International Journal of
Mathematical Modelling & Computations
Vol. 04, No. 04, Fall 2014, 309- 329

An Intelligent Fault Diagnosis Approach for Gears and Bearings Based on Wavelet Transform as a Preprocessor and Artificial Neural Networks

Mahmuod Akbari[a], Hadi Homaei[b] and Mohammad Heidari[c,*]

[a]*Faculty of Engineering, Shahrekord University, P.O.Box 115, Shahrekord, Iran* ;
[b]*Faculty of Engineering, Shahrekord University, P.O.Box 115, Shahrekord, Iran*;
[c]*Abadan Branch, Islamic Azad University, P.O.Box 666, Abadan, Iran.*

Abstract. In this paper, a fault diagnosis system based on discrete wavelet transform (DWT) and artificial neural networks (ANNs) is designed to diagnose different types of fault in gears and bearings. DWT is an advanced signal-processing technique for fault detection and identification. Five features of wavelet transform RMS, crest factor, kurtosis, standard deviation and skewness of discrete wavelet coefficients of normalized vibration signals has been selected. These features are considered as the feature vector for training purpose of the ANN. A wavelet selection criteria, Maximum Energy to Shannon Entropy ratio, is used to select an appropriate mother wavelet and discrete level, for feature extraction. To ameliorate the algorithm, various ANNs were exploited to optimize the algorithm so as to determine the best values for "number of neurons in hidden layer" resulted in a high-speed, meticulous three-layer ANN with a small-sized structure. The diagnosis success rate of this ANN was 100% for experimental data set. Some experimental set of data has been used to verify the effectiveness and accuracy of the proposed method. To develop this method in general fault diagnosis application, three different examples were investigated in cement industry. In first example a MLP network with well-formed and optimized structure (20 : 15 : 7) and remarkable accuracy was presented providing the capability to identify different faults of gears and bearings. In second example a neural network with optimized structure (20 : 15 : 4) was presented to identify different faults of bearings and in third example an optimized network (20:15:3) was presented to diagnose different faults of gears. The performance of the neural networks in learning, classifying and general fault diagnosis were found encouraging and can be concluded that neural networks have high potential in condition monitoring of the gears and bearings with various faults.

Received: 24 August 2013, Revised: 20 Janurary 2014, Accepted: 1 March 2014.

Keywords: Discrete wavelet transform, Artificial neural network, Fault diagnosis, Vibration analysis, Feature.

Index to information contained in this paper

1 **Introduction**
2 **Theory of Artificial Neural Networks (ANNs)**
3 **Data Acquisition Experiments**
4 **Feature Selection**
5 **Features Extraction and Faults Classification**
6 **General Diagnosis of Bearings and Gears**

*Corresponding author. Email: m.heidari1356@yahoo.com.

7 Conclusions

1. Introduction

Condition monitoring of machines is gaining importance in industry because of the need to increase reliability and to decrease possible loss of production due to machine breakdown. The use of vibration and acoustic emission (AE) signals is quite common in the field of condition monitoring of rotating machinery. The vibration monitoring of bearings and gearboxes due to their importance in industry and their vibration signal characteristics has been an interesting topic for researchers in this field. Therefore fault diagnostics and monitoring techniques for bearing and gearboxes have been improved in a short time frame.

Interests in automating the fault detection and diagnosis of machinery and reducing human errors have encouraged researchers to use soft computing methods. Artificial neural networks (ANNs) and fuzzy logic are used for identifying the machinery condition, while the genetic algorithm is used to optimize the monitoring system parameters. Fuzzy logic-based condition monitoring systems require expert's information of machinery faults and their symptoms. Wu and Hsu [21] designed a fuzzy logic- based fault diagnosis system for a gearbox system. However, these systems are fast and close to human inference rules and qualitative measurement techniques. On the other hand, monitoring systems based on ANNs do not require any background on the machinery characteristics and can be trained using a data set of machinery vibrations in different fault conditions.

Rafiee et al. [15] used a multiple- layer perceptron ANN to classify three different fault conditions and one no-fault condition of a gearbox. Also the genetic algorithm has been used as an effective tool for evolving monitoring systems and boosting their accuracy and speed of fault diagnosis process.

One of the most significant issues in intelligent monitoring is related to feature extraction. For this purpose different techniques of vibration analysis such as time, frequency and time-frequency domain are extensively used.

Samantha and Balushi [16] have presented a procedure for fault diagnosis of rolling element bearings through artificial neural network (ANN). The characteristic features of time-domain vibration signals of the rotating machinery with normal and defective bearings have used as inputs to the ANN. Yang et al. [24] have proposed a method of fault feature extraction for roller bearings based on intrinsic mode function (IMF) envelope spectrum. Fault diagnosis of turbo-pump rotor based on support vector machines with parameter optimization by artificial immunization algorithm has been done by Yuan and Chu [25]. Traditional techniques like Fast Fourier Transform (FFT) which are used for analysis of the vibration signals are not appropriate to analyze signals that have a transitory characteristic. Moreover, the analysis is greatly dependent on the machine load and correct identification of much closed fault frequency components requires a very high resolution data [4]. Wavelet transform (WT), a very powerful signal-processing tool can be used to analyze transients signal and thus eliminating load dependency and is capable of processing stationary and non-stationary signals in time and frequency domains simultaneously and can be used for feature extraction (Daubechies,1991). WT can be mainly divided into discrete (DWT) and continuous (CWT) forms. The discrete forms are faster with lower CPU time, but continuous forms generate an awful lot of data, so CWT has not been widely applied in the field of condition monitoring. Lei et al. [11] 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. The effectiveness of wavelet based features for fault diagnosis of gears using support vector machines (SVM) and proximal support vector machines (PSVM) has been revealed by Saravanan et al. [17]. Various artificial intelligence techniques have been used with wavelet transforms for fault detection in rotating machines [1, 5, 12, 14, 19, 22, 23].

Above researches focused on fault diagnosis only for a specific device. In the present study general fault diagnosis of gears and bearings has been investigated, therefore a multiple layer perceptron ANN was designed to classify seven different conditions of gears and bearings. The vibration signals acquired from the test-rig were first preprocessed using discrete wavelet transform and then ANNs were designed to classify different faults. Then the designed ANN was developed for general fault diagnosis of gears and bearings. Three different case studies were investigated in cement industry. The performance of designed ANNs in general fault diagnosis was found encouraging.

