ANN BASED FAULT DIAGNOSIS OF ROLLING ELEMENT BEARING USING TIME-FREQUENCY DOMAIN FEATURE

D.H. PANDYA, S.H. UPADHYAY, S.P. HARSHA

Mechanical & Industrial Engineering Department
Indian Institute of Technology, ROORKEE
Roorkee-247667, INDIA
veddhrumi@gmail.com, shumefme@iitr.ernet.in, spharsha@gmail.com

Abstract :

This paper presents a methodology for an automation of fault diagnosis of ball bearings having localized defects (spalls) on the various bearing components. The system uses the wavelet packet decomposition using ‘rbio5.5’ real mother wavelet function for feature extraction from the vibration signal, recorded for various bearing fault conditions. The decomposition level is determined by the sampling frequency and characteristic defect frequency. Maximum energy to minimum Shannon entropy ratio criteria is used for selection of best node of wavelet packet tree. The two features kurtosis and energy are extracted from the wavelet packet coefficient for selected node of WPT. The total 10 data sets at five different speeds corresponding to each bearing condition are recorded for fault classification. Thus, extracted features are used to train and test neural network with multi layer perceptron to classify the rolling element bearing condition as HB, ORD, IRD, BD and CD. The proposed artificial neural network with multi layer perceptron classifier has overall fault classification rate of 97 %.

Keywords: wavelet packet transform; artificial neural network; rolling element bearing.

1. Introduction

Machine condition monitoring, early fault detection, diagnosis and classification are extremely important topics in the engineering field. Bearings are the most widely used mechanical parts in rotational equipment and are primary cause of breakdowns in machines. Such malfunctions can lead to costly shutdowns, lapses in production, and even human casualties. Fault diagnosis of rolling element bearings using vibration signature analysis is the most commonly used to prevent breakdowns in machinery. To analyze vibration signals, several methods in different domains have been implemented such as time domain, frequency domain and time-frequency domain [McFadden(1985)].

In time-domain, the interpretation of the signal is done through several parameters. Some of the parameters are: RMS, Crest factor, peak, probability density function, second, third and fourth order statistical moments, which can be extracted from vibration signal [Lin and Qu (2000)]. Whereas, in frequency domain based analysis Fourier transformations are employed to transform time domain signals into frequency domain. Further analysis is done on vibration amplitude and power spectrum. The key point in both the analysis techniques is that the direct use of informational content in one domain is excluded when the other domain is employed [Li (1998)]]. A major revolution in the signal processing techniques is brought by the introduction of time-frequency analysis as it is capable of revealing aspects of data that other signal analysis techniques could not. One major advantage afforded by wavelets is the ability to perform analysis for non stationary signals. The major disadvantages of Fourier analysis considered here can be cited as information loss and difficulty in interpreting the signals when moving from time domain to frequency domain, particularly in non stationary signals [Sugumaran *et al.*(2007)]. The Hertzian contact stresses between the rolling elements and the races are one of the basic mechanisms that initiate a localized defect. An impulse is produced, when a rolling element strikes a localized defect, which excites the resonances of the bearing structure [Harsha *et al.*(2007)].

[Nikolaou and Antoniadis (2002)] have used wavelet packet transform to identify the nature of rolling element bearing faults. The wavelet packet transform is used for the analysis of vibration signals resulting from bearings with localized defect. [Samantha and Al-Balushi (2003)] 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 been used as inputs to the ANN [Liu and Mengel (1991)].

Previous research articles have highlighted the advantages of wavelet transforms when applied to fault diagnosis. Fault diagnosis is a type of classification problem, and artificial intelligence techniques based classifiers can be effectively used to classify normal and faulty machine conditions. A machine fault classification problem consists of two main steps. First step is feature extraction from raw vibration signals to extract some features that demonstrate the information of fault from the raw signals and second step is to use these extracted features for fault diagnosis using various artificial intelligence techniques like artificial neural networks, support vector machines, etc.

