



A Hybrid Genetic Algorithm and Back-Propagation Classifier for Gearbox Fault Diagnosis

Sunil Tyagi & S. K. Panigrahi

To cite this article: Sunil Tyagi & S. K. Panigrahi (2017) A Hybrid Genetic Algorithm and Back-Propagation Classifier for Gearbox Fault Diagnosis, Applied Artificial Intelligence, 31:7-8, 593-612, DOI: [10.1080/08839514.2017.1413066](https://doi.org/10.1080/08839514.2017.1413066)

To link to this article: <https://doi.org/10.1080/08839514.2017.1413066>



Published online: 12 Dec 2017.



Submit your article to this journal 



Article views: 580



View related articles 



View Crossmark data 



Citing articles: 2 View citing articles 



A Hybrid Genetic Algorithm and Back-Propagation Classifier for Gearbox Fault Diagnosis

Sunil Tyagi  and S. K. Panigrahi 

Department of Mechanical Engineering, Defense Institute of Advance Technology, Girinagar, Pune, India

ABSTRACT

An Artificial Neural Network (ANN) classifier trained by a hybrid GA-BP method for diagnosis of gear faults is presented here that can be incorporated in an online fault diagnostic system of vital gearboxes. The distinctive features obtained from vibration signals of a running gearbox; that was operated in normal and with faults induced conditions were used to feed the GA-BP hybrid classifier. Time domain vibration signals were divided in 40 segments. From each segment features such as magnitude of peaks in time domain and spectrum along with statistical features such as central moments and standard deviations were extracted to feed the classifier. Based on the experimental results it was shown that the GA-BP hybrid classifier can successfully identify gear condition. It was also shown that the network trained by GA-BP hybrid method performs much better than ANN that is trained by standard BP or GA individually. Further, it was also shown that if prior to extraction of features; the vibration signals are pre-processed by Discrete Wavelet Transform (DWT) then efficacy of the GA-BP hybrid is significantly enhanced.

Introduction

Gears are used in machines to pass on power from one shaft to another with change in its speed and torque. In plant industry Gearboxes are perhaps the most critical machinery fitted in a power transmission system (Radzevich 2012). Due to its high cost, gearboxes are not readily kept in stock for replacement in event of a catastrophic failure. Therefore gearbox failure can potentially lay off the plant for months, due to long lead time of production of a replacement gearbox. The implementation of an online supervisory condition monitoring system for critical gearboxes is an undeniable obligation due to the ever increasing requirement of reliability coupled with decreasing manpower for watch keeping. The online supervisory condition monitoring system should be able to detect fault at the incipient stage and give necessary alarm and/or trip signal. To achieve this it is necessary to

CONTACT S. K. Panigrahi  panigrahisk@diat.ac.in  Department of Mechanical Engineering, Defense Institute of Advance Technology, Girinagar, Pune 411025, India.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/UAAI.

© 2017 Taylor & Francis

hybridize multiple cotemporary techniques/algorithms so as to harness their good points and overcome their shortcomings.

Alteration in vibration signals emanating from machine indicate that the gear pair meshing condition is undergoing a change. Monitoring changes in vibration signal with the help of accelerometer mounted on the gearbox housing is an established method for gear damage assessment. The dominant source of vibration in gears is the interaction of gear teeth. Even when there are no faults present, the dynamic forces that are generated produce both impulsive and broadband vibrations. The discrete, impulsive vibration is associated with the various meshing impact processes, and the broadband vibration is associated with friction, fluid flow, and general gear system structural vibration (Norton and Karczub 2003).

Gear faults generally fall into following two categories:

- (i) Localised faults-discrete gear tooth irregularities such as chipped tooth or missing tooth Gear.
- (ii) Distributed faults-uniform wear on teeth all around the gear.

The main frequency at which gearing induced vibrations will be generated is the gear meshing frequency, f_m . It is given by:

$$f_m = N \times f_s \quad (1)$$

where N is the number of teeth, and the f_s is the rotational speed of the gear shaft. Dominant peak at f_m can be observed in the spectrum and due to the periodic nature of gear meshing, integer harmonics of f_m are also present in spectrum. Increase in vibration levels at the gear meshing frequency and its associated harmonics are typical criteria for fault detection.

The localized fault manifest itself in high vibration levels at the gear mesh frequency, f_m , and its associated harmonics. Also, discrete faults tend to produce low level, flat, sideband spectra (at $\pm$ the shaft rotational speed and its associated harmonics) around the various gear meshing frequency harmonics. Distributed faults such as uniform wear around a whole gear tend to produce high level sidebands (at $\pm$ the shaft rotational speed and its associated harmonics) in narrow groups around the gear meshing frequencies (Norton and Karczub 2003).

The increases at the gear meshing frequency and its various harmonics are associated with fault/wear. However, in a typical gear train several such gear meshing frequencies and their harmonics are present, which coupled with high broadband random noise that is always present in actual working gearbox; makes detection of gear faults a tough task.

To solve this problem, techniques using both time domain and spectral methods have been developed for gear fault diagnosis (Kang, Wang, and Chang 2007; Sharma and Parey 2016). Time domain methods make use of

indicators that are able to capture ‘peakyness’ of signal such as kurtosis, difference histograms, crest factor and peak level/rms value. It is hard to recognize gear defect in simple vibration spectrum or by traditional time domain methods, hence many new techniques have been employed by researchers over the years for diagnosis of gear faults. These include techniques such as Artificial Neural Networks (ANN) (Bangalore and Tjernberg 2015; Dellomo 1999), Wavelet Transform (Tse 2004; Yao et al. 2009; Zhou et al. 2007) and Genetic Algorithm (GA) (Hajnayeb et al. 2011). To make the classification process faster and more effective various researchers have demonstrated the efficacy of hybridizing ANN with Genetic Algorithm (GA) for fault diagnosis of rotary machines (Kamp and Savenije 2006; Samanta 2004).

This article presents a report on effectiveness of GA-BP trained ANN classifier for diagnosis of gear faults using features obtained from time-domain and frequency spectrum of vibration signal. In this article method of acquiring vibration signals and creation of features to feed the classifier was similar to the method employed by Tyagi and Panigrahi (2017). However, the similarity lies only till data acquisition and creation of test and train vectors.

