Using PCA with LVQ, RBF, MLP, SOM and Continuous Wavelet Transform for Fault Diagnosis of Gearboxes

M. Heidari[1], H. Homaei[2], H.Golestanian[3], A.Heidari[4]

[1]Mechanical Engineering Group, Shahrekord University, Shahrekord, Iran [2,3,4] Faculty of Engineering, Shahrekord University, P. O. Box 115, Shahrekord, Iran

Abstract

A new method based on principal component analysis (PCA) and artificial neural networks (ANN) is proposed for fault diagnosis of gearboxes. Firstly the six different base wavelets are considered, in which three are from real valued and other three from complex valued. Two wavelet selection criteria Maximum Energy to Shannon Entropy ratio and Maximum Relative Wavelet Energy are used and compared to select an appropriate wavelet for feature extraction; next, the continuous wavelet coefficients (CWC) are evaluated for some different scales. As a new method, the optimal range of wavelet scales is selected based on the maximum energy to Shannon entropy ratio criteria and consequently feature vectors are reduced. In addition, energy and Shannon entropy of the wavelet coefficients are used as two new features along with other statistical parameters as input of the classifier. To prevent the curse of dimensionality problem, the principal component analysis applies to this set of features. Finally, the gearbox faults are classified using these statistical features as input to machine learning techniques. Four artificial neural networks are used for faults classifications. The test result showed that the MLP identified the fault categories of gearbox more accurately for both real wavelet and complex wavelet and has a better diagnosis performance as compared to the RBF, LVQ and SOM.

Keywords*: gearbox, wavelet, PCA, MLP, RBF, SOM, LVQ.*

1. Introduction

Gearbox is one of the most popular machines in the world. The importance and need of this machine are clear; so, fault diagnosis of them is a core research area in the condition monitoring field. Fault detection and diagnosis of gearboxes [1, 2] is one of the most common and intricate challenges in industries as a result of frequent gear defects in machines [3,4]. Vibration signal processing of gears [5] is categorized as a reliable method in condition monitoring. To analyze vibration signals, various techniques such as time [6,7], frequency [8], and time–frequency domain [9] have been extensively studied. Among these, wavelet transform [10–13] has progressed in the last two decades, and outweighs the other time–frequency methods, although it is lacking in a few aspects as well. Wavelet transform (WT) has attracted many researchers' attention recently. The wavelet transform was utilized to represent all possible types of transients in vibration signals generated by faults in a gearbox [14]. Among the various methods for condition monitoring of machinery, artificial neural network (ANN) have become in the recent decades the outstanding method exploiting their non-linear pattern classification properties, offering advantages for automatic detection and identification of gearbox failure conditions, whereas they do not require an in-depth knowledge of the behavior of the system. A neural network was used to diagnose a simple gear system after the data have been pre-processed by the wavelet transform [15]. Wavelet transform was used to analyze the vibration signal from the gear system with pitting on the gear [16]. Hence, based on the literature review there exist a wide scope to explore machine learning methods like ANN, SVM and PSVM for fault diagnosis of the gearbox. Samantha and Balushi [17] have presented a procedure for fault diagnosis of rolling element bearings through artificial neural networks. The characteristic features of time-domain vibration signals of the rotating machinery with normal and defective bearings have used as inputs to the ANN. Lei et al. [18] have proposed a method for intelligent fault diagnosis of rotating machinery based on wavelet packet transform (WPT), empirical mode decomposition (EMD), dimensionless parameters, a distance evaluation

technique and radial basis function (RBF) network. Paya et al. [11] presented artificial neural network served as a classifier to diagnose faults of rotating machinery. In this study, the method by using multilayer artificial neural networks on the sets of preprocessed data by wavelet transforms single and multiple faults have a direct effect on the diagnostic effectiveness of distinct group faults. Rafiee et al. [19] have developed a procedure which experimentally recognizes gears and bearing faults of a typical gearbox system using a multilayer perceptron neural network. Rafiee and Tse [20] have presented a time–frequency-based feature recognition system for gear fault diagnosis using autocorrelation of continuous wavelet coefficients. It has been shown that the size of vibration signals can be reduced with minimal loss of significant frequency content. Rafiee et al. [21] have shown that the Daubechies 44 wavelet is the most effective for both faulty gears and bearings. Using the genetic algorithm, Rafiee et al. Optimized the decomposition level of the WPT, the order of the Daubechies wavelet function, and number of neurons in the hidden layer of an ANN identify slight-worn, medium-worn and broken-tooth of gears faults perfectly [22]. Kankar et al. [23] have conducted a comparative experimental study of the effectiveness of ANN and SVM in fault diagnosis of ball bearings and concluded that the classification accuracy for SVM is better than of ANN. Saravanan and Ramachandran [24] investigated the usage of discrete wavelets for feature extraction and application of artificial neural network for classification. In this study, a feed-forward multilayer perceptron (MLP) neural network is utilized for classification. Regarding the references that have shown that neural networks can be effectively used in the diagnosis of various gear faults, in this study, a feed-forward multilayer perceptron (MLP) neural network is utilized for classification. In addition, two new neurons are considered as inputs of the network to enhance the accuracy of the ANN. Hajnayeb et al. [25] investigated, a system based on artificial neural networks (ANNs) to diagnose different types of fault in a gearbox. An experimental set of data was used to verify the effectiveness and accuracy of the proposed method. The system was optimized by eliminating unimportant features using a feature selection method (UTA method). This method of feature selection was compared with genetic algorithm results. Bafroui and Ohadi [26] used a feed-forward multilayer perceptron (MLP) neural network for fault diagnosis of an automobile gearbox with Morlt wavelet. In this research, to improve the speed of fault diagnosis with high efficiency performance, the optimal scales of wavelet were selected based on the maximum Energy to Shannon Entropy ratio criteria. Gharavian et al. [27] used from FDA-based and PCA-based features for fault diagnosis of automobile with Gaussian mixture model and K nearest neighbor.

The paper is organized as follows. Section 2 is dedicated to the CWT and wavelet selection criterion methods. Section 3 is dedicated to the feature extraction of CWC. In Section 4, artificial neural network such as MLP, SOM, LVQ and RBF is given. In the following sections, we will describe concisely experimental setup, data collection and fault description. The effectiveness of the proposed method will be investigated through a numerical simulation study in Section 6. The conclusion of this paper is given in Section 7.

2. Continuous Wavelet Transform

The principles of continuous wavelet transform are presented in this section. Wavelet transform supplies a power tool that can be accustomed to detect both stationary and transitory signals. The wavelet technique has special benefits for describing signals at various localization levels in time, in addition to frequency domains [28].

The continuous wavelet transform of a signal x(t) is defined as follows:

$$W_{\psi}(a,b) = \int_{-\infty}^{+\infty} x(t)\psi_{a,b}\,(t)dt \quad (1)$$

Where

$$\psi_{a,b}(t) = |a|^{-\left(\frac{1}{2}\right)}\psi\left(\frac{(t-b)}{a}\right) \quad (2)$$

is called daughter wavelet derived from mother wavelet ψ(t) which in this definition, it must be real and satisfy $\int \psi(t)dt = 0$. In Eq. (2), a and b are real value parameters, denoting scale and translation, respectively. After acquiring the vibration signals in the time domain, it is processed to obtain feature vectors. The continuous wavelet transform is used for obtaining the wavelet coefficients of the signals. The statistical parameters of the wavelet coefficients are extracted, which make up the feature vectors.