2. Wavelet Packet Transform

Wavelet packet transform is a generalization of the wavelet transform. Let us define two functions $W_0(t)=Ø(t)$, $W_1(t)=\psi(t)$ where $Ø(t)$ and $\psi(t)$ are the scaling and translating wavelet functions respectively. Then in an orthogonal case we can write functions $W_n(t)$, n=0,1,2…. and k=0,1,2…

$$W_{2n}^{j}(t)=\sqrt{2}\sum_{k}h(k)W_{n}^{j}(2t-k) \tag{1.a}$$

$$W_{2n+1}^{j}(t)=\sqrt{2}\sum_{k}g(k)W_{n}^{j}(2t-k) \tag{1.b}$$

$$W_{j,n,k}(t)=2^{-j/2}W_{n}(2^{-j}t-k) \tag{1}$$

Where, j is a scale parameter, k is a time localization parameter. The analyzing functions W_j,n,k are called wavelet packet atoms as in equation 1[10]. It is observed that the wavelet tree is a part of the wavelet packet tree. Each node of the WP tree is indexed with a pair of integers (j,m) where j is the corresponding level of decomposition and m is the order of its position in the specific level. In each level j there are 2^j nodes and their order is m=0,1,…,2^j-1. A vector of wavelet packet coefficients $C(j,m)$ corresponds to each node (j,m), according to the basic step procedure. The length of a vector C_j,m is approximately $N_t/2_j$. N_t can be produced, by setting to zero the coefficients of all the other vectors in level j, and by implementing the wavelet packet tree inversely. The sampling rate f_S of the signal is assumed 2.04 kHz. This is explained by the fact that low pass filters may carry information about high frequency content of the signal, due to frequency folding, caused by down sampling.

As a time scale domain technique, the wavelet transform utilizes template functions of different time resolutions at different scales to extract “transient” features embedded in a signal. Such transient features can be generated by the interactions between the rolling elements in bearing and a localized defect like spall on the surface of outer raceway. As the rolling elements periodically roll over the localized defect, the “transient” feature will reoccur at a fundamental frequency, which is a function of the bearing rotational speed. Such relationship will be reflected in a wavelet transform of the bearing vibration signal, in that the measure function $C(s_0,m)$ will retain the same fundamental frequency along the time axis at one of its scales (s_1). As a result, the spectral feature of the transient signal is retained in the wavelet transform, although it is not expressed explicitly. It is difficult to rely on wavelet transform alone to identify hidden pattern as noise and other signals are masked at the same scale with similar spectra characteristics. Such constraint of the wavelet transform can be compensated by wavelet packet transform by auto correlating selected decompose detail wavelet coefficient and original approximate coefficient at level J applying the Fourier transform to the measure function C (j, m) resulting from the wavelet packet transform. Such a post spectral technique reveals the specific frequency location of the transient features. The procedure for extracting coefficient at each level with WPT outline as below:

- Signal decomposition: Choosing a wavelet and decomposition level J, and then performing WPT up to level J on the signal. The decomposition can be implemented using filtering and down sampling, and can be iterated, with successive approximation as in [Kittler (1975)]. The total decomposition levels (J) can be calculated according to the following relationship:

  $$j \geq \log_2(f_s / f_c)+1 \tag{2}$$

  Best suited decomposition levels for outer race characteristic defect frequency detection are shown in Table 1.
- For selected decomposition level from 1 to J, coefficient at each node is taken.

Table 1. Best suited decomposition level for outer raceway defect frequency detection

Shaft speed (rpm)	f_{bpfo} (Hz)	Decomposition Level	Frequency band (Hz) Approximate Coefficient	Frequency band (Hz) Detail Coefficient
1000	59.8	6	0-32	32-64
2000	120	5	0-64	64-128
3000	179	4	0-128	128-256
4000	239	4	0-128	128-256
5000	299	3	0-256	256-512
6000	359	3	0-256	256-512

3.0 Artificial Neural Network (ANN)

ANN is self-possessed large number of neurons working simultaneously to solve a specific problem. ANN consists of a set of nodes organized into layers and connected through weight element called synapses. At each node, the weighted inputs are summed, threshold and then inputted to the activation functions to produce output for that specific node. These operations are shown in Fig.1. Most neurons in Neural Networks convert their inputs using a scalar-to-scalar function called an activation function, yielding a neuron output. The most commonly used activation functions are linear functions, threshold functions, sigmoid functions and bipolar sigmoid functions. The sigmoid function was selected in the NN classifiers used in this paper. Sigmoid function is suitable due to following reasons.