2. Theory of Artificial Neural Networks (ANNs)

An artificial neural network is a nonlinear mapping tool that relates a set of inputs to a set of outputs. It can learn this mapping using a set of training data and then generalize the obtained knowledge to a new set of data [7]. Today, ANNs have a variety of applications. As a classifier, one of the most commonly used ANNs is the Multi-Layer Perceptron (MLP) network. There are three types of layers in any MLP: the output layer, the input layer and the hidden layer. Each layer is comprised of n nodes ($n \geqslant 1$) and each node in any layer is connected to all the nodes in the neighboring layers. Each node can also be connected to a constant number which is called bias. These connections have their individual weights which are called synaptic weights and are multiplied to the node values of the previous layer. Input and output data dimensions of the ANN determine the number of nodes in the input and output layers, respectively, but the number of hidden layers and their nodes is determined heuristically. The number of hidden layers and nodes in an MLP is proportional to its classification power. However, there is an optimum number of hidden layers and nodes for each case and considering more than those amounts leads to over fitting of the classifier and made the computations substantially increased. The value of any node can be computed through Eq. (1):

$$\underline{a}^{l+1} = \underline{f}^{l+1}(W^{l+1}\underline{a}^{l} + \underline{b}^{l+1}) \tag{1}$$

Where $\underline{a}^l, \underline{b}^l andl$ are output vector, bias vector and layer number, respectively. W is the synaptic weights matrix of the MLP. $\underline{f}^l$ is the activation function of the lth layer and can be used to create nonlinear boundaries for the classifier. For example "sigmoid" is an activation function which can be used to bound the node values between 0 and 1. After setting the structure of the MLP ANN, it should be trained. Training an ANN means adjusting the synaptic weights in a way that any particular input leads to the desired output. It can be conducted by different algorithms. One of the most commonly used learning algorithms is resilient back propagation, which is used in this paper. For any learning algorithm, a limit should be defined to stop the learning process, which is called Stopping Criterion and usually consists of the following rules, or all of them simultaneously:

(a) The error root mean square in an epoch becomes less than a predefined value.

(b) Error gradient becomes less than a predefined value.

(c) The number of epochs reaches a predefined number.

The error vector for an MLP is defined as the difference between the network output vector and the desired output vector. Selecting an appropriate structure, initial weights, training algorithm for an MLP and supplying it with enough training datasets, enables the MLP to operate as a powerful classifier. In this study, it classifies the gears and bearings conditions into five faulty (two types of fault will be created on a gear and three types of fault will be created on a bearing) and two healthy conditions, according to the symptoms extracted from the measured vibration signals.

3. Data Acquisition Experiments

The experimental setup to collect dataset consists of a one-stage gearbox with spur gears, a fly wheel and an electrical motor with a constant nominal rotation speed of 1400 RPM. The fly wheel is supported through two roller bearings. Electrical motor, gearbox and fly wheel are attached together through flexible couplings as shown in figure 1.

Tables 1 and 2 depict gears and bearings specifications. Vibration signals were obtained in radial direction by mounting the accelerometer on the top of the gearbox and bearing housing. "Easy viber" data collector and its software, "SpectraPro", were used for data acquisition. Table 3 shows accelerometer probe specifications. The signals were sampled at 16000 Hz lasting 2 s.

In the present study, three pinion wheels whose details are as mentioned in Table 1 were used. One wheel was new one and was assumed to be free from defects. In the other two pinion wheels, defects were created. The raw vibration signals acquired from the gearbox when it is loaded with various pinion wheels discussed above. The vibration signal from the piezoelectric transducer (accelerometer) is captured for the following conditions: Good Spur Gear, Spur Gear with tooth breakage, and Spur Gear with face wear of the teeth.

For bearing vibration signal acquisition four Self-aligning ball bearings whose details are as mentioned in Table 2 were used. One new bearing was considered as good bearing. In the other three bearings, some defects were created and then various bearings were installed and the raw vibration signals acquired on the bearing housing. So the vibration signals were captured for the following conditions: Good Bearing, Bearing with spall on inner race, bearing with spall on outer race, bearing with spall on ball. Fig. 2 depicts some of the acquisition signals in time domain for various gears conditions.

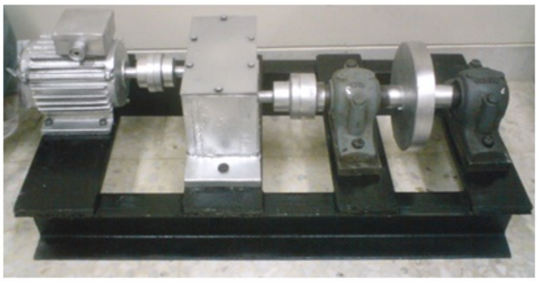


Fig.1. Fault simulator setup.

Table 1: Gear wheel and pinion details.

Parameters	Pinion wheel	Gear wheel
Outer diameter	63	93
No. of teeth	40	60
Module	1.5	1.5
Normal pressure angle	20°	20°
Top clearance	0.3 mm	0.3 mm
Material	C.K.45	C.K.45

Table 2: Bearing details.

Description	Self-aligning ball bearing
Designations Bearings with tapered bore	1209K
Outer diameter	85
Inner diameter	45(taper1:12 on diameter)

Table 3: Accelerometer probe characteristics.

Description	Multi-Purpose Accelerometer, Top Exit Connector / Cable, 100 mV/g
Sensitivity	100 mV/g
Frequency Response (±3dB)	30-900,000 CPM
Dynamic Range	±50 g, peak
Max Temp	121°C

4. Feature Selection

The process of computing some measures which represent a signal is called feature extraction. Wavelet based analysis is an exciting new problem solving tool for the mathematicians, scientists and engineers. It fits naturally with the digital computer with its bias functions defined by summations not integrals or derivatives. Unlike most traditional expansion systems, the basis functions of the wavelet analysis are not solutions of differential equations. In some areas, it is the first truly new tool we have had in many years. Indeed, use of wavelet transforms requires a new point of view and a new method of interpreting representations that we are still learning how to exploit.

In the early studies, Fourier analysis has been the dominating signal analysis tool for fault detection. But, there are some crucial restrictions of the Fourier transform. The signal to be analyzed must be strictly periodic or stationary; otherwise, the resulting Fourier spectrum will make little physical sense [13]. Unfortunately, gears and bearings vibration signals are often non-stationary and represent non-linear processes, and their frequency components will change with time. Therefore, the Fourier transform often can not full fill the gears and bearings fault diagnosis task pretty well. On the other hand, the time-frequency analysis methods can generate both time and frequency information of a signal simultaneously through mapping the one-dimensional signal to a two dimensional time-frequency plane. Among all available time-frequency analysis methods, the wavelet transforms may be the best one and have been widely used for gears and bearings fault detection [20].

4.1 *Theoretical Background of Wavelet Transform*

The wavelet transform (WT) is a time-frequency decomposition of a signal into a set of 'wavelet' basis functions. In this section, we review the continuous wavelet transform (CWT) and the discrete wavelet transform (DWT). Figure 2 shows sample signals in time domain for various gears conditions.