In this article ANN is used as a classifier. The ANN classifier used here was trained by an ingenious hybrid technique where the weights and biases of ANN were optimized initially by Genetic Algorithm followed by back-propagation (BP). The magnitude of the vibration data may change with varying operating conditions such as operating speed, load or location/sensitivity of sensor. Making the signals comparable regardless of differences in the magnitude is achieved by normalizing. In present work normalized signals were used for better generalization and to ensure that the classification results are not influenced by changes in signals due to operating conditions.

The hybrid classifier was fed with features obtained from time-domain and frequency spectrum of vibration signal. The vibration (acceleration) signals were obtained from gearbox that was operated in normal (without fault) condition and with induced faults. These signals were treated with simple processing to extract features that were thereafter used to feed the ANN classifier.

The effects of number of generations in GA loop, comparison of GA-BP hybrid with pure GA and pure BP as well as effect of pre-processing with DWT were studied and their results are presented subsequent paragraphs.

Discrete wavelet transform and multi resolution analysis

Wavelet Transform has been successfully used for diagnosis of machinery faults. Peter et al. (2004) demonstrated how Wavelet Transform is an

effective tool for machinery fault diagnosis. In recent times DWT has been applied for gearbox fault diagnosis (Rafiee et al. 2007; Yao et al. 2009; Zhou et al. 2007). Various researchers have enhanced the effectiveness of traditional classifiers in diagnosis of machinery fault by employing DWT along with the classifier (Paya, Esat, and Badi 1997).

Due to non-stationary nature of machinery vibration signal the Wavelet transforms are most suitable to study them. Further, DWT also provides multi resolution analysis which makes it is possible to obtain good time resolution at high frequency band and good frequency resolution at low frequencies (Chui, 1992). This property makes it most appropriate for analysis of vibration signal which is inherently non-stationary and contains both high and low frequency components.

The Daubechies wavelet of order 44 (db44) has been found to be especially useful in machinery fault diagnosis applications as reported by Rafiee et al. (2009). Hence db44 has been used in this article to carry out the DWT pre-processing of the signals.

Wavelet transform is a mathematical function that is used to divide continuous time signals or functions into different scale (frequency components), and there after analyse each section with resolution matched to its scale.

Let $x(t)$ be a discrete signal. The Continuous Wavelet Transform (CWT) for this discrete time signal can be defined as:

$$W_{\psi}(\tau, s) = \int_{-\infty}^{+\infty} x(t) \cdot \psi_{\tau, s}^*(t) dt \quad (2)$$

where $\psi_{\tau, s}^*$ is the conjugate of $\psi_{\tau, s}(t)$ which is the shifted and scaled form of the transforming function, named as “mother wavelet”, which is defined as:

$$\psi_{\tau, s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t - \tau}{s}\right) \quad (3)$$

The transformed signal obtained in Equation (2) above is a function of s and τ , which are the scale and translation parameters respectively. The DWT obtained by discretization of $W_{\psi}(\tau, s)$ is given by:

$$\text{DWT}(j, k) = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{+\infty} x(t) \psi\left(\frac{1 - 2^j k}{2^k}\right) dt \quad (4)$$

An algorithm to implement the above scheme was proposed by Mallat (1989). Figure 1 illustrates the wavelet decomposition transformation algorithm as a flow diagram. DWT is carried out by sequential decomposition of the signal $\mathbf{x}$ ($\mathbf{t}$) as shown in Figure 2. At each level the signal $\mathbf{x}(\mathbf{t})$ is convolved with the high pass filter $\mathbf{H}$ and the low pass filter $\mathbf{L}$, producing two decomposed vectors $\mathbf{D}$

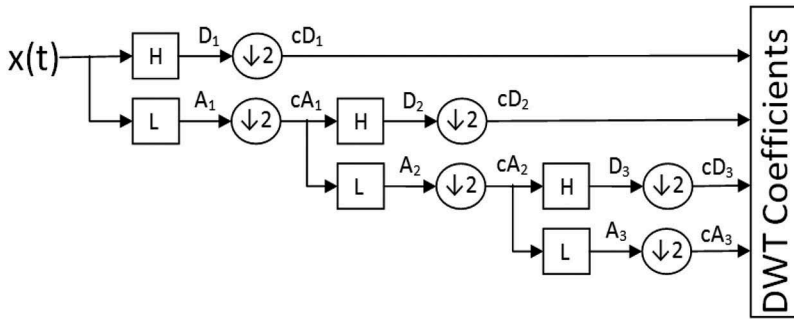


Figure 1. Wavelet Transform decomposition at different levels.

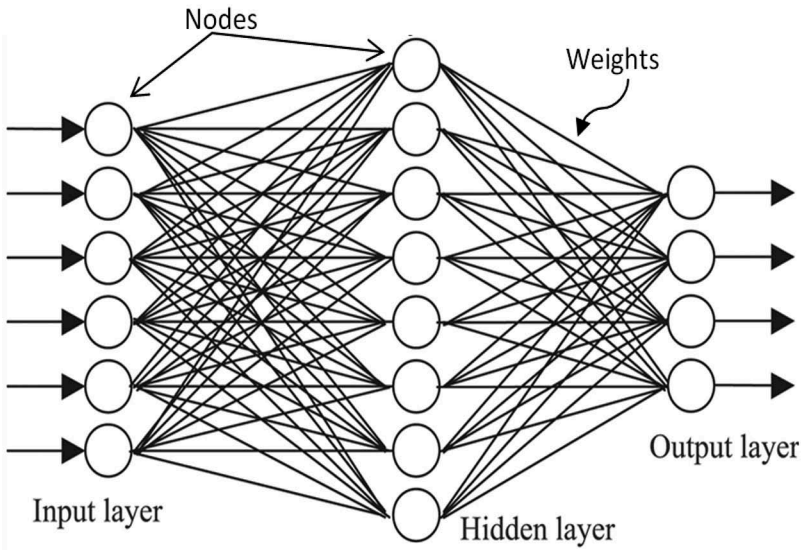


Figure 2. Artificial neural network.

and **A** respectively. The vectors **D** and **A** are thereafter down sampled by half to obtain **cD** and **cA** called detail and approximate coefficients respectively. Down sampling is performed to ensure that the total numbers of coefficients produced in all of the decomposed constituents are equal to number of samples in the input discrete signal **x(t)**. The Multi Resolution Analysis is achieved by repeating the decomposition process with approximate coefficients **Ca** at each level to obtain DWT coefficients at next level. The decomposition is performed till the desired resolution has been attained.