1) Sigmoid function has particular advantages for use in back propagation NN because it is easy to differentiate and thus can dramatically reduces the computational burden for training.
2) Function is suitable for applications whose desired output values are between 0 and 1.
3) Activation functions should be chosen to suit the distribution of target values for output neurons.

Sigmoid function is well suited for target values that are binary [0, 1]. The sigmoid functions have additional advantages for continuous valued targets with a bounded range, provided that either the outputs or the targets are scaled to the range of the output activation function [Paya et al. (1997)].

In feed-forward networks, neurons are organized in layers and these layers are connected in one direction and there is no loop in the network. Through a training process, the parameters of ANN are adjusted for a specific application. There are two kinds of learning algorithm: supervised and unsupervised. In supervised learning, by using target output a neural network are trained in a way that a particular input leads to a specific target output. One of the most important supervised learning algorithms is the Back-propagation learning Algorithm which is mainly used to train multilayer neural networks. In back propagation learning the network weights is updated in the direction of the gradient of the performance function [karry and silva (2004)].

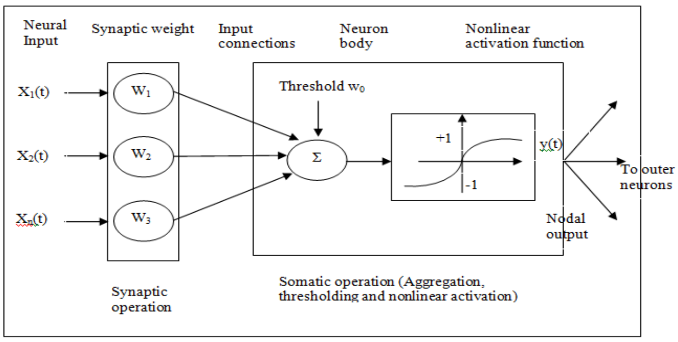


Fig. 1. Operation at a node of neural network

A few aspects, such as number of neural node and layers of neural network need to be consider while designing the structure of neural network classifiers. Large numbers of neurons enable a NN to approximate functions with more complexity. Greater degrees of freedom in between the training points may achieve with more neurons. But, this might cause the test data to fit the function desired poorly. Similarly, single-layer networks are more capable of representing only linearly mapping to any accuracy with sigmoid functions. Multi hidden separable functions which can be approximated any layer, increases the number of local minima. Even if the training algorithm can find the global minima, it is highly possible to be stuck in a local minimum after many time consuming iterations [Rajakarunakaran (2008)]. Considering the limitation and strength of number of hidden layers in NN, only two hidden layers are designed and tested in this study.

4.0 Experimental set up

In the present study, experimentation is carried out on a rotor test rig [Pandya *et al* (2012)]. This consists of a shaft supported on rolling element bearings and driven by a DC motor with external servo controller. A spring type flexible coupling has been used to connect rotor shaft and motor to compensate any misalignment. The SKF 6205 bearings have been taken for the study, which is same as was taken in the theoretical simulation. The horizontal rigid rotor of weight 2.5 kg is used and rig is connected to a data acquisition system through proper instrumentation as shown in Fig. 2. Two piezoelectric accelerometers (bi-axial) are used for picking up the vibration signals from various stations on the test rig. These special piezoelectric pickup type of sensors with frequency range of 1-30 KHz, measurement range +/- 500g peak, resolution of 0.005 g and resonant frequency of 70 KHz. Vibration responses were acquired and analyzed by analyzer with 4 input channels and sampling rate of 0.48 – 102.4 Hz. The localized defects considered are spall on outer race, inner race and rolling elements. Photographic views of defects are in Fig.3. Bearing geometry is described in Tab.2.