2.1. Maximum Relative Wavelet Energy criterion

Relative Wavelet Energy (RWE) is considered as time-scale density that can be used to detect a specific phenomenon in time and frequency planes. RWE gives information about the relative energy with associated frequency bands and can detect the degree of similarity between segments of a signal [29, 30]. The energy at each resolution level n, will be the energy content of the signal at each resolution is given by

$$E(n) = \sum_{i=1}^{m} |C_{n,i}|^2 \quad (3)$$

Where 'm' is the number of wavelet coefficients and $C_{n,i}$ is the ith wavelet coefficient of nth scale. The total energy can be obtained by

$$E_{total} = \sum_n \sum_i |C_{n,i}|^2 = \sum_n E(n) \quad (4)$$

Then, the normalized values, which represent the Relative Wavelet Energy, is the energy probability distribution, defined as [31]

$$p_n = \frac{E(n)}{E_{total}} \quad (5)$$

Where $\sum_n p_n = 1$, and the distribution, p_n, is considered as a time scale density. The total energy is calculated for each scale and for vibration signals at different rotor speed 1800, 2100, 2400, 2700 and 3000 RPM and for different loading conditions using healthy and faulty bearings. Relative Wavelet Energy is calculated for each scale by taking the ratio of total energy of nth scale and sum of total energy of all scales.

2.2. Maximum Energy to Shannon Entropy ratio criterion

An appropriate wavelet is selected as the base wavelet, which can extract the maximum amount of energy while minimizing the Shannon entropy of the corresponding wavelet coefficients. A combination of the energy and Shannon entropy content of a signal's wavelet coefficients is denoted by energy to Shannon entropy ratio [31] and is given as

$$\zeta(n) = \frac{E(n)}{S_{entropy}(n)} \quad (6)$$

Where entropy of signal wavelet coefficients is given by

$$S_{entropy}(n) = -\sum_{i=1}^{m} p_i \cdot \log_2 p_i \quad (7)$$

Where p_i is the energy probability distribution of the wavelet coefficients, defined as

$$p_i = \frac{|C_{n,i}|^2}{E(n)} \quad (8)$$

With $\sum_{i=1}^{m} p_i = 1$, and in the case of $p_i = 0$ for some *i*, the value of $p_i . \log_2 p_i$ is taken as zero.

3. Feature extraction

In a classification task, a very high-dimensional feature vector, e.g. the features obtained from CWT analysis, typically causes the curse of dimensionality problem; moreover, using such features for training classifiers may be a time consuming procedure. In such a case, high dimension data could be projected on a manifold with a lower dimension. There are various approaches for dimension reduction in the literature, which one of them is introduced in the present study.

3.1. Principle Component Analysis

Principal component analysis (PCA) is an important method developed by Hotelling in 1936 (see the monograph on PCA by Jolliffe [32] or books [33, 34]. As stated in [32] and recently confirmed by He et al. [35], the PCA method has been used extensively in the past and still remains as one of the most popular methods used for data mining, feature extraction, data reduction and visualization. Two various definitions for PCA are introduced: The maximum variance and minimum error formulations. In the first definition, PCA is assumed as a transformation that projects data to a new space with lower dimension than the original data in which, the data variances on the new axes are maximized. In the second interpretation, it projects data into a reduced-dimension space that by reconstructing original samples, the minimum error occurs. Let's assume xn: n=1: N is a set of data sample vectors in RD which their mean is zero. This data can be described with a linear combination of a set of orthonormal basis vectors {ui : i =1,…, D} as given by

$$x_n = \sum_{i=1}^{D} \alpha_{ni} u_i \quad (9)$$

Where αni is obtained from the inner product of xn and ui ($\alpha_{ni} = x_n^T u_i$ coefficients for any data sample). Data samples could be described by the basis vectors, as given by

$$x_n = \sum_{i=1}^{D} (x_n^T u_i) u_i \quad (10)$$

It is assumed that the first M dimensions describe a new basis. Without loss of generality, xn can be projected to a new subspace ($\tilde{x}_n$)

$$\tilde{x}_n = \sum_{i=1}^{M} z_{ni} u_i + \sum_{i=M+1}^{D} b_i u_i \quad (11)$$

In Eq. (11), zni (i=1,…, M) is calculated for each data sample individually and bi are constant for all of the data samples. Choosing appropriate values for ui and bi is performed such that the distortions introduced by dimension reduction is minimized (minimization of $\{J = \frac{1}{N}\sum_{n=1}^{N} x_n - \tilde{x}_n^2\}$).

This minimization is straight forward using Lagrange multipliers, as described in [34]. ui is the first M eigenvectors of covariance matrix of data points (S), corresponding to the M largest eigenvalues (λi) of S matrix. The parameter d is the percent of distortion occurs by the dimension reduction as defined in Eq. (12)

$$d = \frac{\sum_{i=M+1}^{D} \lambda_i}{\sum_{i=1}^{D} \lambda_i} \times 100\% \quad (12)$$

In practice, M is chosen such that d is less than one percent. In this research, the PCA method is used for the dimension reductions [27].

4. Artificial Neural Network (ANN)

After the dimension reduction, a proper classifier should be trained for the fault sorting. In this paper, four different artificial neural network classifiers are exploited. ANN is one of the approaches to forecast and validate using computer models with some of the architecture and processing capabilities of the human brain [36]. The technology that attempts to achieve such results is called neural computing or artificial neural networks. ANN mimics biological neurons by simulating some of the workings of the human brain. An ANN is made up of processing elements called neurons that are interconnected in a network. The artificial neurons receive inputs that are analogous to the electro-chemical signals that natural neurons receive from other neurons. By changing the weights given to theses signals, the network learns in a process that seems similar to that found in nature. i.e., neurons in ANN receive signals or information from other neurons or external sources, perform transformations on the signals, and then pass those signals on to other neurons. The way information is processed and intelligence is stored depends on the architecture and algorithms of ANN. A main advantage of ANN is its ability to learn patterns in very complex systems. Through learning or self-organizing process, they translate the inputs into desired outputs by adjusting the weights given to signals between neurodes.

4.1 Multi-Layer Perceptron (MLP)

In this paper, a feed-forward MLP neural network is used for classification of the extracted features from CWT. A multilayer neural network that is a network with multiple layers of neurons (usually three layers), namely an input layer that obtains signals from a particular supplier, a hidden layer that processes the data and an output layer that sends processed data to outside [37, 38]. The back propagation of an ANN assumes that there is a supervision of learning of the network. The method of adjusting weights is designed to minimize the sum of the squared errors for a given training data set:

j – Identifies a receiving node,

i – denotes the node that feeds a second node,

I – denotes the input to a neuron,

O – Denotes output of a neuron,

Wij – denotes the weights associated with the nodes.