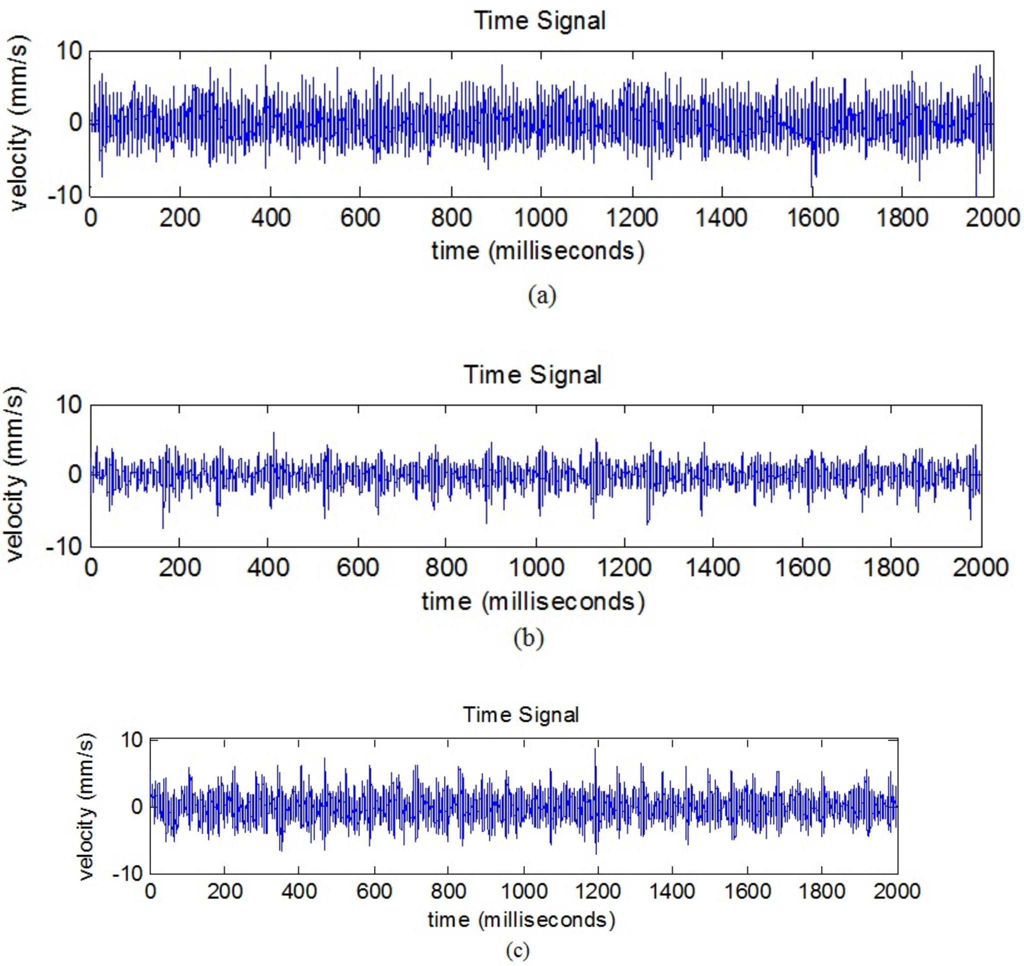


Fig. 2 Sample signals in time domain for various gears conditions. (a) Good Spur Gear, (b) Spur Gear with tooth breakage, (c) Spur Gear with face wear of the teeth

4.1.1 Continuous wavelet transform (CWT)

The continuous wavelet transform of a time function f(t) is given by the equation:

$$T(a,b) = \int_{-\infty}^{+\infty} f(t)\psi^*_{(a,b)}(t)dt \tag{2}$$

where * denotes complex conjugation and

$$\psi_{(a,b)}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a})(a,b \in R, a \neq 0) \tag{3}$$

is a member of the wavelet basis, derived from the basic analysis wavelet $\psi(t)$ through translation and dilation. As seen in Eq. (3), the transformed signal T(a,b) is defined on the a-b plane, where a and b are used to adjust the frequency and the time location of the wavelet in Eq. (3). A small a produces a high-frequency (contracted) wavelet when high-frequency resolution is needed. The WT's superior time- localization properties stem form the finite support of the analysis wavelet: as b increases, the analysis wavelet transverses the length of the input signal, and an increases or decreases in response to changes in the signal's local time and frequency content. Finite support implies that the effect of each term in the wavelet representation is purely localized. This sets the WT apart from the Fourier Transform, where the effects of adding higher frequency sine waves are spread throughout the frequency axis.

4.1.2 Discrete wavelet transform (DWT)

Discrete methods are required for computerized implementation of the WT. The DWT is derived from the CWT through discretization of the wavelet $\psi_{(a,b)}(t)$. The most common discretization of the wavelet is the dyadic discretization, given by:

$$\psi_{(j,k)}(t) = \frac{1}{\sqrt{2^j}}\psi(\frac{t-2^jk}{2^j}) \tag{4}$$

Where a has been replaced by 2^j and b by 2^jk[19-20]. Under suitable conditions Eq. (4) is an orthonormal basis of L2R, and the original time function can be expressed as:

$$f(t) = \sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} C_(j,k)\psi_{(j,k)}(t) \tag{5}$$

$$c_(j,k) = \int_{-\infty}^{+\infty} f(t)\psi^*_{(j,k)}(t)dt \tag{6}$$

Where $c_{j,k}$ are referred to as wavelet coefficients. A second set of basis function $\phi(t)$ called scaling function is then obtained by applying multi-resolution approximations to obtain the orthonormal basis of $\psi(t)$:

$$\phi_{(j,k)}(t) = \frac{1}{\sqrt{2^j}}\phi(\frac{t-2^jk}{2^j}) \tag{7}$$

The original time function can now be written as:

$$d_(j,k) = \int_{-\infty}^{+\infty} f(t)\phi^*_{(j,k)}(t)dt \tag{8}$$

Here, the $d_{j,k}$, which are called the scaling coefficients, $d_{0,k}$is the sampled version of f(t), represent a jth order resolution discretization of $f(t)$. The scaling coefficients and the wavelet coefficients for resolutions of order greater than j can be obtained iteratively by

$$d_{j+1,k} = \sum_{i=-\infty}^{\infty} h(i-2k)d_{j,k} \tag{9}$$

$$c_{j+1,k} = \sum_{i=-\infty}^{\infty} g(i-2k)c_{j,k} \tag{10}$$

The sequences h and g are low-pass and high-pass filters derived from the original analyzing wavelet $\psi(t)$. The scaling coefficients $d_{j,k}$ represent the lower frequency approximations of the original signal, and the wavelet coefficients $c_{j,k}$ represent the distribution of successively higher frequencies. The inverse DWT yields a difference series representation for the input signal $d_{0,k}$ in terms of the filters h and g and the wavelet coefficients $c_{j,k}$:

$$d_{j,k} = \sum_{i=-\infty}^{\infty} h(k-2)d_{j+1,i} + \sum_{i=-\infty}^{\infty} g(k-2i)c_{j+1,i} \tag{11}$$