Table 2. Details of rotor-bearing experimental set up

Bearing specification	SKF 6205
Mass of the rotor (m_r)	2.5 kg
Diameter of inner race with point of contact with the rolling element (d)	31.12 mm
Diameter of outer race with point of contact with the rolling element (D)	47 mm
Radius of each rolling element (ρ_r)	3.965 mm
Radial load (W)	25 N
Outside diameter	52 mm
Number of rolling elements (N_b)	9
Angular separation between elements (β)	40 deg
Spall a seeded defect (micron)	300X300X20

Fig. 2 experimental setup

Fig. 3 Localized defect seeded in the bearing elements

The FFT diagram of all bearing condition in vertical and horizontal directions at 1000 rpm is as shown in Fig.4(a)-(e) and Fig.5(a)-(e) respectively. The vibrations response of healthy bearing is excited at 2X due to misalignment present in the set up, as shown in Fig.4(a) & Fig 5 (a). The vibrations response of bearing with ORD (spall) is excited at (ω_{bpfo} = 59.8 Hz), as shown in Fig.4 (b) & Fig 5 (b). The vibrations response of inner race defect (spall) is excited, mainly due to interaction between cage frequency (ω_{cage}=6.68 Hz) and inner race passage frequency on the outer race (ω_{bpfi} = 90.8 Hz) due to spall on inner race, as shown in Fig.4(c) & Fig 5 (c). The significant response peaks are recorded at 2X and 2(ω_{bpfi} + ω_{cage}). The vibrations response of ball defect (spall) is excited, mainly due to ball pass frequency (ω_{bpf} = 39.3 Hz) due to spall on ball, as shown in Fig.4(d) & Fig.5(d). The dominant response at (ω_{bpf} =39.3 Hz) clearly shown in Fig.4 (d) and horizontal response elaborate real condition of ball defect. The vibrations response of combined defect (spall) is excited, mainly due to interaction between cage frequency (ω_{cage} = 6.68 Hz), inner race passage frequency (ω_{bpfi} = 90.8 Hz), ball pass frequency (ω_{bpf} = 39.3 Hz) and outer race frequency (ωbpfo = 239 Hz) due to spall on outer race, as shown in Fig.4(e) & Fig 5 (e).

