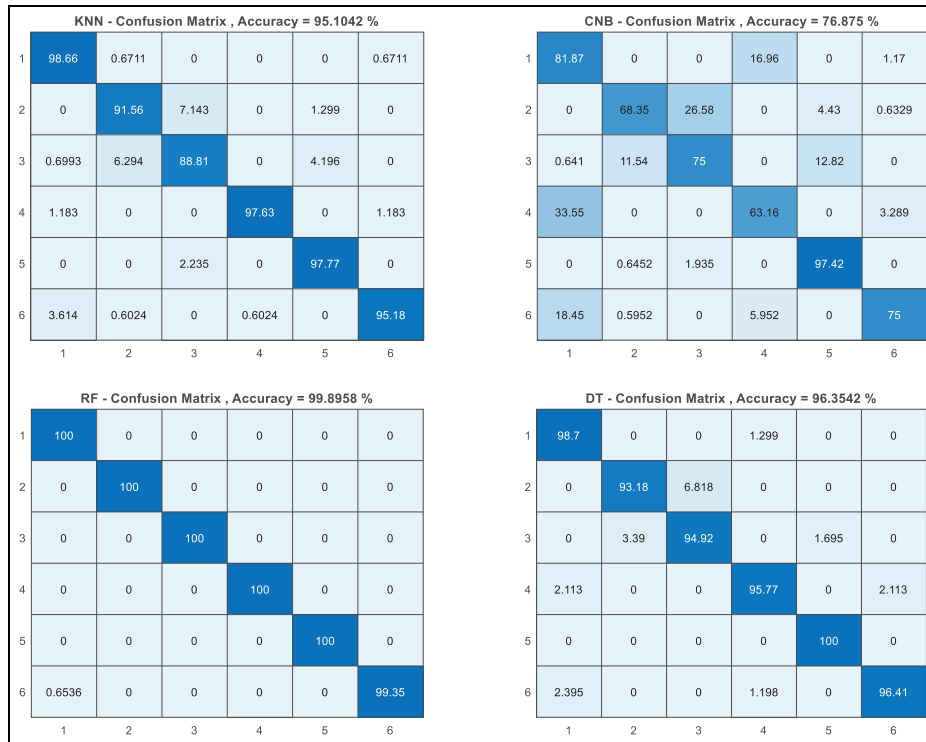


Figure 11. Confusion matrixes with 15 features.