The wavelet filters adopted determine the quality of the wavelet analysis. For example the for Daubechies wavelets of length 2:

$$h(0) = \frac{1}{\sqrt[n]{2}}\ ,\ h(1) = \frac{1}{\sqrt[n]{2}}\ ,\ g(0) = h(1)\ ,\ g(1) = -h(0) \tag{12}$$

Since the input signal f(t) is discretized into N samples, Eqs. (9) and (10) can be written in the form of matrix:

$$\begin{pmatrix} d_{1,1} \\ c_{1,1} \\ d_{1,2} \\ c_{1,2} \\ \vdots \\ d_{1,N/2} \\ d_{1,N/2} \end{pmatrix} = TN \begin{pmatrix} d_{0,1} \\ c_{0,2} \\ d_{0,3} \\ c_{0,4} \\ \vdots \\ d_{0,N-1} \\ d_{0,N} \end{pmatrix}$$

$Where$

$$TN = \begin{pmatrix} h(0) & h(1) & 0 & \dots & \square \\ g(0) & g(1) & 0 & \dots & \square \\ 0 & h(0) & h(1) & 0 & \square \\ 0 & g(0) & g(1) & 0 & \square \\ \square & \square & 0 & \vdots & \square \\ \vdots & \vdots & 0 & h(0) & h(1) \\ \square & \square & 0 & g(0) & g(1) \\ h(1) & 0 & \vdots & 0 & h(0) \\ g(1) & 0 & \square & 0 & g(0) \end{pmatrix} \tag{13}$$

The scaling coefficients$d_{j+1,k}(k = 1 \sim N/2^{j+1})$and the wavelet coefficients$C_{j+1,k}$ of the $(j+1)$th order resolution can be obtained by applying the$N/2^j \times N/2^j$ matrix $T_{N/2^j}$ to the scaling coefficients of the jth order $d_{j,k}(k = 1 \sim N/2^j)$. When the number of data points is $N = 2^i$, all of the wavelet coefficients are obtained after $(i-1)$ iterations of Eq. (13). The inverse DWT is performed in a similar manner by straight forward inversion of the orthogonal matrix T_N.
The wavelet analysis has the advantage of better performance for non-stationary signals, representing a time signal in terms of a set of wavelets. They are constituted for a family of functions which are derived from a single generating function called mother wavelet, although dilation and translation processes. Dilation is related with size, and it is also known as scale parameter while translation is the position variation of the selected wavelet along the time axis. This process is illustrated in figure 3.

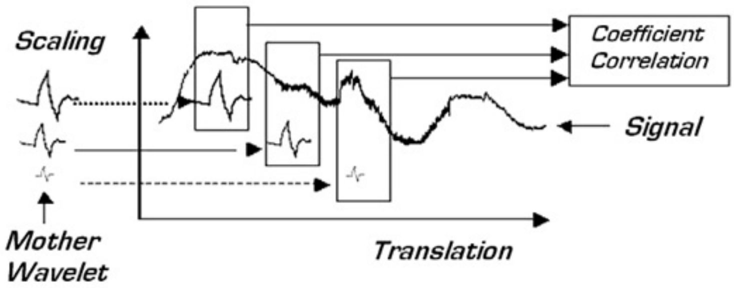


Fig. 3.Wavelet transform execution

For gears and bearings fault detection, the frequency ranges of the vibration signals that are to be analyzed are often rather wide; and according to the Shannon sampling theorem, a high sampling speed is needed, and sequentially, large size samples are needed for the gears and bearings fault detection. Therefore, it is expected that the desired method should have good computing efficiency. Unfortunately, the computing of continuous wavelet transform (CWT) is somewhat time consuming and is not suitable for large size data analysis and on-line fault diagnosis. The discrete wavelet transform (DWT), which is based on sub-band coding, is found to yield a fast computation of Wavelet Transforms. It is easy to implement and reduces the computation time and resources required. Hence, it is taken up for this study.

4.2 *Mother WaveletSelection*

One of the most significant issues in wavelet transform is related to mother wavelet selection. For this purpose researchers have used various methods, such as the genetic algorithm [15], decision tree [18], etc. Entropy is a common concept in many fields, mainly in signal processing [26]. In the present study, the "shannon entropy" will be used in various fault conditions after data preprocessing of wavelet transform and a wavelet selection criteria are used to select an appropriate mother wavelet for feature extraction.

4.2.1 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 [3, 8] and is given as

$$\xi(n) = \frac{E(n)}{S_{entropy}(n)} \tag{14}$$

Where the energy at each resolution level n is given by

$$E(n) = \sum_{i=1}^{m} |C_{n,i}|^2 \tag{15}$$

The total energy can be obtained by

$$E_{total} = -\sum_{n}\sum_{i}^{\square}|C_{n,i}|^2 = \sum_{n}^{\square}E(n) \tag{16}$$

Where $'m'$ is the number of wavelet coefficients and $C_{n,i}$ is the ith wavelet coefficient of nth scale. Entropy of signal wavelet coefficients is given by

$$S_{entropy}(n) = -\sum_{i=1}^{m} p_i log_z p_i \tag{17}$$

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

$$p_i = \frac{|C_{n,i}|^2}{E(n)} \tag{18}$$

With $\sum_{i=1}^{m} p_i = 1$, and in the case of $p_i = 0$ for some i, the value of $p_i log_z p_i$ is taken as zero. The following steps explain the methodology for selecting a base wavelet for the vibration signals under study:

1. In this study, Good Spur Gear, Spur Gear with tooth breakage, Spur Gear with face wear of the teeth, good bearings, bearings with spall in outer race, inner race and ball defects are considered. Total 98 vibration signals in time domain are obtained in vertical directions for each gear and bearing conditions.

2. For healthy and faulty bearings, discrete wavelet coefficients (DWT) of vibration signals are calculated using 36 different mother wavelets. Haar, Daubechies ($db2 \sim db10$), Symlet ($sym2 \sim sym11$), Coiflet ($coif1 \sim coif5$), Bi-orthogonal (bior1.1, bior1.2, bior3.3, bior3.1, bior2.8, bior2.6, bior2.4, bior2.2, bior1.5, bior1.3, bior3.5), where discrete approximation of Meyer was selected.

3. Wavelet selection criterion was used to select an appropriate mother wavelet using Energy to shannon entropy ratio as:

The total energy and total shannon entropy of DWT in third and fourth decomposition levels are calculated for vibration signals at different conditions using healthy and faulty gears and bearings. The total energy to total shannon entropy ratio for each wavelet is calculated as shown in figure 4.

Compare of total energy to total shannon entropy ratio for different mother wavelets shows little change from third to fourth levels as shown in figure 5 for Daubechies mother wavelets. Therefore, third decomposition level was determined to be the most appropriate level for this case study.