Each non-input node has an output level O_j where

$$O_j = \frac{1}{1+e^{-I_j}}\ ,\ I_j = \sum W_{ij} O_j \quad (13)$$

Where Oi is each of the signals to node j (i.e., the output of node of i). The derivation of the back propagation formula involves the use of the chain rule of partial derivatives and equals:

$$\delta_{ij} = \frac{\partial SSE}{\partial W_{ij}} = \left(\frac{\partial SSE}{\partial O_j}\right)\left(\frac{\partial O_j}{\partial I_j}\right)\left(\frac{\partial I_j}{\partial W_{ij}}\right) \quad (14)$$

Where by convention the left-hand side is denoted by δij, the change in the sum of squared errors (SSE) attributed to Wij. Now error is given by

$$e_i = \left(D_j - O_j\right), \qquad SSE = \sum (D_j - O_j)^2 \quad (15)$$

Therefore,

$$\left(\frac{\partial SSE}{\partial O_j}\right) = -2\sum D_j - O_j \quad (16)$$

From the output of the output node, we obtain,

$$\left(\frac{\partial O_j}{\partial I_j}\right) = O_j(1 - O_j) \quad (17)$$

The input to an input node is $I_j = \sum W_{ij}\, O_i$. Therefore the change in the input to the output node, resulting from the previous hidden node, *i*, is

$$\left(\frac{\partial I_j}{\partial W_{ij}}\right) = O_i \quad (18)$$

Thus, from the above equations, the *j*th delta is

$$\delta_{ij} = 2e_j O_j\left(1 - O_j\right)O_i \quad (19)$$

Now the old weight is updated by the following equation:

$$\Delta W_{ij}(new) = \eta \delta_{ij} O_j + \alpha \Delta W_{ij}(old) \quad (20)$$

For the hidden layers, the calculations are similar. The only change is how the ANN output error is back propagated to the hidden layer nodes. The output error at the *i*th hidden node depends on the output errors of all nodes in the output layer. This relationship is given by

$$e_i = \sum W_{ij} e_j \quad (21)$$

After calculating the output error for the hidden layer, the update rules for the weights in that layer are the same as the previous update [39, 40].

4.2. Self-organizing maps (SOM)

As shown in Fig. 1, the SOM neural network structure is composed by input layer and competitive layer. Input layer is a one-dimensional vector, and the competition level is a two-dimensional planar array. The whole network is realized by connecting the neurons of the input layer and the competitive layer.

SOM can find similarity of the input data automatically, and put them together on the network. The learning algorithm is as follows.

(1) Choose a smaller random value for initial weight vector $w_j(0)(j = 1, 2, \dots, l)$. Each $w_j(0)$ is different from others [41], and l is the number of neurons in the network.

(2) Input vector $X = (x_1, x_2, x_3, \dots, x_m)^T$ to the input layer.

(3) Calculate the Euclidean distance between the weight vectors of the map layer and the input vector. The distance between the jth neuron and the input vector is expressed as follows:

$$d_j = \|X - W_j\| = \sqrt{\sum_{i=1}^{m}(x_i(t) - \omega_{ij}(t))^2} \quad (22)$$

Where ωij is the weight between neuron i of the input layer and neuron j of the map layer. A neuron called winning neuron, which is the nearest neuron from the input vector, is obtained by calculation, and it is denoted as j*. The collection of neighboring neurons is also acquired simultaneously.

(4) Fix weight of the output neuron j* and its neighboring neurons as follows:

$$\Delta\omega_{ij} = \omega_{ij}(t+1) - \omega_{ij}(t) = \eta(t)(x_i(t) - \omega_{ij}(t)) \quad (23)$$

Where $\eta(t)$ is a constant between 0 and 1.

$$\eta(t) = \frac{1}{t} \; or \; \eta(t) = 0.2\left(1 - \frac{t}{10000}\right) \quad (24)$$

(5) Calculation of output as:

$$o_k = f(min_j\|X - W_j\|) \quad (25)$$

Where f (*) is a function or other non-linear function between *0* and *1*.

(6) The calculation is finished if the results can meet the requirements, otherwise it should return to step (2) and repeat the process [41].

4.3. Learning vector quantization (LVQ)

Learning vector quantization [42] is a supervised machine learning technique in which the structure of the input space is exploited so that the size of the input data can be reduced which results in less computational time. LVQ is based on vector quantization in which an input space is divided into a number of distinct regions and for each region a reconstruction vector is defined. When a new vector is presented to the quantizer, the region in which the vector lies is first determined, and is then represented by the reproduction vector for that region. The collection of possible reproduction vectors is called the code book of quantizer, and its members are called code words. The SOM algorithm provides an approximate method for computing the Voronoi vectors in an unsupervised manner, with the approximation being specified by the synaptic weight vectors of the neurons in the feature map. Therefore, we can say that firstly SOM can be employed for the computation of the feature map and secondly LVQ is applied, which provides a mechanism for the final tuning of a feature map. Hence LVQ is said to be supervised version of SOM [43]. Although there are separate algorithms in the literature on LVQ, the LVQ1 and LVQ2 algorithm have been used in this paper [44].

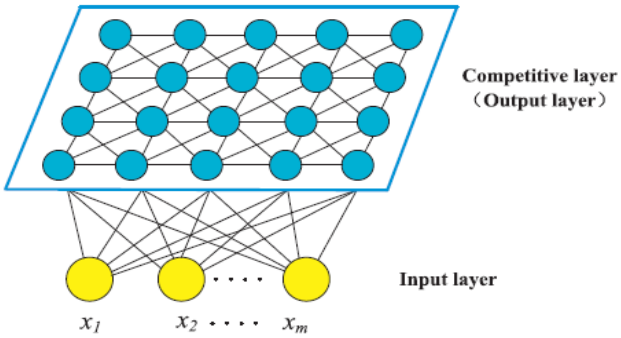

Fig1. Structure of SOM neural network [41].

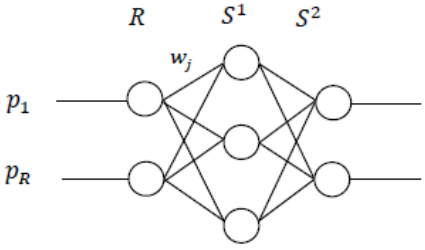

Fig2. RBF neural network architecture: R=number of elements in input vector, S1=number of neurons in layer 1, S2=number of neurons in layer 2, wj=number of weights in layer 1 [46].

4.4. Radial basis function neural network

The RBF neural network [45] is a feed forward network with its architecture as shown in Fig. 2. It consists of three layers: an input layer of *R* neurons, a hidden radial basis layer of S^1 neurons and an output linear layer of S^2 neurons. The information of the input neurons will transfer to the neurons in the hidden layer. The RBF in the hidden layer will respond to input information, and then generates the outputs in the neurons of the output layer. The advantage of the RBF network is that the hidden neurons will have non-zero response if the minimum of a function is in the pre-defined, limited range of the input values, otherwise, the response will be zero. Therefore, the number of active neurons is smaller and the time required in training the network is less. Hence, the RBF network is also referred to as the local range network. In the RBF neural network, the transfer function of the hidden layer is a Gaussian function

$$A_j = \exp\left[-\frac{(p-c_j)^T(p-c_j)}{2\sigma_j^2}\right], j = 1, 2, \ldots, S^1, \qquad (26)$$

Where Aj is the output of the jth neuron in the hidden layer, p is the input mode, cj is the center of the jth neuron Gaussian function. σ_j^2 is the unitary parameter, and S1 is the neurons in hidden layer 1. There are two steps in the training of the RBF network. In the first step, depending on the information contained in the input samples, the neurons in the hidden layer S1; the center of Gaussian function cj and the unitary parameter σ_j^2 will be determined. The most frequently used methods to determine the Gaussian function is the K-means aggregation algorithm.

In the second step, according to the parameter in the hidden layer, input samples and the target values, the weight wj will be determined and adjusted by the principle of least squares.

5. Experimental set-up, data collection and faults description

Rolling element bearings and gears are the most common and important components used in rotating machinery such as gearboxes. Faults occurring on the surface of these components could cause unexpected machine breakdown. Therefore, it is necessary to develop an effective intelligent gearbox fault diagnosis method. To verify the effectiveness of the proposed method, gearbox datasets provided by the Prognostics and Health Management Society are analyzed.