Artificial neural networks

Artificial Neural Networks (ANN) are defined as a highly connected array of elementary processors called neurons. ANN takes inspiration from

biological learning process of human brain (Jain, Mao, and Mohiuddin 1996). ANN have been widely used in recent years for practical applications such as pattern recognition, classification, function approximation, etc. (Rabuñal and Dorado 2006). ANN have also been used successfully for machinery fault detection (Paya, Esat, and Badi 1997; Tyagi and Panigrahi 2017) and recently researchers have employed ANN for gearbox fault diagnosis (Bangalore and Tjernberg 2015; Dellomo 1999; Saravanan and Ramachandran 2010).

A typical ANN consists of one input layer, one output layer and one or more hidden layers. Figure 2 schematically illustrates the structure of neuron. Each node shown in Figure 2 represents a neuron. There are several neurons in each layers and neuron in layers are connected to the neurons in the adjacent layer with different connection weights. The neurons of input layer are fed with input feature vectors. Each neuron of hidden and output layers receives signals from the neurons of the previous layer multiplied by weights of the interconnection between neurons. The neuron there after produces output by passing the summed signal through a transfer function.

Multilayer neural networks are traditionally trained by the supervised learning methods. In supervised learning network is provided with set of input vector and the targeted output that is desired from the network. Let for example the inputs and desired target to a neural network are:

$$\{p_1, t_1\}, \{p_2, t_2\}, \dots, \{p_Q, t_Q\}$$

where p_Q is an input vector to network and t_Q is its corresponding target vector that is desired from the ANN (Park et al. 1991).

The network is trained iteratively. In each iteration Mean Square Error (MSE) between target and network output is calculated. The MSE function $F(x)$ at the k th iteration is given by:

$$F(x) = (t_k - a_k)^2 \quad (5)$$

where t_k and a_k are the targets and output vectors at the k^{th} iteration.

The training of network is achieved by adjusting the weights and biases so that the MSE function $F(x)$ is minimized. Back-Propagation (BP) using Levenberg-Marquardt algorithm (LM) is commonly employed to train ANN (Park et al. 1991). BP is a gradient descent method that uses the calculated MSE at each layer to adjust the value of the interconnecting weights to the neuron to achieve minimum MSE. This process of layer wise weight adjustment is repeated until the minimum of the error function is achieved or the MSE has reached to such a low value, which would be able to classify the input vectors correctly. The weights at m^{th} layer in $k + 1$ iteration are estimated by:

$$\mathbf{w}_{i,j}^m = \mathbf{w}_{i,j}^m(\mathbf{k}) - \alpha \frac{\partial F(\mathbf{x})}{\partial \mathbf{w}_{i,j}^m} \quad (6)$$

Being a gradient descent process BP is a fast and efficient algorithm that achieves convergence very quickly. However, BP like any other gradient search techniques gives inconsistent and unpredictable performance when applied to complex nonlinear optimization problems such as ANN (Curry and Morgan 1997). Due to the intricate nature of training ANN, the error surfaces are very complex. Since the nature of BP is to converge locally, the solutions are highly dependent upon the initial random draw of weights. Due to which BP algorithm is likely get trapped in a local solution that may or may not be the global solution. This local convergence and inability to come out of local minima could present serious problems when using ANN for practical applications.

GA-ANN hybrid algorithm

Genetic Algorithms (GA) is an optimization method that was introduced by Holland (1992), it imitates the process followed in natural evolution. Laws of natural selection and genetics are followed by GA search process. Genetic Algorithm consists of three processes namely; selection, genetic operation, and replacement. The population is set of chromosomes each one of which can provide the solution. To check survival of candidate chromosomes the fitness values of complete population (chromosomes) are evaluated. Fitness value is obtained by means of an objective function which is the performance criteria (phenotype). A particular set of parents are selected from the population (by a preselected method) to generate offspring. The next generation is produced by observing genetic operations of crossover and mutation and the chromosomes in the current population are then replaced by the offspring, based on chosen replacement strategy. Such a GA cycle is iteratively repeated until the desired termination criterion is achieved. At the end of the process of this simulated evolution, the best chromosome (individual) in the final population will be highly evolved to give best possible solution to the given problem.

To overcome the local convergence problem of ANN trained by BP, a number of different optimization techniques have been developed these include global search techniques like evolutionary programming, simulated annealing (Salchenberger, Cinar, and Lash 1992). Many researchers have applied GA to train the ANN by updating weights and biases to get optimized solution (Bi et al. 2000; Dorsey and Mayer 1994).

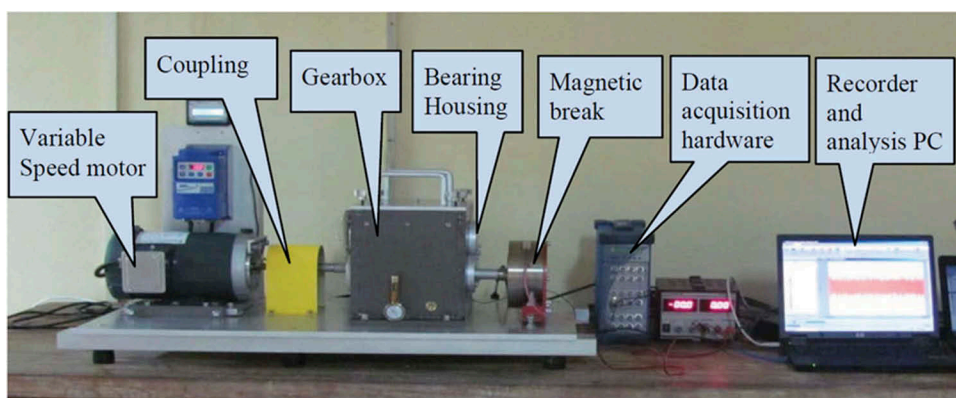


Figure 3. Experimental setup.

Experimental setup

The setup that was used to carry out experiments is shown in [Figure 3](#) was used to carry out the experiments. The setup was similar to the one used by Tyagi and Panigrahi (2017). The experimental setup comprised of a variable speed AC motor driving a gearbox shaft through flexible couplers. The gearbox consists of a two stage parallel shaft gearbox with four spur gears having 32, 80, 48, and 64 teeth. A magnetic break driven by belt was used to create torsional load on the system.