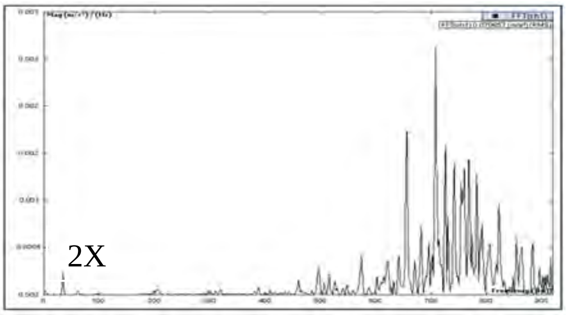


Fig. 4.(a) vertical response with HB

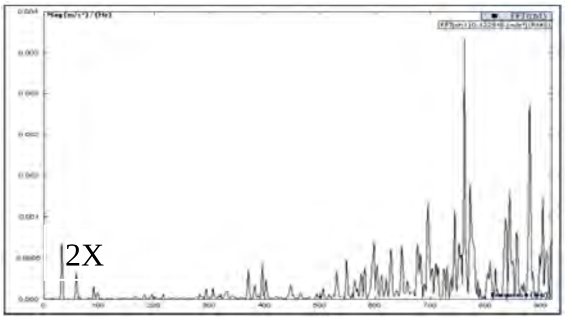


Fig 5. (a) horizontal response with HB

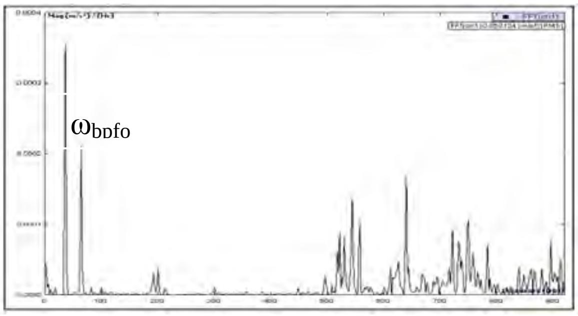


Fig. 4. (b) vertical response with ORD

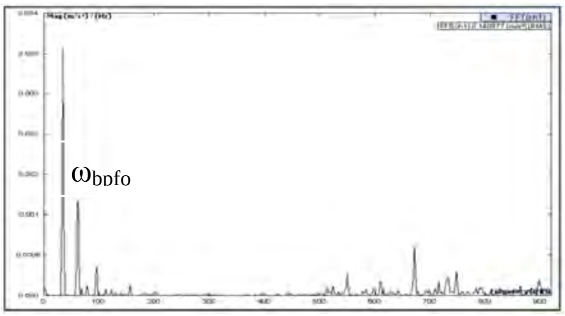
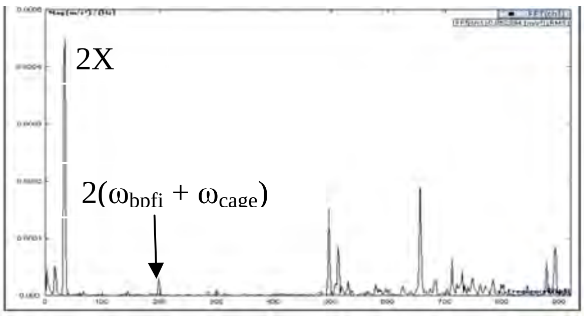


Fig. 5. (b) horizontal response with ORD



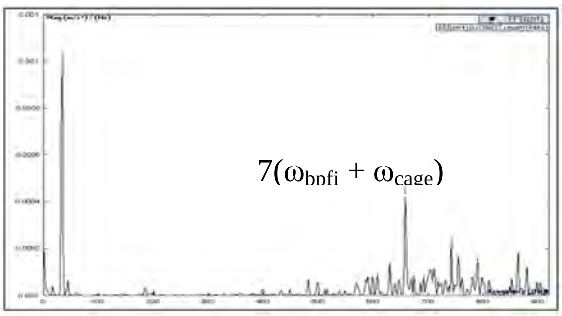
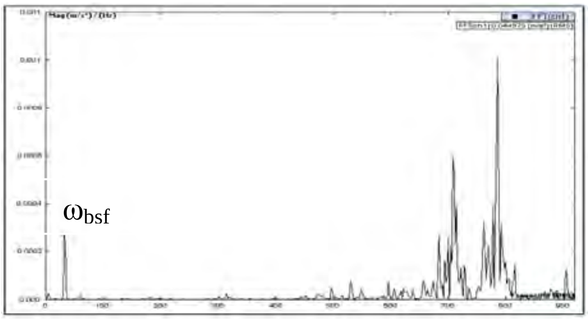


Fig. 4. (c) vertical response with IRD

Fig. 5. (c) horizontal response with IRD



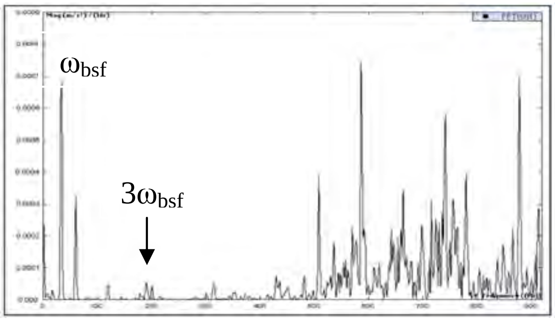
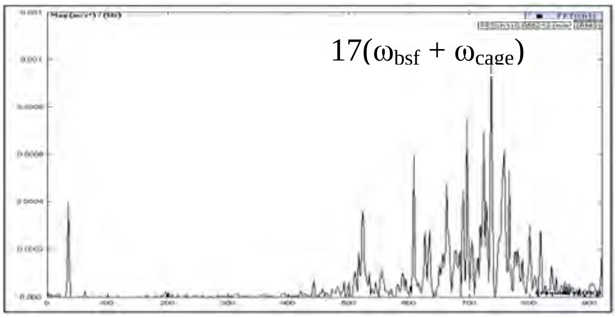