5.1. Prognostics and Health Management Society

Data collected in this section come from public datasets distributed by Prognostics and Health Management (PHM) Society under 2009 PHM challenge competition [47]. The data are representative of a generic industrial gearbox shown in figure 3. Data were sampled synchronously from accelerometers mounted on both the input and output shaft retaining plates of the gearbox. An attached tachometer generates 10 pulses per revolution, providing very accurate zero crossing information. Data were collected at 30, 35, 40, 45 and 50 Hz shaft speed under high and low loading. The test runs include seven different combinations of faults and one fault-free reference run. The signals were sampled with sampling frequency 66.666 kHz and the sampling horizon was 4s long.

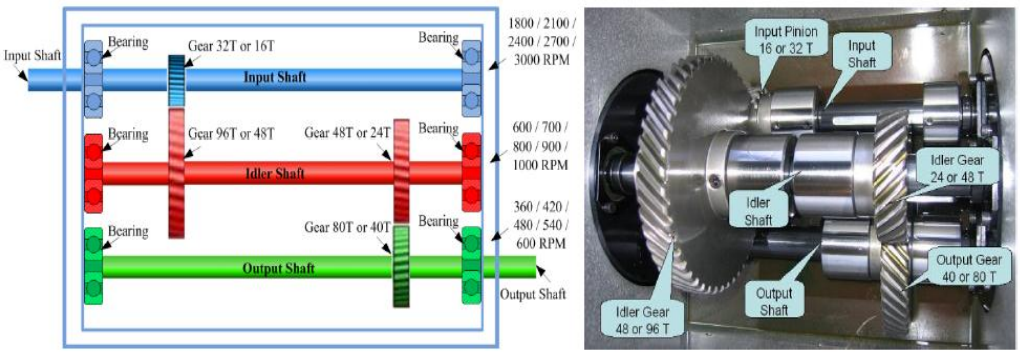

Fig3. PHM society gearbox data collection setup [47].

Table 1. Comparison of parameters for wavelet selection

Wavelet type	Energy to Shannon Entropy ratio	Maximum Relative Wavelet Energy
Complex Morlet	25.516	0.054756
Complex Gaussian	130.348	0.015358
Shannon	98.139	0.035353
Meyr	137.630	0.014681
Sym2	122.561	0.015962
rBio5.5	143.992	0.011401

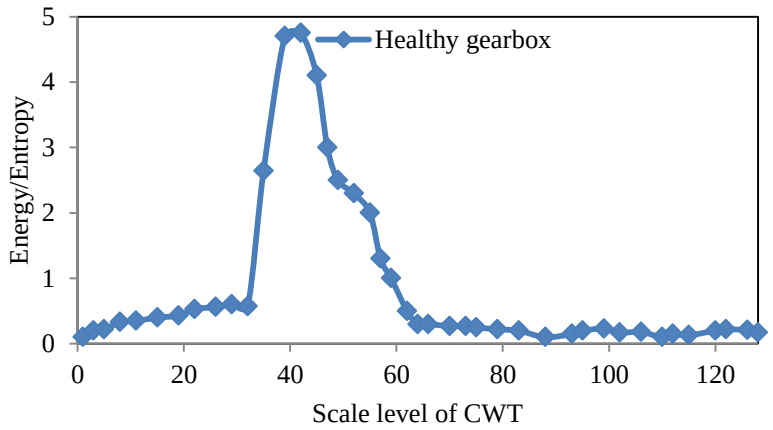

Fig4. Plot between Energy to Shannon Entropy ratio vs. Scale number for rBio5.5 wavelet (Maximum Relative Wavelet Energy criterion).

6. Results and discussions

In the present study, training and testing of the classifiers as MLP, RBF, LVQ and SOM have been carried out. For healthy and faulty bearings and gears, continuous wavelet coefficients (CWC) of vibration signals are calculated using six different mother wavelets in which there are real valued as Meyer, rBio5.5, Symlet2 wavelets and other three are complex valued as complex Gaussian, complex Morlet and Shannon wavelets. The total energy and total Shannon entropy of CWC is calculated for vibration signals at different rotor speed 1800, 2100, 2400, 2700 and 3000 rpm and for different loading conditions using healthy and faulty gearbox. The total energy to total Shannon entropy ratio for each wavelet is calculated as shown in Table 2. The total energy is calculated for each scale and for vibration signals at different rotor speed and for different loading conditions using healthy and faulty conditions. Relative wavelet energy is calculated for each scale by taking the ratio of total energy of nth scale and sum of total energy of all scales. Maximum value of RWE is selected for each wavelet as shown in Table 1.

The Energy to Shannon Entropy ratio (ESER) is obtained maximum for rBio5.5 wavelet. Hence, rBio5.5 wavelet is considered as the base wavelet to extract features for fault diagnosis. Similarly, the Complex Morlet wavelet is selected based on Maximum Relative Wavelet Energy (MRWE) criterion. The statistical parameters of the wavelet coefficients are extracted in 1–128 scale of CWT basic wavelet. In this study, the Morlet wavelet and rBio5.5 wavelet are selected based on two selection wavelet criteria. Seven statistical features are selected like standard deviation, crest factor, absolute mean amplitude value, variance, kurtosis, skewness and fourth central moment of wavelet coefficients of vibration signals. Some statistical features are defined as follows [48, 49]:

Mean value: mean value is the average of the signal.

Standard deviation (STD): Standard deviation is a measure of energy content in the vibration signal

$$Standard\ Deviation = \sqrt{\frac{n\sum x^2-(\sum x)^2}{n(n-1)}} \quad (27)$$

Skewness: skewness characterizes the degree of asymmetry of a distribution around its mean. Skewness can come in the form of negative or positive skewness

$$Skewness = \frac{n}{(n-1)(n-2)}\sum(\frac{x_i-\bar{x}}{s})^3 \quad (28)$$

Kurtosis: Kurtosis is a dimensionless, statistical measure that characterizes the flatness of a signal's probability density function. A signal containing many sharp peaks will result in a high kurtosis value. The mathematical definition of the kurtosis value can be found in equation (29). The kurtosis value is used to identify broken or chipped gears as well as the bearing faults. If the signal is normally distributed the parameter value will be close to 3.

$$Kurtosis = \left\{\frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum\left(\frac{x_i-\bar{x}}{s}\right)^4\right\} - \frac{3(n-1)^2}{(n-2)(n-3)} \quad (29)$$

For feature reduction, the optimal amounts of scales should be extracted. Wavelet entropy and energy are used to select the optimal scales of the basic Morlet wavelet and rBio5.5 wavelet. The proper scales are chosen when the Shannon entropy of the corresponding wavelet coefficients is minimum and the energy is maximum. To search for optimal scales, the distribution of energy to Shannon entropy with rBio5.5 wavelet and complex Morlet wavelet belongs to each scale level from scale 1 to 128 for healthy gears and bearings are shown in figures 4 and 5. From figures 4 and 5, the optimal scales with two criteria are between 35-62 and 52-70 respectively.