In present work the gear with 32 teeth was analyzed. It was mounted on load end side of the shaft to enable easy replacement. Vibration signals in form of acceleration were acquired by the accelerometer that was mounted with help of stud on the housing of bearings that supported defective gear shaft. Three different gear faults such as missing tooth, chipped tooth and surface wear were artificially created. Four vibration signals (acceleration) were collected for different conditions of gearbox such as gear in good condition and gearbox with defective gear having missing tooth, chipped tooth and surface wear defects respectively. During data recording the torsional load on the gearbox was set to 164 N-m and shaft was rotated at 20 Hz speed and the vibration signals were collected for the duration of two seconds each, at a sampling rate of 51.2 KSa/s, thus each acquired signal had 102400 samples. Following four different vibration signals were collected:

- (i) Gear in normal condition.
- (ii) Gear with Missing Tooth Fault (MTF).
- (iii) Gear with Chipped Tooth Fault (CTF).
- (iv) Gear with Surface Wear Fault (SWF).

Feature extraction and training/test vectors creation for GA+BP hybrid classifier

Feature extraction

The magnitude of the vibration data may vary with different kind of faults under varying operating conditions. Making the signals comparable regardless of differences in the magnitude is achieved by normalizing by standardization. The signal is normalized by using its mean and standard deviation as per Equation (7).

$$y_i = \frac{(S_i - \mu)}{(\sigma)} \quad (7)$$

Where S_i denotes samples, μ is mean and σ is standard deviation of signal. The vibration signals (i) to (iv); each having 102400 samples were thereafter divided in 40 non-overlapping bins of 2560 samples (y_i). Simple statistical and spectral features were extracted from these 40 bins in the similar fashion as reported by Tyagi and Panigrahi (2017). Apart from ten statistical features that were used in (Tyagi and Panigrahi 2017); two additional features that is amplitude of both side bands to the highest peak in spectrum were also extracted in present research. Details of twelve features extracted are given in Table 1.

The features extracted from the acquired vibration signals of defective gears and gears in good condition showed good numerical dissimilarity with each other and therefore an ANN classifier that is trained by these features they is likely to be effective.

Creation of training and test vectors

All four types of signals (as explained in Para 5) were divided in 40 bins each. Out of these 40 bins 24 bins were used as training bins and remaining 16 were used as test bins. The classifier was exposed only to the training bins during training of classifier. 16 test bins that were not exposed to classifier during training were used to test the performance of the trained classifier. The training feature vectors were created by extracting features from bins of

Table 1. Description of feature numbers and their parameter.

Feature number	Parameter
1 to 5	Amplitude of five highest peaks of the bin.
6	Amplitude of highest peak in spectrum.
7 & 8	Amplitude of side bands to the highest peak in spectrum.
9	Standard deviation σ of the bin.
10	Skewness γ_3 (third central moment) of the bin.
11	Kurtosis γ_4 (fourth central moment) of the bin.
12	Sixth central moment γ_6 of the bin.

defective gears and bins of normal gear and placed them alternatively. Therefore three sets of 48 training feature vectors denoted as MTF, CTF and SWF were created for gear with missing tooth fault, chipped tooth fault and surface wear fault respectively. In similar manner three sets of 32 test vectors were also formed. Since 12 features were used, therefore a training matrix of size 12×144 and test matrix of 12×96 sizes was created.

Creation of GA-BP hybrid trained classifier

In this article ANN with configuration of 12:15:1 is used as a classifier (details are given in Para 7.1). The weights and biases of ANN classifier are updated by an ingenious GA–BP hybrid technique. The network was trained by four steps as given below:

The framework of the GA–BP hybrid training of the neural network classifier for gear fault diagnosis is shown in Figure 4. The GA loop shown in Figure 4 starts with randomly selected weights and biases. Step 1 and step 2 of hybrid training process are performed in GA loop. The step 3 and 4 of hybrid training is performed in BP loop shown in Figure 4.

ANN classifier

A simple Artificial Neural Network with one hidden layer was created. The number of neurons in input layer was 12 (equal to length of feature vectors). Hidden layer nodes were varied from 10 to 30 and the output layer nodes varied between one and two. It was found that best performance was obtained with ANN configuration of 12:15:1. The output layer was made with only one neuron and in training stage it was set to {1} for good gear and {0} for defective gear. The network was first trained by GA as explained in Para 7.2 followed by training BP that is explained in Para 7.3.

Step	Procedure
1st step	• Weights and biases of ANN are optimized by Genetic Algorithm. • The GA optimizes weights and biases so as to minimize the error function $F(x)$ (refer Equation 5). • Details of GA parameters used are given in Para 7.2
2nd step	Updating of weights and biases of ANN is stopped as per predetermined number of generations of GA.
3rd step	• The weights and biases attained at the end of step 2 are saved • Saved weights & biases are provided as initial start point for updating by BP in step 4.
4th step	• Resume updating of weights and biases to minimise the error function $F(x)$ by using Back Propagation. • The details of algorithm, learning rate and stopping criteria used is given in Para 7.4.

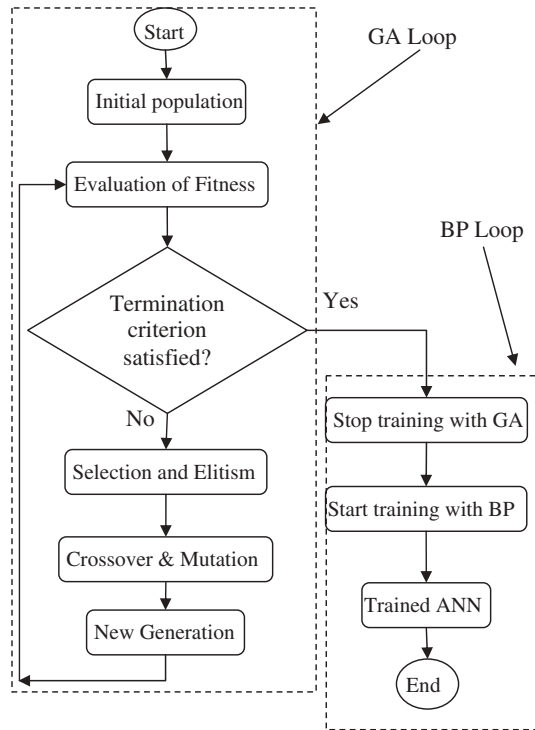


Figure 4. Framework of GA – BP hybrid training of ANN classifier.