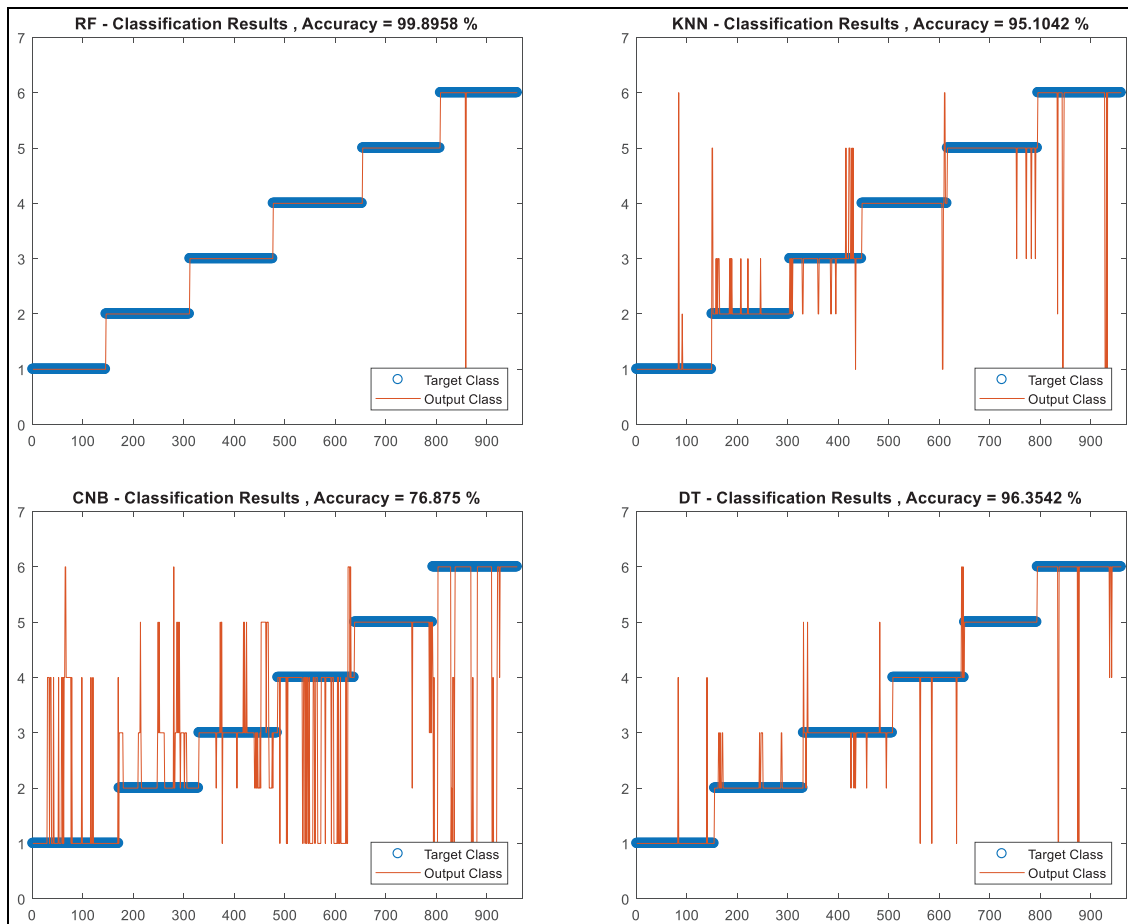


Figure 12. Classification result with 15 features.

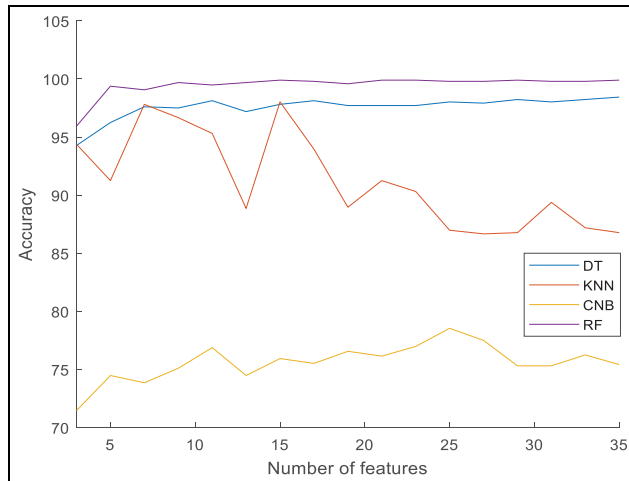


Figure 13. Classification accuracy variation according to the selected number of features.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Chemseddine Rahmoune  <https://orcid.org/0000-0003-0388-7455>