In case A, statistical parameters in all scales (1–128) of continuous wavelet transform are considered as feature sets which means that no feature reduction is used. Case B is related to the condition that statistical features in optimal scales, which has been extracted based on the criteria of maximum energy to Shannon entropy ratio, are considered. In case C, in addition to statistical features in optimal scales, energy and Shannon entropy factors are used as two new features as features sets. In case D, in addition the condition of case C, the PCA is applied to CWC for more reduction in feature sets. After finding the optimal scales based on maximum energy to Shannon entropy, which results in 65% and 59% feature reduction with MRWE and ESER respectively. For better classification, the energy and Shannon entropy factors are used as two new features in the input of the classifier. There are ten neurons in the input layer for this case, out of which seven are statistical parameters, number of rotation and the other two are energy and Shannon entropy features. As shown in Table 3, the performance of the ANN classifier for machinery fault diagnosis is acceptable. It can be seen that the adoption of maximum energy to Shannon entropy ratio for features reduction (case B-optimal scales) leads the diagnosis method to higher accuracy, about 11–21 percentage points greater than the case with no feature reduction (case A). In addition, the training time of ANN with feature reduction is about 2.45 times less than the corresponding time of before feature reduction. On the other hand, Table 3 indicates that the usage of energy and Shannon entropy as two new features along with other statistical features in optimal scales (case C) increases the accuracy of the diagnosis method by 20–30 percentage points in comparison with the no feature reduction case (case A). Also, this idea improves the performance of the classifier between 5 to 15 percentage points in comparison with case B, in which the energy and Shannon entropy have not been considered as the input of the classifier; however, there is no significant difference between the training time of the two cases B and C. Also, using PCA in feature reduction (Case D), improves the performance of the classifier between 1 to 3 percentage points in comparison with case C. The results of classification with Maximum Energy to Shannon Entropy ratio criterion and using MLP, SOM, LVQ and RBF have been shown in Tables 3, 4, 5 and 6 respectively. From these Tables we found that the MLP network has the best accuracy in classification of gearbox faults in comparison of SOM, LVQ and RBF. Table 3 shows the classification rates for each of the implemented datasets. As expected, the use of the PCA leads to perfect classification. According to Tables 3, 4, 5 and 6, it is clearly seen that the PCA projects the conditions samples into new spaces leads to the highest accuracy. Tables 7, 8, 9 and 10 show the recognition rates for the combination faults of gears and bearings of gearbox with Maximum Relative Wavelet Energy criterion. The results show the validity of the proposed method. The overall performance for MLP (case D) and with MESE and MRWE criteria is 98.70% and 96.378%, respectively. From Tables 3 and 7, it is noticeable to be mentioned that the time needed to implement the MLP classifier with MRWE is bigger than the training time needed for the MESE criteria. As can be seen in Tables 3 and 7, the best result of classification can be obtained is 100% and the number of hidden neurons is 30. Tables 6 and 10 show the performance of the RBF network with two wavelet criteria and four datasets. It shows that the RBF network employing the K-means aggregation algorithm used 50.97 s and 315 steps in training, whilst the MLP network used only 46.74 s and 253 steps in training. Hence, the performance of the MLP network is better than the RBF network, particularly in the computational resource for training.

As it can be seen in Tables 5 and 9, the maximum probability of correct classification for the healthy and faulty gearbox condition is about 98%. These Tables demonstrate the results by applying LVQ1 and LVQ2 algorithm. After applying LVQ2 algorithm, the results show that there is an improvement between 1% to 3% in the classification percentage as shown in Tables 5 and 9. In these Tables, Tr. A., Te. A., and T. T. Is training accuracy, testing accuracy and training time respectively. In all cases the training time of LVQ1 is bigger than LVQ2. The accuracy of the LVQ classifier in fault diagnosis of gearbox is after of the MLP and RBF. From Table 5, PCA overall performance is about 90.97% and 93.13% for the LVQ1 and LVQ2 classifiers respectively. Similarly from Table 9, when we apply the Maximum Relative Wavelet Energy criterion to LVQ1 algorithm and LVQ2 algorithm, the overall performance classification (case D) is 89.02% and 91.33%, respectively.

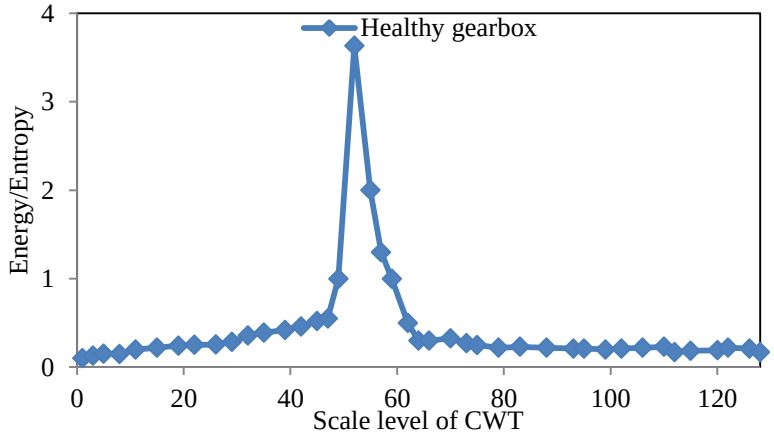

Fig5. Plot between Energy to Shannon Entropy ratio vs. Scale number for Complex Morlet wavelet (Maximum Energy to Shannon Entropy ratio criterion).

Table 2. Four different cases as input for ANN

Case name	Feature sets
A	Statistical parameters in all scales
B	Statistical parameters in optimal scales
C	Statistical parameters and energy and Shannon entropy in optimal scales
D	Statistical parameters and energy and Shannon entropy in optimal scales with PCA

Table 3. Classification rate for gearbox with MLP (Maximum Energy to Shannon Entropy ratio criterion)

Classifier (inputs)		Gear faults				Bearing faults				
		Healthy	Chipped	Broken	Eccentric	Healthy	Inner	Ball	Outer	Combine
Case A	Tr. A. (%)	80.64	70.50	78.89	76.64	82.96	81.70	83.52	80.19	84.66
	Te. A. (%)	79.64	70.02	77.42	74.58	80.29	80.91	81.66	79.08	82.81
	T. T. (s)	100.66								
Case B	Tr. A. (%)	93.13	84.68	89.93	88.63	92.50	91.50	93.79	92.21	94.39
	Te. A. (%)	91.61	83.77	87.50	86.31	90.42	89.48	91.70	90.17	92.90
	T. T.(s)	40.94								
Case C	Tr. A. (%)	97.78	97.80	98.47	93.59	97.58	96.07	98.01	97.53	98.89
	Te. A. (%)	95.89	96.01	96.73	91.97	95.64	94.39	96.85	95.14	96.22
	T. T.(s)	50.19								
Case D	Tr. A. (%)	99.24	99.12	98.89	95.96	98.49	98.63	99.61	98.40	100
	Te. A. (%)	97.10	97.84	96.46	93.60	96.39	96.18	98.49	96.10	98.89
	T. T.(s)	46.74								

Table 4. Classification rate for gearbox with SOM (Maximum Energy to Shannon Entropy ratio criterion)

Classifier (inputs)		Gear faults				Bearing faults				
		Healthy	Chipped	Broken	Eccentric	Healthy	Inner	Ball	Outer	Combine
Case A	Tr. A. (%)	74.20	69.39	74.18	71.85	75.62	76.53	77.38	78.41	78.36
	Te. A. (%)	72.28	68.06	72.62	70.23	73.19	74.21	75.91	76.31	76.10
	T. T. (s)	116.67								
Case B	Tr. A. (%)	86.39	77.79	83.80	82.49	86.48	85.51	87.55	88.42	88.59
	Te. A. (%)	84.29	75.36	81.94	80.44	84.35	83.58	86.47	86.22	86.37
	T. T. (s)	60.18								
Case C	Tr. A. (%)	91.67	90.43	92.55	86.74	90.49	89.06	92.38	91.54	92.19
	Te. A. (%)	89.05	88.66	90.29	84.98	88.85	87.18	90.64	89.28	90.59
	T. T. (s)	65.73								
Case D	Tr. A. (%)	93.18	93.28	92.70	88.61	92.38	93.29	94.19	94.20	95.27
	Te. A. (%)	91.29	91.64	90.49	86.55	90.59	91.08	93.32	92.79	93.46
	T. T. (s)	62.80								