A rather stiff criterion was used to classify the output of the GA-ANN classifier during simulation. The input vector from defective was considered to be correctly classified only if the output from output layer lies within $[0 \pm 0.1]$ similarly the input vector from gear in good condition was considered to be correctly classified if output lies within $[1 \pm 0.1]$ range.

GA loop

The use of GA needs consideration of six basic issues: chromosome (genome) representation, selection function, genetic operators like mutation and crossover for reproduction function, creation of initial population, termination criteria, and the evaluation (fitness) function. The fitness function of GA was the MSE of the network. The total number of weights and biases were 211 as neurons of ANN were arranged in 12:15:1 configuration. A population size of 450 individuals was used starting with randomly generated genomes. This size of population was chosen to ensure relatively high interchange among different genomes within the population and to reduce the likelihood of convergence within the population. The GA used single point crossover with crossover and mutation fractions of 0.8 and 0.2 respectively. In addition,

the selection scheme of GA loop taken was the roulette method. The termination criterion for total number of generation was varied from 10 to 10,000.

Effects of number of generations in GA loop

The termination criteria of GA loop was varied from 10 to 10,000 generations to check the effect of number of generations on identification of gear condition. The results are presented in Figures 5 and 6.

Post training the ANN with GA, the network was simulated with training vectors. Total MSE post GA loop was computed for all training vectors as given below:

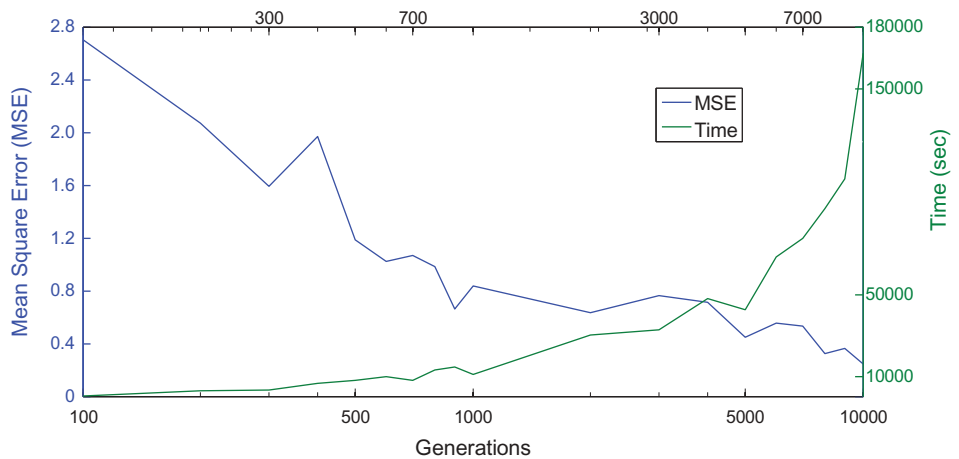


Figure 5. Variation of MSE and processing time in GA loop.

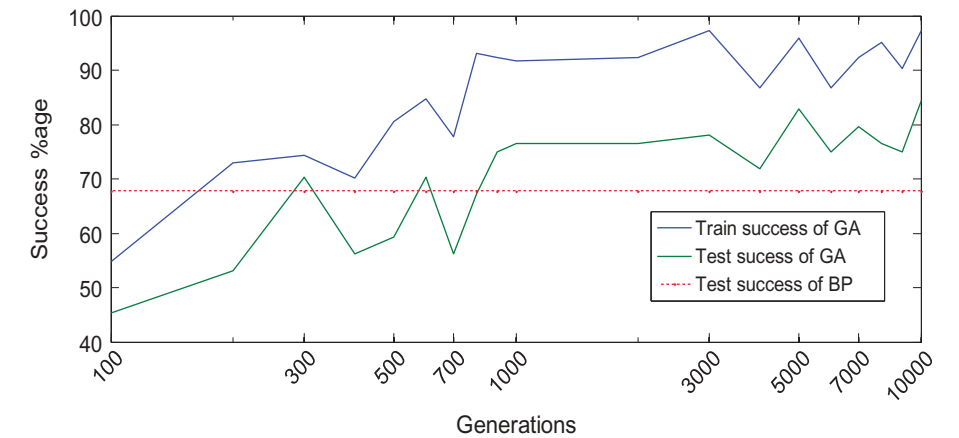


Figure 6. Success achieved in GA loop.

$$MSE = \sum_{k=1}^{144} (t_k - a_k)^2 \quad (9)$$

Where 144 are the total number of training vectors, t_k and a_k are the target and output of the network when it is fed with k th training vector. Figure 5 plots the variation of MSE with that of number of generations. Figure 5 helps to gauge how the classification efficiency of network varies with number of generations in GA loop. The total MSE reduces when number of generations is increased indicating that as GA is allowed to run for more number of generations the network is able to converge closer to the optimum solution. The MSE reduced significantly initially, but reduction in MSE decreased as generations were increased beyond 300.

Further, Figure 5 also presents time taken (in seconds) by GA to produce number of generations. It is evident that as the number of generation is increased the process time taken by GA also increased significantly. The processing time taken by GA rose quite significantly after 300 generations. These results also proves that GA is a slow algorithm (in agreement with Louis et al. (1993)) as it took about 46.5 hrs to reach convergence in case of 10,000 generations.

To check the efficacy of the GA trained network; the network was fed with training and test vectors and the results that were obtained are presented in Figure 6. The success percentage obtained when. It was found; that the network was able to correctly classify with high success rate when it was fed with the training vectors. The success rate of correctly classified test vectors was about 10% lower than training vectors. The performance of network improved with increase in number of generations as success percentage in both test and train vectors increased with increase in number of generations. The performance of network improved rapidly with generations till 900 generations. Although, there was increase in success percentage when number of generations were increased from 900 generations to 10,000 generations, but the increase was not as rapid and significant as that obtained when generations were increased from 100 to 900.