Fig. 4. (d) vertical response with BD

Fig. 5. (d) horizontal response with BD



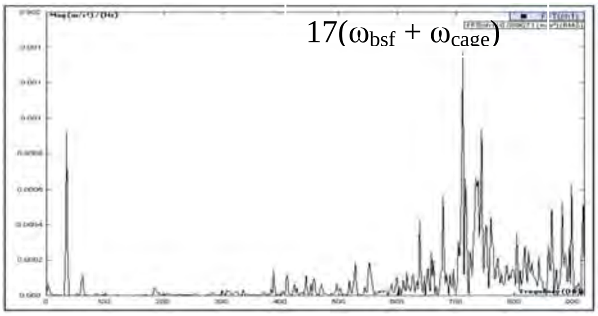


Fig. 4. (e) vertical response with CD

Fig. 5. (e) horizontal response with CD

5.0 Implementation of Proposed Methodology

In present work, a ‘rbio5.5’ real wavelet has been selected based on energy to Shannon entropy criteria to extract the features at appropriate level of decomposition using wavelet packet transform. Level of decomposition is based on sampling frequency and characteristic defect frequency of rolling element at different rotating speed. The two features kurtosis and sub band energy are extracted, based on Linear Discriminant Analysis. Normalized data of selected features for vertical and horizontal responses are then input into neural network with twelve input nodes. A FFNN comprises of two hidden layers with fourteen nodes and output layer with five nodes, which use to classify the bearing conditions. The training and testing of the NN is done for various bearing condition determination of healthy bearing (HB), outer race defect (ORD), inner race defect (IRD), ball defect (BD) and combine defect (CD). The block diagram of proposed method is shown in Fig 6.

5.1 Features extraction and selection

WPT allows for feature extraction from sub frequency bands of the decomposed signal where the features are concentrated, thereby directing the computation to where it is most needed. Feature vectors constructed based on the energy and kurtoses of wavelet coefficients are more significant combination to diagnose and classify defects.

The energy content of a signal provides a quantitative measure for the signal, which uniquely characterizes the signal. The total amount of energy contained in the signal is equal to the sum of the energy in each sub band and expressed as equation 3.

$$E_{x(t)} = \sum_{m=0}^{2^j-1} \int |x_j^m(t)|^2 \, dt \qquad (3)$$

Since the energy content of each sub band of the signal is directly related to the severity of the defect, it presents an indicator or feature of the bearing’s condition.

Kurtosis of sub band coefficient gives another metric which can identify whether coefficients at a level are transiently varying over time or not. In other way, Kurtosis is a dimensionless, statistical measure that characterizes the flatness of a signal’s probability density function. An impulsive signal that is peaked has a larger kurtosis value than a signal that is flat and varies with time slowly. Mathematically, it is defined as the forth order moment of signal data and kurtosis of sub band coefficient is expressed as in Equation 4.

$$K_j^m = \frac{\sum\{[x_j^m(t) - \bar{x}_j^m(t)]^4\}}{\sigma^4{}_{x_j^m(t)}} \quad (4)$$

Since, the energy content of a signal provides a robust indicator of the signal, but is not sensitive in characterizing incipient defects, whereas the kurtosis value has high sensitivity to incipient defects but has low stability these two features can be combined instead of being used alone to improve the signal characterization[Yan and Gao (2004)].

The approach introduced here for efficient feature selection is to evaluate the discriminant power of each individual feature within a class pair. Features that have a low discriminant power are excluded from the data analysis process, as they contain little useful information. Such an approach can be realized by examining the rank order. Based on linear discriminant criteria four condition pairs are considered in this work with reference data of healthy bearing condition. The first three key node features in each condition pair are selected for each feature from vertical and horizontal vibration signals. Thus total 12 input nodes per data is design for the proposed NN. One input vector for vertical response is as shown in Table. 3.