References

1. Yang L and Chen H. Fault diagnosis of gearbox based on RBF-PF and particle swarm optimization wavelet neural network. *Neural Comput Appl* 2019; 31: 4463–4478.
2. Hamadache M, Lee D and Veluvolu KC. Rotor speed-based bearing fault diagnosis (RSB-BFD) under variable speed and constant load. *IEEE Trans Ind Electron* 2015; 62: 6486–6495.
3. Goyal D and Pabla BS. The vibration monitoring methods and signal processing techniques for structural health monitoring: a review. *Arch Comput Methods Eng* 2016; 23: 585–594.
4. Yang Q, Liu T, Wu X, et al. Gear backlash detection and evaluation based on current characteristic extraction and selection. *IEEE Access* 2020; 8: 107161–107176.
5. Wang J, Gao RX and Yan R. Integration of EEMD and ICA for wind turbine gearbox diagnosis. *Wind Energy* 2014; 17: 757–773.
6. Wang L, Xiang J and Liu Y. A time–frequency-based maximum correlated kurtosis deconvolution approach for detecting bearing faults under variable speed conditions. *Meas Sci Technol* 2019; 30: 125005.
7. Han D, Guo X and Shi P. An intelligent fault diagnosis method of variable condition gearbox based on improved DBN combined with WPEE and MPE. *IEEE Access* 2020; 8: 131299–131309.
8. Zheng X, Wang S and Qian Y. Fault feature extraction of wind turbine gearbox under variable speed based on improved adaptive variational mode decomposition. *Proc IMechE, Part A: J Power and Energy* 2020; 234: 848–861.
9. Gougam F, Rahmoune C, Benazzouz D, et al. Bearing fault diagnosis based on feature extraction of empirical wavelet transform (EWT) and fuzzy logic system (FLS) under variable operating conditions. *J Vibroengineering* 2019; 21: 1636–1650.
10. Shi P, Yang W, Sheng M, et al. An enhanced empirical wavelet transform for features extraction from wind turbine condition monitoring signals. *Energies* 2017; 10: 972–985.
11. Li Z, Ma Z, Liu Y, et al. Crack fault detection for a gearbox using discrete wavelet transform and an adaptive resonance theory neural network. *J Mech Eng* 2015; 61: 63–73.
12. Sanz J, Perera R and Huerta C. Gear dynamics monitoring using discrete wavelet transformation and multi-layer perceptron neural networks. *Appl Soft Comput* 2012; 12: 2867–2878.
13. Chen ZS, Rhee SH and Liu GL. Empirical mode decomposition based on Fourier transform and band-pass filter. *Int J Nav Archit Ocean Eng* 2019; 11: 939–951.
14. Yang Y, He Y, Cheng J, et al. A Gear fault diagnosis using Hilbert spectrum based on MODWPT and a comparison with EMD approach. *Measurement* 2009; 42: 542–551.
15. Shan P-W and Li M. Nonlinear time-varying spectral analysis: HHT and MODWPT. *Math Probl Eng* 2010; 2010: 1–14.
16. Huang H, Baddour N and Liang M. Bearing fault diagnosis under unknown time-varying rotational speed conditions via multiple time-frequency curve extraction. *J Sound Vib* 2018; 414: 43–60.
17. Fedala S, Rémond D, Zegadi R, et al. Contribution of angular measurements to intelligent gear faults diagnosis. *J Intell Manuf* 2018; 29: 1115–1131.
18. Boualem M, Chemseddine R, Djamel B, et al. Gear fault feature extraction and classification of singular value decomposition based on Hilbert empirical wavelet transform. *J Vibroengineering* 2018; 20: 1603–1618.
19. Marcano A. Feature selection using sequential forward selection and classification applying artificial metaplasticity neural network. In: *IECON 2010 – 36th Annual Conference on IEEE Industrial Electronics Society*, Glendale, AZ, USA, 7–10 November 2010, New York: IEEE.
20. Song LY, Wang HQ, Pan J, et al. Application of Ant colony algorithm in fault diagnosis of roller bearing. *Adv Mater Res* 2011; 291–294: 1957–1960.
21. Colomi, A., Dorigo, M. and Maniezzo, V. Distributed Optimization by Ant Colonies. In: Varela, F. and Bourgine, P. (Eds.) *Proceedings of the European Conference on Artificial Life, ECAL'91*. Paris: Elsevier Publishing, Amsterdam, 1991, pp. 134–142.

22. Dorigo M, Maniezzo V and Colorni A. Positive feedback as a search strategy. *Technical report* 91-016, Dipartimento di Elettronica, Politecnico di Milano, Italy, 1991.
23. Dorigo M and Blum C. Ant colony optimization theory: a survey. *Theor Comput Sci* 2005; 344: 243–278.
24. Breiman L. Random forests. *Mach Learn* 2001; 45: 261–277.
25. Bonissone P, Cadenas JM, Carmen Garrido M, et al. A fuzzy random forest. *Int J Approx Reason* 2010; 51: 729–747.
26. Pal M. Random forest classifier for remote sensing classification. *Int J Remote Sens* 2005; 26: 217–222.