The network was also trained just by BP alone using Levenberg-Marquardt algorithm and the test success percentage obtained was 68.2%. Although, the success %age of network that is trained by BP alone do not depend on number of GA generations but it is also shown plotted (in red dotted line) in Figure 6 for the sake of comparison only. It can be seen from Figure 6 that performance of ANN that was trained with GA for number of generations 800 and above performed better than the performance of ANN trained by BP.

Training	• Using Back propagation with Levenberg-Marquardt algorithm. • Learning rate = 10^{-7}
Stopping criteria	• Number of iteration < 15,000, • Gradient > 10^{-6} • Mean Square error $\leq 10^{-8}$

Training the ANN with back-propagation post GA

Post training the Artificial Neural Network with GA as explained in Para 7.2 and 7.3 the network's weights and biases were saved. The ANN was further trained by BP using the saved weights and biases. Details of training by BP are given as follows:

The results of success obtained post training the ANN with GA followed by BP are presented in Figure 7. If one compares the result of GA-BP trained ANN with the results obtained by pure GA trained network that is given in Figure 6 it can be seen that training of network by BP post GA has significantly improved the performance of the network. The success percentage for all generations was more than 94% and 88% for train vectors and test vectors respectively. Whereas GA trained network could manage test success in range 75% to 85% (refer Figure 6). Since pure BP trained ANN could manage test success of 68.2% only hence the performance of hybrid GA-BP ANN is better than pure BP.

Although the training of network by adding BP module post GA has greatly improved the performance but the addition of BP hasn't caused any significant increase in processing time. There was only 85 seconds (on average) increase in processing time due to addition of BP module (refer Table 3) which is insignificant since GA loop took 1.67×10^5 seconds to process 10,000 generations. Negligible increase in total processing time on

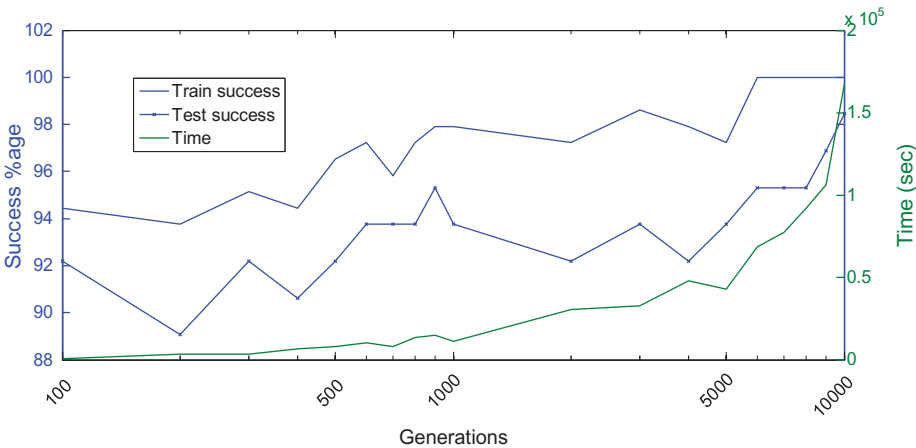


Figure 7. Success of GA-BP hybrid trained network.

account of performing BP post GA loop was expected as BP is a gradient descent method which is an efficient and a fast process.

Pre-processing by discrete wavelet transform

Results obtained post pre-processing the acquired vibration signals by DWT is presented here. Raw vibration signals obtained from Gear in good condition and defective gears were pre-processed with DWT using db44 (Daubechies wavelet of order 44). The db44 wavelet was chosen because the shape of db44 wavelet is most similar to the vibration produced by defective gear as reported by Rafiee et al. (2009). The wavelet packets were obtained after performing DWT decomposition of signal till level 5. Wavelet packets of the low frequency approximates at level 1–5 (A1–A5) and the high frequency detail signals at level 1–5 (D1–D5) were obtained after doing the pre-processing (decomposition). Features were extracted from these approximates and details as was done earlier for raw vibration signal and the test/training vectors were created with features as was done for raw signal.

The GA-BP hybrid network was trained from the training vectors obtained from details (D1–D5) and approximates (A1–A5). Table 2 presents frequency range of details and approximates. The D1 details had highest frequency components the associated frequency components decreased progressively from D1 to D5. Approximates were comprised of low frequency components. Figure 8 shows the cumulative (all 19 cases) success obtained by the classifier when input signal was pre-processed by DWT for each type of approximate and detail along with raw signal (not pre-processed by DWT). Inspection of Figure 8 reveals that when GA-BP hybrid network was trained by D1, D2, and D3 it resulted in high training and test success. This indicates that gear defects signals have distinctive influence in high frequency range >4 kHz as the associated frequency range of the D1, D2, and D3 were between 4 and 24 KHz (refer Table 2). The test, and training performance of D1, D2, and D3 were better than the raw signal by almost 4%. The performance of details at level 4 and 5 (D4, D5) and approximates at levels 1–5 (A1–A5) which had frequency spread predominantly in low frequency range (10 Hz–5 KHz) were not good as good as obtained by D1, D2, and D3. Best results were obtained

Table 2. Frequency range of Approximates and details at various levels of DWT.

Details/approximate	Range	
	Min (KHz)	Max(KHz)
D1	10	24
D2	5	11
D3	3	7
D4	1	4
D5	0.7	1.5
A1 to A5	0	4

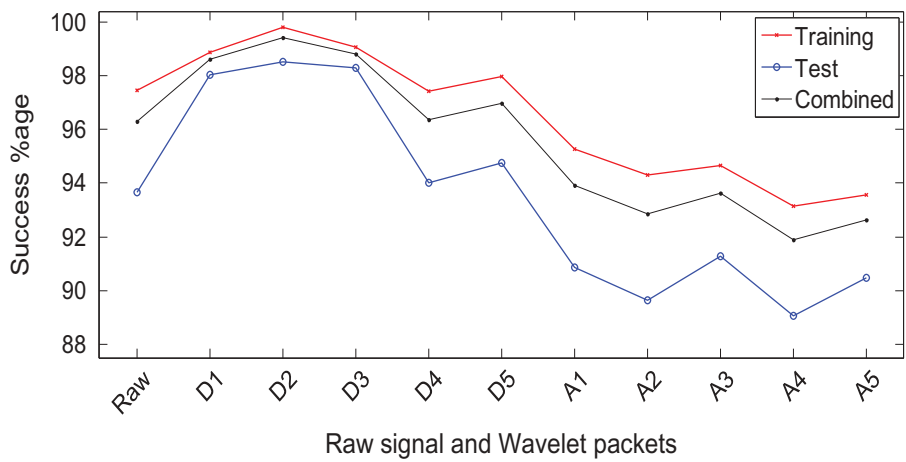


Figure 8. Success obtained by raw signal and DWT packets.