4. The wavelet having maximum energy to shannon entropy ratio is considered for fault diagnosis of gears and bearings.

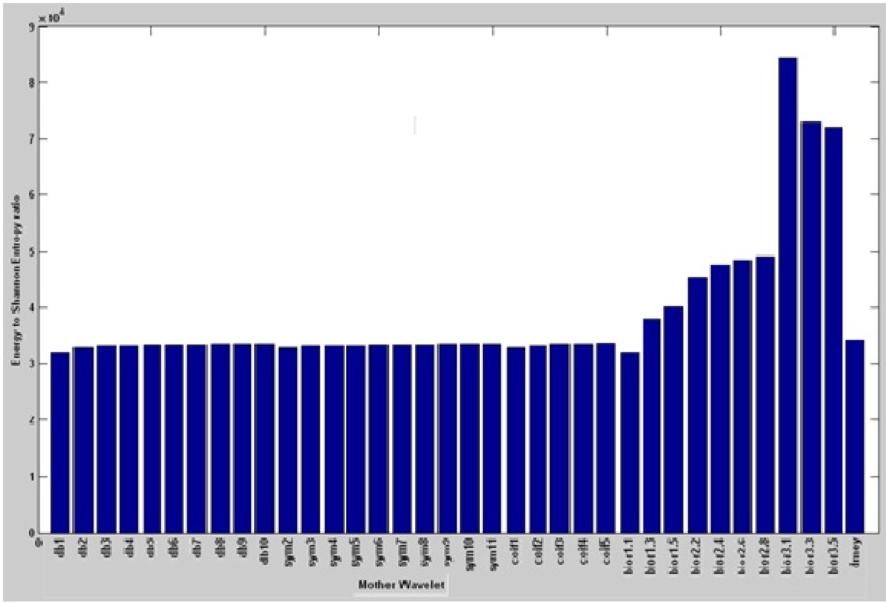


Fig 4 Total energy to total shannon entropy ratio for 36 mother wavelet

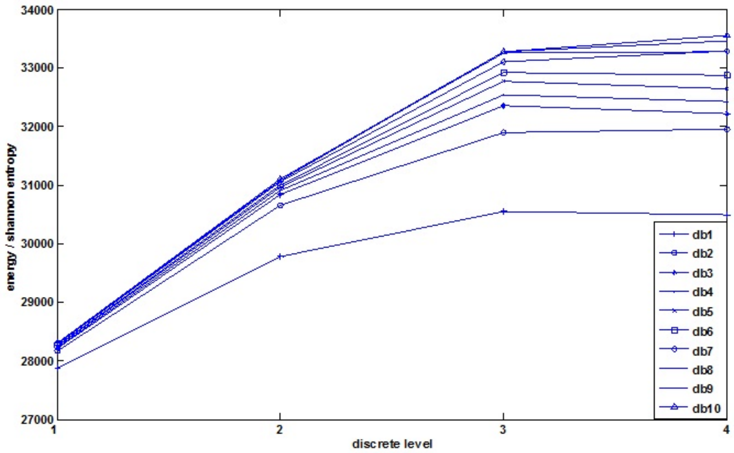


Fig. 5. Total Energy to Total Shannon Entropy ratio for 1~4 decomposition levels of Daubechies mother wavelets

5. Features Extraction and Faults Classification

Based on wavelet selection criteria, Bi-orthogonal (bior3.1) wavelet is selected as the best base wavelet among the other wavelets considered. The vibration signal associated with various conditions of gears and bearings with different gears and bearings explained in Section 3 have been decomposed using "bior3.1" wavelet. The approximated and detailed coefficients for various conditions of gears with various fault gears are shown in figure 6. From figures 6a-c, the signal 's' represents the actual vibration signal where as 'a3' represents the approximation at level 3 of bior3.1 wavelet and 'd1' to 'd3' represents the coefficients details at level 1 3, respectively. The wavelet tree representation of the vibration signals gives a clear idea about how the original signal is reconstructed using the approximations and

details at various levels. The wavelet tree representation of the good pinion wheel is shown in figure 6e. The coefficients obtained using this wavelet transforms are further subjected to statistical analysis and the statistical features are extracted for all the approximation and details coefficients of DWT. Root mean square (RMS) value, crest factor, kurtosis, skewness, standard deviation, mean, shape factor, etc., are most commonly used statistical measures used for fault diagnosis of gears and bearings [6, 8, 18]. Statistical moments like kurtosis, skewness and standard deviation are descriptors of the shape of the amplitude distribution of vibration data collected from a bearing or gear. Therefore in present paper, author's use RMS, crest factor and statistical moments like kurtosis, skewness and standard deviation as features to effectively indicate early faults occurring in gears and bearings. These features are briefly described as follows:
RMS: is a statistical measure of the magnitude of a varying quantity.

$$RMS = \sqrt{\frac{1}{n}\sum_{1}^{n} x_i^2} \tag{19}$$

Crest Factor: The crest factor or peak-to-average ratio (PAR) is a measurement of a waveform, calculated from the peak amplitude of the waveform divided by the RMS value of the waveform.

$$CrestFactor = \frac{Peaklevel}{RMS} \tag{20}$$

Standard deviation: Standard deviation is measure of energy content in the vibration signal.

$$s = \sqrt{\frac{n\sum_1^n x_i^2 - (\sum_1^n x_i)^2}{n(n-1)}} \tag{21}$$

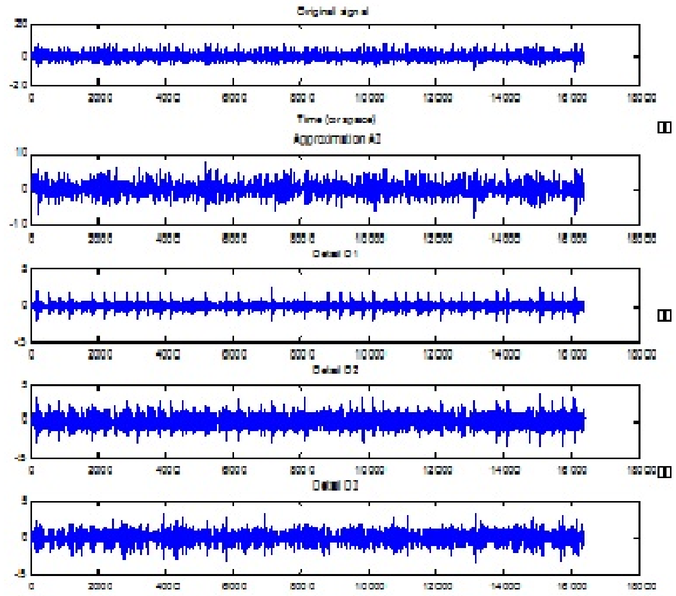
Kurtosis: A statistical measure used to describe the distribution of observed data around the mean. Kurtosis is defined as the degree to which a statistical frequency curve is peaked.

$$Kurtosis = \{\frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i}^{n}(\frac{x_i - \overline{x}}{s})^4 - \frac{3(n-1)^2}{(n-2)(n-3)}\} \tag{22}$$