Table 5. Classification rate for gearbox with LVQ (Maximum Energy to Shannon Entropy ratio criterion)

Classifier (inputs)			Gear faults				Bearing faults				
			Healthy	Chipped	Broken	Eccentric	Healthy	Inner	Ball	Outer	Combine
Case A	Tr. A. (%)	LVQ1	74.20	66.37	73.29	70.10	74.17	76.35	77.92	78.03	78.18
		LVQ2	76.18	70.90	75.36	73.76	77.97	78.09	79.54	80.51	80.20
	Te. A. (%)	LVQ1	72.52	66.05	71.75	70.05	72.64	74.36	75.06	75.48	76.33
		LVQ2	74.20	68.38	74.11	72.51	75.25	76.27	77.21	78.16	78.03
	T. T. (s)	LVQ1	131.83								
		LVQ2	125.91								
Case B	Tr. A. (%)	LVQ1	86.49	77.22	83.90	82.35	86.12	85.04	87.25	88.18	88.59
		LVQ2	88.46	80.39	85.29	84.73	88.59	87.30	89.11	90.37	90.64
	Te. A. (%)	LVQ1	83.92	76.16	81.05	80.27	84.71	83.19	85.59	86.05	87.41
		LVQ2	86.17	78.52	83.72	82.39	86.24	85.36	87.44	88.10	89.05
	T. T. (s)	LVQ1	66.48								
		LVQ2	60.18								
Case C	Tr. A. (%)	LVQ1									
		LVQ2	93.42	92.19	94.38	88.37	92.61	91.49	94.27	93.71	94.30
	Te. A. (%)	LVQ1	89.04	88.47	90.10	84.82	88.32	88.63	90.59	88.60	90.84
		LVQ2	91.28	90.18	92.57	86.48	90.19	90.81	92.55	91.20	92.10
	T. T. (s)	LVQ1	75.19								
		LVQ2	70.59								
Case D	Tr. A. (%)	LVQ1									
		LVQ2	95.31	95.39	95.38	90.51	94.62	95.21	96.94	96.33	97.62
	Te. A. (%)	LVQ1	91.95	90.49	91.46	86.75	90.42	91.08	92.73	91.52	92.36
		LVQ2	93.11	93.50	93.71	88.12	92.47	93.62	94.10	94.23	95.33
	T. T. (s)	LVQ1	70.27								
		LVQ2	64.39								

Table 6. Classification rate for gearbox with RBF (Maximum Energy to Shannon Entropy ratio criterion)

Classifier (inputs)		Gear faults				Bearing faults				
		Healthy	Chipped	Broken	Eccentric	Healthy	Inner	Ball	Outer	Combine
Case A	Tr. A. (%)	78.90	72.76	77.51	75.24	80.72	80.41	81.52	82.39	82.44
	Te. A. (%)	76.52	71.70	75.79	73.94	78.06	77.98	79.74	80.05	80.88
	T. T. (s)	104.63								
Case B	Tr. A. (%)	90.90	82.81	87.16	86.71	90.10	89.98	91.53	92.08	92.98
	Te. A. (%)	88.06	80.42	85.25	84.90	88.59	88.36	89.54	90.73	90.41
	T. T. (s)	48.36								
Case C	Tr. A. (%)	95.27	94.14	96.19	90.82	94.68	93.09	95.59	95.90	96.97
	Te. A. (%)	93.19	92.11	94.45	89.27	92.08	91.10	93.39	92.90	94.81
	T. T. (s)	57.49								
Case D	Tr. A. (%)	97.66	97.83	97.17	92.47	96.19	96.84	98.57	98.84	99.06
	Te. A. (%)	95.23	94.59	95.55	90.73	94.90	94.18	96.43	95.94	97.92
	T. T. (s)	50.97								

Table 7. Classification rate for gearbox with MLP (Maximum Relative Wavelet Energy criterion)

Classifier (inputs)		Gear faults				Bearing faults				
		Healthy	Chipped	Broken	Eccentric	Healthy	Inner	Ball	Outer	Combine
Case A	Tr. A. (%)	79.28	68.71	77.36	75.20	80.17	79.81	81.63	78.48	82.18
	Te. A. (%)	77.19	69.92	75.27	75.01	78.55	78.71	79.46	76.39	80.12
	T. T. (s)	110.37								
Case B	Tr. A. (%)	91.72	82.16	87.03	86.24	90.19	89.33	91.13	90.22	92.44
	Te. A. (%)	89.27	81.63	86.95	84.77	88.19	87.79	89.15	89.31	90.29
	T. T. (s)	49.09								
Case C	Tr. A. (%)	95.37	95.89	96.90	91.72	95.27	94.36	96.93	95.27	96.68
	Te. A. (%)	93.15	94.06	94.66	90.11	93.81	92.60	94.53	93.25	94.76
	T. T. (s)	59.81								
Case D	Tr. A. (%)	97.49	96.33	96.62	93.15	96.44	96.18	97.19	96.38	97.61
	Te. A. (%)	95.54	94.39	94.16	91.33	93.90	94.78	96.08	94.29	96.74
	T. T. (s)	53.76								

Table 8. Classification rate for gearbox with SOM (Maximum Relative Wavelet Energy criterion)

Classifier (inputs)		Gear faults				Bearing faults				
		Healthy	Chipped	Broken	Eccentric	Healthy	Inner	Ball	Outer	Combine
Case A	Tr. A. (%)	72.17	68.39	72.79	99.18	73.25	74.12	75.27	76.33	76.94
	Te. A. (%)	70.19	66.09	70.12	68.27	71.95	72.91	73.62	74.22	74.16
	T. T. (s)	122.72								
Case B	Tr. A. (%)	84.36	75.27	81.33	80.28	84.44	83.11	85.67	86.70	86.19
	Te. A. (%)	82.87	73.39	79.90	78.51	82.34	81.11	85.06	83.26	83.56
	T. T. (s)	68.59								
Case C	Tr. A. (%)	89.29	88.05	90.11	84.36	88.92	87.41	90.03	89.77	90.71
	Te. A. (%)	87.27	86.18	88.95	82.15	86.69	85.60	88.19	87.79	88.73
	T. T. (s)	73.48								
Case D	Tr. A. (%)	90.18	91.82	90.22	86.38	90.98	91.19	92.28	92.81	93.71
	Te. A. (%)	89.62	89.71	88.73	85.17	88.55	89.67	91.38	90.18	91.81
	T. T. (s)	70.79								

Table 9. Classification rate for gearbox with LVQ (Maximum Relative Wavelet Energy criterion)