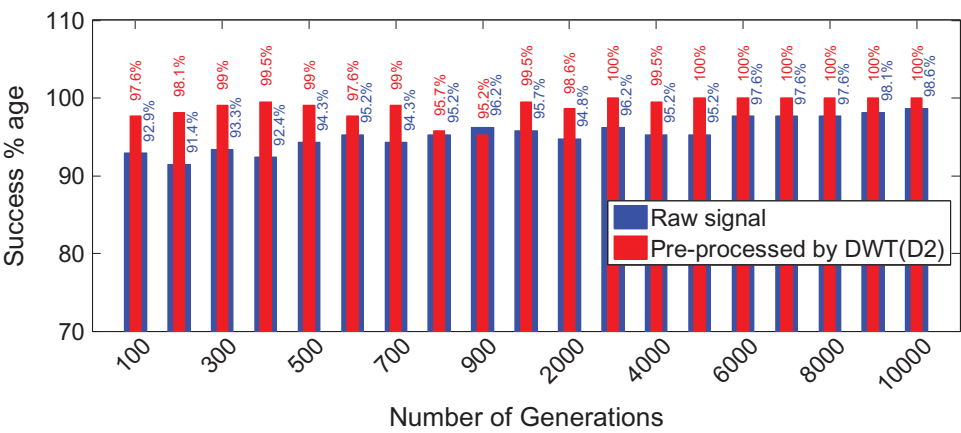


Figure 9. Effect of pre-processing by DWT.

when network was trained with features extracted from D2. With D2 the total success for all 19 cases achieved were very high. Hybrid GA-BP training of network by features obtained from D2 achieved training success of 99.7% and test successes of 98.5%.

Figure 9 presents the combined (training + test) success achieved by the GA-ANN Hybrid classifier, when it was fed with features extracted from raw signal and from D2 (i.e. level 2 details of signal pre-processed with DWT). Figure 9 presents the performance of the classifier by using different generations in GA loop. It is evident from Figure 9 that pre-processing with DWT significantly improves the performance of the GA-BP hybrid trained classifier as combined success has increased in 18 out of 19 cases. The maximum increase (about 7.1%) was in case where 400 generations were used in GA loop, whereas in one case

Table 3. Training time taken by GA and BP for raw and D2 signal.

GA No. of generations	Time taken by GA (Sec)		Time taken (Sec) by BP post GA (D2)
	Raw signal	Pre-processed by DWT (D2)	
100	850	1110	152.8
200	3210	3240	152.3
300	3570	3150	144.1
400	6860	6790	131.5
500	8220	5600	128.1
600	10250	10440	114.9
700	8200	7990	115.5
800	13450	13550	103.7
900	14900	14770	99.3
1000	11280	10570	86.2
2000	30250	21990	82.5
3000	32900	44330	69.5
4000	48100	60180	62.3
5000	42730	42330	53.3
6000	68450	90230	43.4
7000	77340	67020	33.8
8000	91780	91050	27.2
9000	106230	104970	17.1
10000	167600	91870	13.1

where 900 generations were used showed a decrease. It is noticeable that seven cases (when numbers of generations were 4000 and more) achieved 100% success percentage. Apart from increasing the test performance the pre-processing the vibration signal with DWT has significantly reduced the training time taken in GA loop. Table 3 presents the time taken (in seconds) for training by GA loop when it was trained by features extracted from raw signal and with features obtained from signal that was pre-processed by DWT. When the classifier was trained with features obtained from D2 (i.e. signal pre-processed with DWT details at level 2); the reduction in processing time took place in 12 out of 19 cases. The Table 3 also presents the time taken (in seconds) by the BP to train post completion of GA loop using features obtained from DWT pre-processed signal (D2). It is evident that the time taken by BP is very little (average 85 seconds) compared to GA (average 3925 seconds). There was significant improvement in performance of the classifier without much increase in process time when the input signal data was pre-processed with DWT with Debauches 4 Wavelet and using the details at level 2.

Conclusion

A method to identify gear condition by using a hybrid GA-BP training technique to train the ANN classifier is presented here. Simple features created from magnitude of highest peaks in time domain, peaks in frequency spectrum and statistical central moments of time domain vibration signal. It was shown that hybrid classifier is able to correctly diagnose

gearbox condition with high accuracy when trained with these simple features.

Vibration data from gear in good conditions and with defects induced were obtained and features extracted from vibration signals were fed to the ANN classifier that was trained first by GA to certain predetermined number of generations and thereafter the network was trained by BP. It was found that performance of the classifier to classify training as well as test vectors improves as numbers of generations in GA are increased. Although, performance improves with increase in number of GA generations but at the same time the time taken also increases significantly when number of generations is increased beyond 1000. The training of the classifier by BP post training it with GA is able to improve the performance without any corresponding increase in process time.

It was shown that the performance of the hybrid GA-BP trained ANN is better than the ANN that was trained by BP alone or by GA alone. Further, it has also been shown that if the vibration signal is pre-processed with DWT by using Debauches 4 Wavelet details at level 2 that is D2, it not only improves the performance of classifier, but it also reduces the process time taken in training by the GA-BP hybrid classifier.

Nomenclature

a_k	Output vectors at the kth iteration
$DWT(j, k)$	Discrete wavelet transform coefficients
f_m	Gear meshing frequency
f_s	Rotational speed of the gear shaft per second
$F(x)$	Mean square error function
N	Number of teeth on Gear
p_Q	Input vector to network
s	Scale parameter
S_i	ith sample of signal
t_k	Targets vectors at kth iteration
t_Q	Target vector
$W_\psi(\tau, s)$	Continuous wavelet transform coefficient
$w_{i,j}^m$	Weight of ith neuron at mth layer in jth iteration
$x(t)$	Acquired vibration signal
y_i	ith sample of normalised signal
$\Psi_{\tau,s}(t)$	Mother wavelet
σ	Standard deviation of bin
τ	Translation parameter

Declaration of interest

The authors declare that there is no conflict of interest regarding the publication of this article.