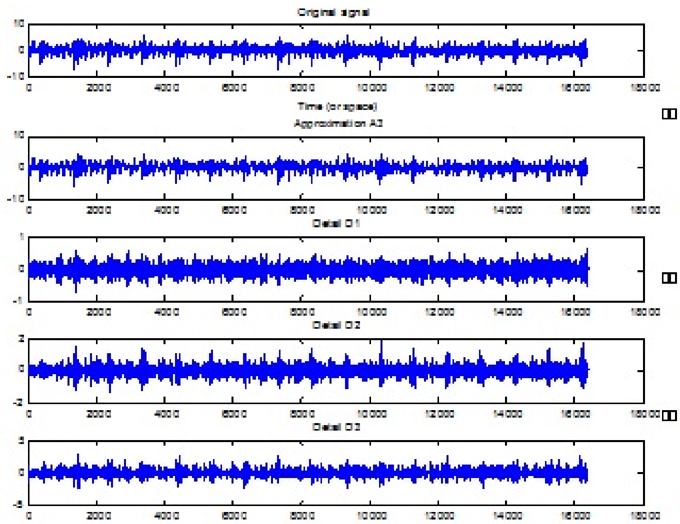
Skewness: Skewness characterizes the degree of asymmetry of distribution around its mean. Skewness can be negative or positive.

$$Skewness = \frac{n}{(n-1)(n-2)}\sum_{i}^{n}(\frac{x_i - \overline{x}}{s})^3 \tag{23}$$

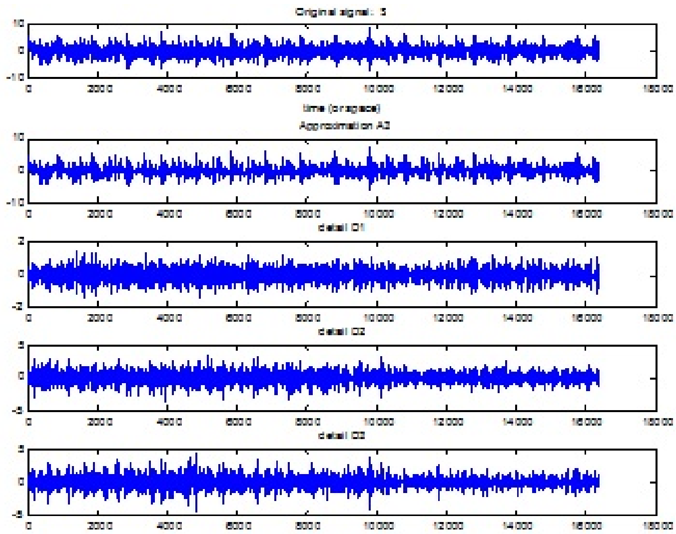
In the above equations x_i is vibration signal data, $\overline{x}$ is mean of vibration signal data and n is the number of vibration signal data. These statistical features are fed as input to ANN, for faults classification.


Original signal
Time (or space)
Approximation A3
Detail D1
Detail D2
Detail D3

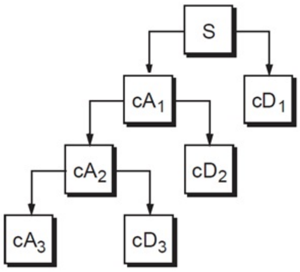
(a)


Original signal
Time (or space)
Approximation A3
Detail D1
Detail D2
Detail D3

(b)



(c)



(e)

Fig.6. Actual vibration signal (S), approximation (A3) and details coefficients (D1~D3) of wavelet decomposition (level 3 of bior3.1 wavelet). (a) Good Spur Gear, (b) Spur Gear with tooth breakage, (c) Spur Gear with face wear of the teeth (e) Wavelet tree.

6. General Diagnosis of Bearings and Gears

To develop this method in the general diagnosis of bearings and gears a computer program called "fault detector" was provided. An image of this program depicted in figure 7. Some of the program menus are briefly described as follow:

"Identify new machine" and "Select existent machine" menus are used to define a machine and train its corresponding network.

"Load signal" menu is used to upload a signal in time domain

"DWT" menu is used to apply discrete wavelet transform on the signal and extract feature vector

"Analyze" menu depicts result of fault diagnosis by applying feature vector to ANN.

In gears and bearings industrial applications three categories of machines are used: machines with gears, machines with bearings and machines with both of gears and bearings. For fault diagnosis of each category ANN should only cover corresponding faults. For example ANN of machine with gear only covers Good Gear, Gear with tooth breakage, and Gear with face wear of the teeth. The performance of the proposed method for each of these categories is evaluated using three practical examples in cement industry.

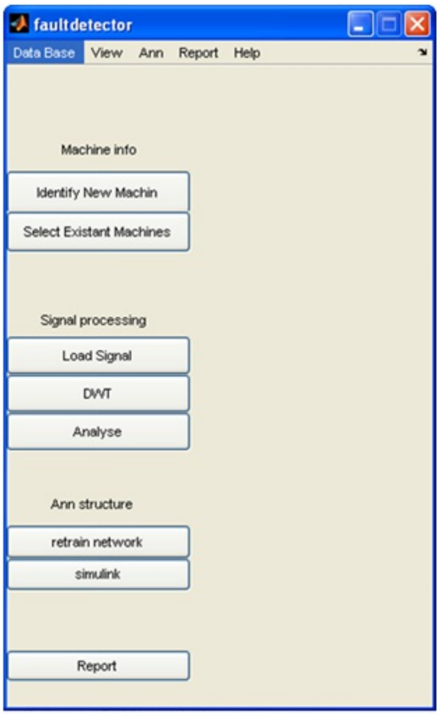


Fig.7 "fault detector" program

6.1 *Fault Diagnosis of Test-Rig Set*

In first step we consider the fault diagnosis of test rig set that we have used for data acquisition and ANN training.

6.1.1 Application of ANN for problem at hand and results

For each faults namely, Good Spur Gear, Spur Gear with tooth breakage, Spur Gear with face wear of the teeth Good Bearing, Bearing with spall on inner race, bearing with spall on outer race and bearing with spall on ball, 10 feature vectors consisting of 20 feature values as mentioned before were calculated from the experimental vibration signals (Sec. 3). Five samples in each class were used for training and five samples are reserved for testing ANN. Training was done by selecting three layers neural network, of that one is input layer, one hidden layer and one output layer.