Classifier (inputs)			Gear faults				Bearing faults				
			Healthy	Chipped	Broken	Eccentric	Healthy	Inner	Ball	Outer	Combine
Case A	Tr. A. (%)	LVQ1	70.39	67.08	70.11	69.44	72.46	73.09	74.18	76.28	77.94
		LVQ2	74.75	68.41	73.35	71.90	75.29	76.47	77.19	78.39	78.59
	Te. A. (%)	LVQ1	70.03	65.21	70.17	66.62	70.51	73.81	73.19	74.51	74.39
		LVQ2	72.24	67.18	72.93	70.76	73.25	74.30	75.37	76.79	76.66
	T.T. (s)	LVQ1	145.84								
		LVQ2	136.51								
Case B	Tr. A. (%)	LVQ1	80.10	73.49	77.03	78.41	82.18	80.57	83.55	84.26	83.44
		LVQ2	82.14	75.77	81.83	80.56	84.20	83.29	85.42	86.71	86.58
	Te. A. (%)	LVQ1	86.33	86.97	87.73	83.53	86.16	86.04	88.19	87.72	88.42
		LVQ2	89.46	88.31	90.02	84.81	88.73	88.49	90.62	89.06	90.41
	T.T. (s)	LVQ1	75.96								
		LVQ2	69.49								
Case C	Tr. A. (%)	LVQ1	88.05	88.24	90.31	84.74	88.06	86.06	90.62	89.07	90.32
		LVQ2	91.54	90.09	92.37	86.27	90.33	89.79	92.14	91.73	92.44
	Te. A. (%)	LVQ1	86.29	87.11	87.78	82.03	86.42	86.27	88.10	87.34	88.52
		LVQ2	89.46	88.31	90.02	84.81	88.73	88.49	90.62	89.06	90.41
	T.T. (s)	LVQ1	85.52								
		LVQ2	78.19								
Case D	Tr. A. (%)	LVQ1	89.29	90.04	90.61	86.11	90.19	91.27	92.03	92.15	93.28
		LVQ2	92.41	93.63	94.10	88.32	92.54	93.75	94.16	94.80	95.57
	Te. A. (%)	LVQ1	89.07	89.64	89.16	84.15	88.25	89.46	90.19	90.05	91.25
		LVQ2	91.53	91.66	91.74	86.78	90.09	91.52	92.44	92.64	93.65
	T.T. (s)	LVQ1	78.52								
		LVQ2	72.95								

Table 10. Classification rate for gearbox with RBF (Maximum Relative Wavelet Energy criterion)

Classifier (inputs)		Gear faults				Bearing faults				
		Healthy	Chipped	Broken	Eccentric	Healthy	Inner	Ball	Outer	Combine
Case A	Tr. A. (%)	76.10	70.39	75.50	73.82	78.20	77.52	79.33	80.81	80.29
	Te. A. (%)	74.16	70.22	73.61	71.15	76.62	75.54	77.70	78.43	78.26
	T. T. (s)	115.22								
Case B	Tr. A. (%)	88.62	80.21	85.61	84.39	88.52	87.71	89.24	90.39	90.18
	Te. A. (%)	86.52	78.91	83.29	82.71	86.53	86.16	87.51	88.60	88.33
	T. T. (s)	50.63								
Case C	Tr. A. (%)	93.25	92.80	94.29	88.44	92.18	91.24	93.25	93.66	94.03
	Te. A. (%)	91.11	90.62	92.28	87.61	90.12	89.94	91.53	90.28	92.24
	T. T. (s)	61.74								
Case D	Tr. A. (%)	93.25	93.77	92.73	88.19	92.46	92.56	94.18	96.60	97.32
	Te. A. (%)	93.33	92.44	93.15	88.36	92.14	92.73	94.16	93.15	95.38
	T. T. (s)	54.59								

7. Conclusion

This study presents, a methodology for detection of gearbox faults by classifying them using four artificial intelligence techniques, like MLP, RBF, SOM and LVQ. First, CWT applied over the signal, employing the six mothers wavelet. Two wavelet selection criteria Maximum Energy to Shannon Entropy ratio and Maximum Relative Wavelet Energy are used and compared to select an appropriate wavelet for feature extraction. Results obtained from the two criteria show that the wavelet selected using Maximum Energy to Shannon Entropy ratio criterion gives better classification efficiency. Though all four techniques performed well, but the results of faults classification with MLP are better than another. To cover all the physical and meaningful frequencies of the vibration signal of the gearbox, CWC must be calculated in several scales which cause the curse of dimensionality. To solve this problem, the PCA dimension reduction method was used. To find very efficient features for classification, maximum energy to Shannon entropy ratio was employed to search for the optimal scales of CWT and consequently the features were reduced. This feature reduction method improved the performance of the classifiers in fault detection of the gearbox between11 to 21 percentage points in comparison with the case of no feature reduction. Moreover, the training time of classifiers with optimal scales reduced 2.45 times. LVQ being the supervised version of SOM the classification accuracy obtained 87.20% (LVQ2), which is better than SOM (85.227%). As a new idea, energy and Shannon entropy have been applied as two new features along with statistical parameters as input of the classifier. The obtained results indicate that the accuracy of the classifier has been increased between 5 to15 percentage points by considering these two features. Also, using PCA in feature reduction, improves the performance of the classifier between 1 to 3 percentage points in comparison with the case of no PCA.

References

[1]. Boulahbal D., Golnaraghi M. F., Ismail F., Amplitude and phase wavelet maps for the detection of cracks in geared systems. Mechanical Systems and Signal Processing, Vol. 13, Issue 3, 1999, p. 423–436.

[2]. Zhan Y., Makis V., Jardine A. K. S., Adaptive state detection of gearboxes under varying load conditions based on parametric modeling. Mechanical Systems and Signal Processing, Vol. 20, Issue 1, 2006, p. 188–221.

[3]. Baydar N., Ball A., Detection of gear failures via vibration and acoustic signals using wavelet transform. Mechanical Systems and Signal Processing, Vol. 17, Issue 4 , 2003, p. 787–804.

[4]. Wang W. Q., Farid Golnaraghi M., Ismail F., Prognosis of machine health condition using neuro-fuzzy systems. Mechanical Systems and Signal Processing, Vol. 18, Issue 4 , 2004, p. 813–831.

[5]. Bartelmus W., Zimroz R., Vibration condition monitoring of planetary gearbox under varying external load. Mechanical Systems and Signal Processing, Vol. 23, Issue 1, 2009, p. 246–257.

[6]. McFadden P. D., A revised model for the extraction of periodic waveforms by time domain averaging. Mechanical Systems and Signal Processing, Volume 1, Issue 1, 1987, p. 83-95.

[7]. Combet F., Gelman L., An automated methodology for performing time synchronous averaging of a gearbox signal without speed sensor. Mechanical Systems and Signal Processing, Vol. 21, Issue 6 ,2007, p. 2590–2606.

[8]. Minamihara H., Nishimura M., Takakuwa Y., Ohta M., A method of detection of the correlation function and frequency power spectrum for random noise or vibration with amplitude limitation. Journal of Sound and Vibration, Vol. 141, Issue 3, 1990, p. 425–434.

[9]. Wang W.J., McFadden P.D., Early detection of gear failure by vibration analysis I. Calculation of the time–frequency distribution. Mechanical Systems and Signal Processing, Vol. 7, Issue 3, 1993, p. 193–203.

[10]. Staszewski W. J., Tomlinson G. R., Application of the wavelet transform to fault detection in a spur gear. Mechanical System and Signal Processing, Vol. 8, Issue 3, 1994, p. 289–307.

[11]. Paya B. A., Esat I. I., Badi M. N. M., Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor. Mechanical Systems and Signal Processing, Vol. 11, Issue 5, 1997, p. 751–765.

[12]. Tse P. W., Yang W. X., Tam H. Y., Machine fault diagnosis through an effective exact wavelet analysis. Journal of Sound and Vibration, Vol. 277, Issues 4–5, 2004, p. 1005–1024.

[13]. Wu J. D., Liu C. H., An expert system for fault diagnosis in internal combustion engines using wavelet packet transform and neural network. Expert Systems with Applications, Vol. 36, Issue 3, Part 1, 2009, p. 4278–4286.

[14]. Wang W. J., McFadden P. D., Early detection of gear failure by vibration analysis. II. Interpretation of the time-frequency distribution using image processing techniques. Mechanical Systems and Signal Processing, Vol. 7, 1993, Issue 3, p. 205–215.