ORCID

Sunil Tyagi  <http://orcid.org/0000-0001-5897-7955>

S. K. Panigrahi  <http://orcid.org/0000-0001-6624-0988>

References

- Bangalore, P., and L. B. Tjernberg. 2015. An artificial neural network approach for early fault detection of gearbox bearings. *IEEE Transactions on Smart Grid* 6 (2):980–87.
- Bi, T. S., Y. X. Ni, C. M. Shen, and F. F. Wu. 2000. A novel ANN fault diagnosis system for power systems using dual GA loops in ANN training. In *Power Engineering Society Summer Meeting*, 2000. IEEE, Vol. 1, pp. 425–30.
- Chui, C. K. 1992. Wavelets: A tutorial in theory and applications. In: *Wavelet Analysis and its Applications*, edited by C. K. Chui, 1. San Diego, CA: Academic Press.
- Curry, B., and P. Morgan. 1997, Feb 28. Neural networks: A need for caution. *Omega* 25 (1):123–33.
- Dellomo, M. R. 1999, Jul 1. Helicopter gearbox fault detection: A neural network based approach. *Journal of Vibration and Acoustics* 121 (3):265–72.
- Dorsey, R. E., and W. J. Mayer. 1994. Optimization using genetic algorithms. *Advances in Artificial Intelligence in Economics, Finance, and Management* 1:69–91.
- Hajnayeb, A., A. Ghasemlooia, S. E. Khadem, and M. H. Moradi. 2011, Aug 31. Application and comparison of an ANN-based feature selection method and the genetic algorithm in gearbox fault diagnosis. *Expert Systems with Applications* 38 (8):10205–09.
- Jain, A. K., J. Mao, and K. M. Mohiuddin. 1996, Mar 1. Artificial neural networks: A tutorial. *IEEE Computer* 29 (3):31–44.
- Holland, J. H. 1992. *Adaptation in Natural and Artificial Systems An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. 2nd ed., p. 232. Cambridge, MA: MIT Press.
- Kamp, R. G., and H. H. Savenije. 2006, Sep 7. Optimising training data for ANNs with genetic algorithms. *Hydrology and Earth System Sciences Discussions* 10 (4):603–08.
- Kang, Y., C. C. Wang, and Y. P. 2007. Chang Gear fault diagnosis in time domains by using Bayesian networks. In: *Analysis and Design of Intelligent Systems using Soft Computing Techniques*. Springer Berlin Heidelberg, pp. 618–27.
- Louis, S. J., and G. J. Rawlins. 1993, Jan 2. Predicting convergence time for genetic algorithms. *Foundations of Genetic Algorithms* 141–61.
- Mallat, S. G. 1989. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11 (7):674–93.
- Norton, M. P., and D. G. Karczub. 2003, Oct 16. *Fundamentals of noise and vibration analysis for engineers*. New York: Cambridge University Press.
- Park, D. C., M. A. El-Sharkawi, R. J. Marks, L. E. Atlas, and M. J. Damborg. 1991, May. Electric load forecasting using an artificial neural network. *IEEE Transactions on Power Systems* 6 (2):442–49.
- Paya, B. A., I. I. Esat, and M. N. Badi. 1997, Sep 30. Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor. *Mechanical Systems and Signal Processing* 11 (5):751–65.
- Peter, W. T., W. X. Yang, and H. Y. Tam. 2004, Nov 5. Machine fault diagnosis through an effective exact wavelet analysis. *Journal of Sound and Vibration* 277 (4):1005–24.

- Rabuñal, J. R. and J. Dorado. 2006. Artificial neural networks in real-life applications. p. 375. Hershey: Idea Group Publishing.
- Radzevich, S. P. 2012. Dudley's Handbook of Practical Gear Design and Manufacture, 2nd Edition. p. 874. CRC Press.
- Rafiee, J., F. Arvani, A. Harifi, and M. H. Sadeghi. 2007, May 31. Intelligent condition monitoring of a gearbox using artificial neural network. *Mechanical Systems and Signal Processing* 21 (4):1746–54.
- Rafiee, J., M. A. Rafiee, N. Prause, and P. W. 2009. Tse application of daubechies 44 in machine fault diagnostics. 2nd International Conference on Computer, Control and Communication, Karachi, IEEE, pp. 1–6. doi: [10.1109/IC4.2009.4909247](https://doi.org/10.1109/IC4.2009.4909247).
- Salchenberger, L. M., E. Cinar, and N. A. Lash. 1992, Jul 1. Neural networks: A new tool for predicting thrift failures. *Decision Sciences* 23 (4):899–916.
- Samanta, B. 2004. Gear fault detection using artificial neural networks and support vector machines with genetic algorithms. *Mechanical Systems and Signal Processing* 18 (3):625–44.
- Saravanan, N., and K. I. Ramachandran. 2010, Jun 30. Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN). *Expert Systems with Applications* 37 (6):4168–81.
- Sharma, V., and A. Parey. 2016. A review of gear fault diagnosis using various condition indicators. *Procedia Engineering* 144:253–63.
- Tse, P. W. 2004. Yang W and Tama HY. Machine fault diagnosis through an effective exact wavelet analysis. *Journal of Sound and Vibration* 277:1005–24.
- Tyagi, S., and S. K. Panigrahi. 2017, Apr 10. A DWT and SVM based method for rolling element bearing fault diagnosis and its comparison with artificial neural networks. *Journal of Applied and Computational Mechanics* 3 (1):80–91.
- Yao, X., C. Guo, M. Zhong, Y. Li, G. Shan, and Y. Zhang 2009 Wind turbine gearbox fault diagnosis using adaptive morlet wavelet spectrum. In Intelligent Computation Technology and Automation. ICICTA'09. Second International Conference on 2009 Oct 10. IEEE, Vol. 2, pp. 580–83.
- Zhou, G. H., C. C. Zuo, J. Z. Wang, and S. X. Liu Gearbox fault diagnosis based on wavelet-AR model. In 2007 International Conference on Machine Learning and Cybernetics 2007 Aug 19 (Vol. 2, pp.1061–65). IEEE.