6.1.2 Results of ANN

The architecture of the artificial neural network is as follows:
Network type: Forward neural network trained with feedback propagation
No of neurons in input layer: 20
No of neurons in hidden layer: Varied from 6 to 21
No of neurons in output layer: 7
Transfer function: Tangent-sigmoid transfer function in hidden layer and sigmoid in output layer
Training rule: Back propagation

For hidden layer the necessary and sufficient number of neurons must be selected. One of the problems that occur during neural network training is over fitting. The error on the training set is driven to a very small value, but when new data is presented to the network the error is large. The network has memorized the training examples, but it has not learned to generalize to new situations. One

Table 4: Results of ANN classifiers with different number of hidden-layer neurons for various gears and bearing conditions

Machine condition	No. of Samples for training	No. of Samples for testing	No. of hidden-layer neurons											
			6		9		12		15		18		21	
			Correct diagnosis	Misdiagnosis	Correct diagnosis	Misdiagnosis	Correct diagnosis	Misdiagnosis	Correct diagnosis	Misdiagnosis	Correct diagnosis	Misdiagnosis	Correct diagnosis	Misdiagnosis
Good Gear	5	5	5	0	5	0	5	0	5	0	5	0	5	0
Gear with tooth breakage	5	5	5	0	5	0	5	0	5	0	5	0	5	0
Gear with wear of the teeth	5	5	4	1	4	1	5	0	5	0	5	0	5	0
Good Bearing	5	5	5	0	5	0	5	0	5	0	5	0	5	0
Bearing with spall on inner race	5	5	4	1	4	1	5	0	5	0	4	1	5	0
bearing with spall on outer race	5	5	5	0	5	0	5	0	5	0	5	0	5	0
bearing with spall on ball	5	5	4	1	4	1	4	1	5	0	5	0	4	1
Total	35	35	32	3	32	3	34	1	35	0	34	1	34	1

Table 5: Total performance of ANN classifiers with different number of hidden-layer neurons

	No. of hidden-layer neurons					
	6	9	12	15	18	21
Results of correct fault diagnosis%	91.4%	91.4%	97.1%	100%	97.1%	97.1%

method for improving network generalization is to use a network that is just large enough to provide an adequate fit. In this study the numbers of neurons in the hidden layer were selected by trial and error. A total of six networks with different number of hidden layers with characteristics mentioned above were created for classifying the faults. The training was done with 35 data set attributes and the cross validation was done using 35 data sets. The efficiency of classification of gears and bearings faults using above networks has been reported in tables 4 and 5. At first, as shown in these tables, increase of neurons in the hidden layer improves the efficiency of classification. Number of neurons in the hidden layer is optimized in 15 neurons and greater number of hidden layer neurons (18 or 21) will not affect on efficiency. Therefore an artificial neural network with 20:15:7 layers was utilized for fault diagnosis. The overall average efficiency of entire classification using ANN was found to be 100%.

6.2 *Fault Diagnosis of a Gearbox*

In this example vibration signal of a three axes gearbox is surveyed. This gearbox is used to rotate a ball mill in cement industry. A section of the gearbox depicted in figure 8. Six vibration signals were collected in 6 points (1, 2, 3, 4, 5 and 6 in figure 8).

In the next step, the neural network which designed for test rig set was used for ball mill gearbox trouble shooting. For this purpose six feature vectors were extracted from six vibration signals. As mentioned above these signals were collected in 6 points of ball mill gearbox. The feature vectors were applied to the neural network, respectively. Results of fault diagnosis were as follow:
Point 1: face wear of the teeth
Point 2: face wear of the teeth
Point 3: safe gear
Point 4: safe gear
Point 5: safe gear
Point 6: safe gear
These findings indicate the teeth wear on the first axis of gearbox. The accuracy of the results was confirmed after gearbox inspection.

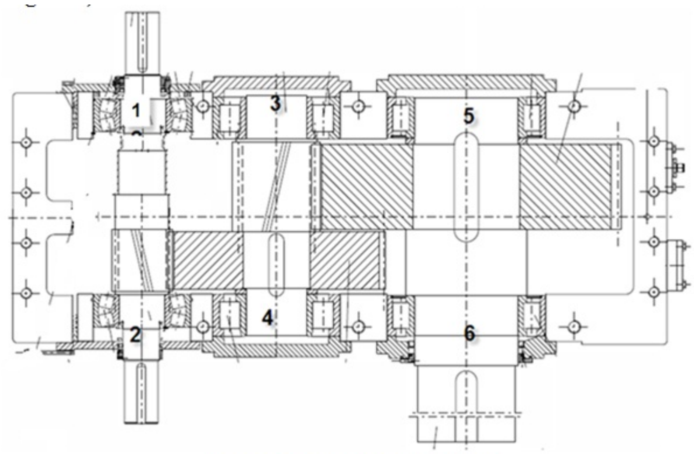


Fig. 8. Section of the gearbox

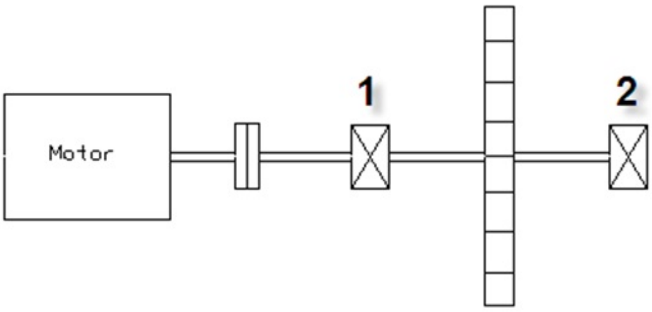


Fig.9 schematic of the ID fan

6.3 *Fault Diagnosis on Bearings*

In this example vibration signal of a fan is surveyed. This fan is used as ID FAN in cement industry. Figure 9 shows a schematic diagram of the fan. Vibration signal was collected in point 1 because of temperature rise in this point.

6.3.1 Application of ANN for this problem

For each faults namely, Good Bearing, Bearing with spall on inner race, bearing with spall on outer race and bearing with spall on ball, 10 feature vectors consisting of 20 feature values were calculated from the experimental vibration signals (Sec.3). 5 samples in each class were used for training and 5 samples are reserved for testing ANN. Training was done by selecting three layers neural network, of that one is input layer, one hidden layer and one output layer.

6.3.2 Results of ANN for bearing fault diagnosis

The architecture of the artificial neural network is as follows:
Network type: Forward neural network trained with feedback propagation
No of neurons in input layer: 20
No of neurons in hidden layer: 15
No of neurons in output layer: 4
Transfer function: Tangent-sigmoid transfer function in hidden layer

Table 6: Girth gear and pinion characteristics.

	Speed(RPM)	Teeth no.	Module	Outer diameter
Girth gear	15	238	30	7200mm
Pinion	115	31	30	990mm

and sigmoid in output layer
Training rule: Back propagation

An optimization was done like first example. Eventually an artificial neural network with 20:15:4 layers was utilized for fault diagnosis. Then the neural network was used for ID FAN bearing trouble shooting. For this purpose the feature vector was extracted from vibration signal and applied to the neural network. Result of fault diagnosis was as follow:
Point 1: spall on ball of the bearing
The accuracy of the result was confirmed after bearing inspection.

6.4 *Fault Diagnosis on Girth Gear*

In this example vibration signal of a girth gear is checked. This girth gear is used to rotate a ball mill in cement industry. Table 6 shows characteristics of this large gear. A vibration signal was collected on journal bearing of its pinion.