[15]. Petrille O., Paya B., Esat I. I., Badi M. N. M., in: Proceedings of the Energy-sources Technology Conference and Exhibition: Structural Dynamics and Vibration PD-Vol. 70, 1995, p. 97.

[16]. Boulahbal D., Golnaraghi M. F., Ismail F., in: Proceedings of the DETC'97, 1997, ASME Design Engineering Technical Conference DETC97/VIB-4009, 1997.

[17]. Samantha B., Al-Balushi K. R., Artificial Neural Networks based fault diagnostics of rolling element bearings using time domain features. Mechanical Systems and Signal Processing, Vol. 17, Issue 2, 2003, p. 317–328.

[18]. Lei Y., He Z., Zi Y., Application of an intelligent classification method to mechanical fault diagnosis. Expert Systems with Applications, Vol. 36, Issue 6, 2009, p. 9941–9948.

[19]. Rafiee J., Arvani F., Harifi A., Sadeghi M. H., Intelligent condition monitoring of a gearbox using artificial neural network. Mechanical Systems and Signal Processing, Vol. 21, Issue 4, 2007, p. 1746–1754.

[20]. Rafiee J., Tse P. W., Use of autocorrelation of wavelet coefficients for fault diagnosis. Mechanical System and Signal Processing, Vol. 23, Issue 5, 2009, p. 1554–1572.

[21]. Rafiee J., Rafiee M. A., Tse P. W., Application of mother wavelet functions for automatic gear and bearing fault diagnosis. Expert Systems with Applications, Vol. 37, Issue 6 , 2010, p. 4568–4579.

[22]. Rafiee J., Tse P. W., Harifi A., Sadeghi M. H., A novel technique for selecting mother wavelet function using an intelligent fault diagnosis

[23]. system. Expert Systems with Applications, Vol. 36, Issue 3, Part 1 , 2009, p. 4862–4875.
[24]. Kankar P. K., Sharma S. C., Harsha S. P., Fault diagnosis of ball bearings using machine learning methods. Expert Systems with Applications, Vol. 38, Issue 3, 2011, p. 1876–1886.
[25]. Saravanan N., Ramachandran K. I., Incipient gearbox fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN). Expert Systems with Applications, Vol. 37, Issue 6, 2010, p. 4168–4181.
[26]. Hajnayeb A., Ghasemloonia A., Khadem S. E., Moradi M. H., Application and comparison of an ANN-based feature selection method and the genetic algorithm in gearbox fault diagnosis, Expert Systems with Applications, Vol. 38, Issue 8 , 2011, p. 10205-10209.
[27]. Heidari Bafroui H., Ohadi A., Application of wavelet energy and Shannon entropy for feature extraction in gearbox fault detection under varying speed conditions. Neurocomputing, Vol. 133, 2014, p. 437–445.
[28]. Gharavian M.H., Almas Ganj F., Ohadi A.R., Heidari Bafroui H., Comparison of FDA-based and PCA-based features in fault diagnosis of automobile gearboxes, Neurocomputing, Vol. 121, 2013, p. 150–159.
[29]. Wu J., Chen J., Continuous wavelet transform technique for fault signal diagnosis of internal combustion engines. NDT & E Int., Vol. 39, Issue 4 , 2006, p. 304–311.
[30]. Yan R., Base wavelet selection criteria for non-stationary vibration analysis in bearing health diagnosis, Electronic Doctoral Dissertations for UMass Amherst, Paper AAI3275786, http://scholarworks.umass.edu/dissertations/AAI3275786, January 1, 2007.
[31]. Rosso O.A., Figliola A., Order/disorder in brain electrical activity. Revista Mexicana De Fisica, Vol. 50, 2004, p. 149–155.
[32]. Rosso O. A., Blanco S., Yordanova J., Kolev V., Figliola A., Schurmann M., Basar E., Wavelet entropy: a new tool for analysis of short duration brain electrical signals, Journal of Neuroscience Methods, Vol. 105, Issue 1, 2001, p. 65–75.
[33]. Jolliffe I. T., Principal Component Analysis, 2nd Edition, Springer, New York, 2002.
[34]. Duda R. O., Hart P. E., Stork D. G., Pattern Classification, 2nd Edition, Wiley, New York, 2001.
[35]. Bishop C. M., Pattern Recognition and Machine Learning, Springer, Singapore, 2006.
[36]. He Q., Yan R., Kong F., Du R., Machine condition monitoring using principal component representations. Mechanical Systems and Signal Processing, Vol. 23, Issue 2, 2009, p. 446–466.
[37]. Saravanan N., Siddabattuni V., Ramachandran K., Fault diagnosis of spur bevel gearbox using artificial neural network (ANN) and proximal support vector machine (PSVM). Applied Soft Computing, Vol. 10, Issue 1 , 2010, p. 344–360.
[38]. Haykin S., Neural Networks: A Comprehensive Foundation, Second ed, Pearson Education, 1998.
[39]. Bartelmus W., Zimroz R., Batra H., Gearbox vibration signal pre-processing and input values choice for neural network training, Artificial Intelligence Methods, Gliwice, Poland, 5–7 November, 2003.
[40]. Heidari M., Homaei H., Estimation of acceleration amplitude of vehicle by back propagation neural networks, Advances in Acoustics and Vibration, Vol. 2013, 2013, 7 pages, doi:10.1155/2013/614025.
[41]. Talebi H. A., Abdollahi F., Patel R. V., Khorasani K., Neural Network-Based State Estimation of Nonlinear Systems. Springer, New York Dordrecht Heidelberg London, 2010.
[42]. Ren J. H., Chen J. C., Wang N., Visual analysis of SOM network in fault diagnosis. Physics Procedia, Vol. 22, 2011, p. 333–338.
[43]. Swiercz E., Classification of parameter changes in a dynamic system with the use of wavelet analysis and neural networks. Advances in Engineering Software, Vol. 45, Issue 1, 2012, p. 28–41
[44]. Cheng G., Cheng Y. L., Shen L. H., Qiu J. B., Zhang S., Gear fault identification based on Hilbert Huang transform and SOM neural network. Measurement, Vol. 46, Issue 3 , 2013, p. 1137–1146.
[45]. Kohonen T.,The self organizing map, Proc. IEEE, Vol. 78, Issue 9, 1990, p.1464–1480.
[46]. Heidari M., Heidari A., Homaei H., Analysis of pull-in instability of geometrically nonlinear microbeam using radial basis artificial neural network based on couple stress theory, Computational Intelligence and Neuroscience, Vol. 2014, 2014, 11 pages, doi:10.1155/2014/571632
[47]. Wuxing L., Tse P. W., Guicai Z., Tielin S., Classification of gear faults using cumulants and the radial basis function network. Mechanical Systems and Signal Processing, Vol. 18, Issue 2 , 2004, p. 381–389.
[48]. Data Analysis Competition - 2009", Prognostics and Health management Society. [Online].

[49]. Available [Cited: December, 2012]: http://www.phmsociety.org/competition/PHM/09/apparatus.
[50]. Kankar P. K., Sharma S. C., Harsha S. P., Rolling element bearing fault diagnosis using wavelet transform. Neurocomputing, Vol. 74, Issue 10 , 2011, p. 1638–1645.
[51]. Akbari M., Homaei H., Heidari M., Faults diagnosis of a girth gear using discrete wavelet transform and artificial neural networks. Int. J. Advanced Design and Manufacturing Technology, Vol. 7, No. 3, 2014, p. 45-56.