6.4.1 Application of ANN

For each faults namely, Good Spur Gear, Spur Gear with tooth breakage, Spur Gear with face wear of the teeth,10 feature vectors consisting of 20 feature values were calculated from the experimental vibration signals (Sec.3). five samples in each class were used for training and 5 samples are reserved for testing ANN. Training was done by selecting three layers neural network, of that one is input layer, one hidden layer and one output layer.

6.4.2 Results of ANN for fault diagnosis

The architecture of the artificial neural network is the same as first and second examples; only number of neurons in output layer was selected 3.ANN optimization was done like first example. Eventually an artificial neural network with 20:15:3 layers was utilized for fault diagnosis. Then this neural network was used for girth gear trouble shooting. For this purpose the feature vector was extracted from vibration signal and applied to the neural network. Diagnostic results indicated breakage and wear of the teeth. The accuracy of the result was confirmed after girth gear inspection.

7. Conclusions

This paper has outlined the definition of the discrete wavelets transform and then demonstrated how it can be applied to the analysis of the vibration signals produced by gears and bearings in various conditions and faults. A wavelet selection criterion "Maximum Energy to Shannon Entropy ratio" was used to select an appropriate wavelet and Bi-orthogonal wavelet (bior3.1) was selected for feature extraction. Five statistical features (Root mean square (RMS) value, crest factor, kurtosis, skewness, standard deviation) are extracted for all the approximation and details coefficients of DWT. These features were fed as input to neural network for classification of various faults of the gears and bearings. Three different examples were investigated in cement industry. In first example a MLP network with well-formed and optimized structure (20:15:7) and remarkable accuracy was presented providing the capability to identify different faults of gears and bearings.

In second example a neural network with optimized structure (20:15:4) was presented providing the capability to identify different faults of bearings and in third example an optimized network (20:15:3) was presented to identify different faults of gears. The performance of the neural network in learning, classifying and general fault diagnosis was found encouraging and can be concluded that neural networks and wavelet transform have high potential in condition monitoring of the gears and bearings with various faults.

References

[1] S. Abbasion, A. Rafsanjani, A. Farshidianfar and N. Irani., *Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine. Mechanical Systems and Signal Processing* **21** (2007) 2933-2945.

[2] F. Arvani, A. Harifi and M.H. Sadeghi, *Intelligent condition monitoring of a gearbox using artificial neural network. Mechanical Systems and Signal Processing* **21** (2007) 1746-1754.

[3] M. Barakat, F. Druaux, D. Lefebvre, M. Khalil and O. Mustapha, *Self adaptive growing neural network classifier for faults detection and diagnosis, Neurocomputing,* **74** (2011) 3865-3876.

[4] W. R. Finley, M. M. Hodowance and W.G. Holter, *An analytical approach to solving motor vibration problems. IEEE Transactions on Industry Applications* 36 (5) (2000) 1467-1480

[5] X. Z. Gao and S. J. Ovaska, *Soft computing methods in motor fault diagnosis. Applied Soft Computing* **1** (2001) 73-81

[6] A. Hajnayeb, A. Ghasemloonia, S. E. Khadem and M. H. Moradi, *Application and comparison of an ANN-based feature selection method and the genetic algorithm in gearbox fault diagnosis, Expert Systems with Applications,* **38** (2011) 10205-10209.

[7] S. Haykin, *Neural networks: A comprehensive foundation (2nd ed.). Pearson Education*1998.

[8] P. K. Kankar, C. S. Satish and S. P. Harsha, *Fault diagnosis of ball bearings using continuous wavelet transform. Applied Soft Computing* **11** (2011) 2300-2312.

[9] S. Mallat *A Wavelet Tour of Signal Processing* . Academic Press, New York, 1998.

[10] K. Mori, N. Kasashima, T. Yoshioka and Y. Ueno, *Prediction of spalling on ball bearing by applying the discrete wavelet transform to vibration signals. Wear* **195** (1996) 162-168.

[11] Y. Lei, Z. He and Y. Zi, *Application of an intelligent classification method to mechanical fault diagnosis. Expert Systems with Applications* **36** (2009) 9941-9948

[12] B. A. Paya, I. I. Esat and M. N. M. Badi, *Artificial neural networks based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor. Mechanical Systems and Signal Processing* **11** (1997) 751-765.

[13] Z. K. Peng, F. L. Chu, *Application of the Wavelet transform in machine condition monitoring and fault diagnostics: A review with bibliography. Mechanical Systems and Signal Processing* **18** (2004) 199-221.

[14] S. Prabhakar, A. R. Mohanty and A. S. Sekhar, *Application of discrete wavelet transform for detection of ball bearing race faults. Tribology International* **35** (2002) 793-800.

[15] J. Rafiee, P. W. Tse, A. Harifi and M. H. Sadeghi, *A novel technique for selecting mother wavelet function using an intelligent fault diagnosis. Expert Systems with Applications* **37** (2010) 4168-4181.

[16] B. Samantha and K. R. Al-Balushi, *Artificial Neural Networks based fault diagnostics of rolling element bearings using time domain features. Mechanical Systems and Signal Processing* **17** (2003) 317-328.

[17] N. Saravanan, V. N. S. Kumar and K. I. Ramachandran, *A comparative study on classification of features by SVM and PSVM extracted using Morlet Wavelet for fault diagnosis. Expert Systems with Applications* **35** (2008) 1351-1366.

[18] N. Saravanan and K. I. Ramachandran, *, Incipient gearbox fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN). Expert Systems with Applications* **37** (2010) 4168-4181.

[19] W.J. Staszewski, G.R. Tomlinson, *, Application of the wavelet transforms to fault detection in spur gear. Mechanical Systems and Signal Processing*8(1994) 289-307

[20] W. J. Wang, *Applications of wavelets to gearbox vibration signals for fault detection. Journal of Sound and Vibration* **192 (5)** (1996) 927-939.

[21] J. D. Wu and C. C. Hsu, *Fault gear identification using vibration signal with discrete wavelet transform technique and fuzzy-logic inference. Expert Systems with Applications* **36 (2)** (2009) 3785-3794.

[22] G. Xian and B. Zeng, *An intelligent fault diagnosis method based on wavelet packet analysis and hybrid support vector machine. Expert Systems with applications* **22** (2009) 1048-1060

[23] R. Yan, *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.

[24] Y. Yang, Y. Dejie, J. Cheng, *A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM. Measurement* **40** (2007) 943-950.

[25] S. Yuan, F. Chu, *Fault diagnosis based on support vector machines with parameter optimization by artificial immunization algorithm. Mechanical Systems and Signal Processing* **21** (2007) 1318-1330.

[26] J. Zhang, G. G. Walter, Y. Miao and W. N. W. Lee, *CWavelet neural networks for function learning. IEEE Transactions on Signal Processing* **43** (1995) 1485-1497.