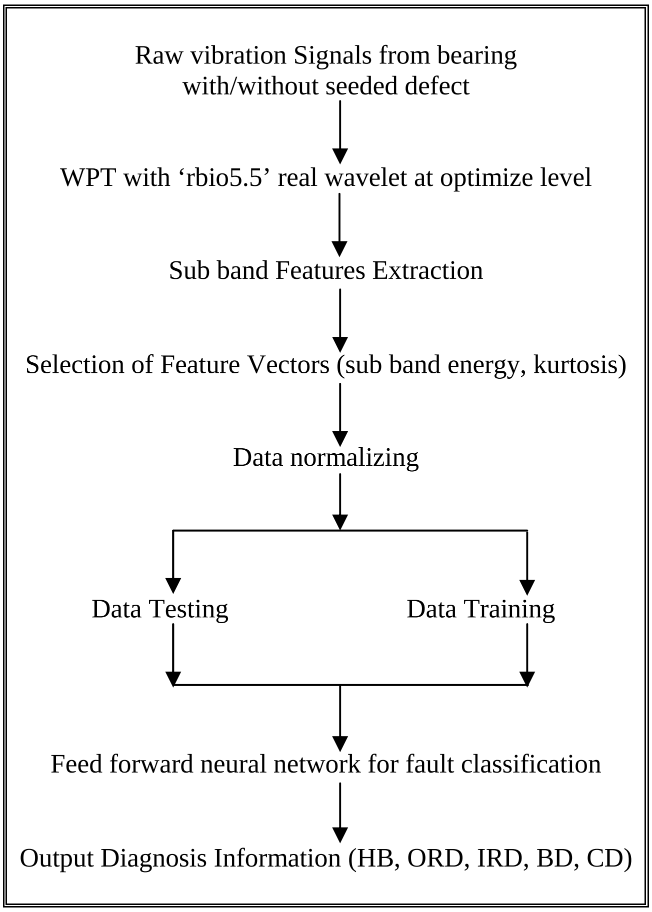


Fig. 6. Block diagram for proposed ANN with multi layer perceptron

Table 3. A input vector extruded thru WPT

Sr. No.	**Bearing Condition**	**Energy**			**Kurtosis**		
1	HB	16.38	7.55	5.37	4.17	3.62	3.14
2	ORD	19.50	6.55	2.83	0.93	0.83	0.60
3	IRD	11.37	6.31	4.47	6.11	3.43	2.90
4	BD	11.77	8.00	4.65	3.50	2.68	2.62
5	CD	13.23	7.42	3.69	5.84	4.82	3.97

5.2 DATA NORMALIZATION

As shown in table 3 the energy vectors contain the higher values then the kurtosis. Hence, during training of the neural network, higher valued input variables may tend to suppress the influence of the smaller one. To overcome this problem and in order to make neural networks perform well, the data must be well processed and properly scaled before inputting to the ANN. The Colum data are normalized in the range 0.1 to 0.9 to minimize the effect of input variable. The range 0.1 and 0.9 is selected instead of zero and one because zero and one can not be realized by the activation function (sigmoid function). All the component of feature vector is normalized using the following equation.

$$X = 0.8\left(\frac{x - \min(x)}{\max(x) - \min(x)}\right) + 0.1 \tag{5}$$

Where, X is normalized data, x actual data, min(x) & max (x) minimum and maximum in the respective data set.

6.0 RESULTS & DISCUSSION

In this work, FFNN with two hidden layers, twelve input nodes and a output layer with five nodes were designed. The neural network is trained using an error back propagation algorithm. The training can cease according to the criteria of either mean square error (MSE) reach to certain value or that the epoch of training reaches certain value. In our application a target mean square error of 10^{-5} and a maximum iteration number (epoch) of 400 is setup. The training process would stop if any of these conditions were met. The initial weights and biases of the network were generated automatically by the program. The target value of output layer node based on normalization data is shown in Tab. 4.

Table 4 Target vector for output layer

Sr. No	Bearing Condition	Node 1	Node 2	Node 3	Node 4	Node 5
1	HB	0.9	0.1	0.1	0.1	0.1
2	ORD	0.1	0.9	0.1	0.1	0.1
3	IRD	0.1	0.1	0.9	0.1	0.1
4	BD	0.1	0.1	0.1	0.9	0.1
5	CD	0.1	0.1	0.1	0.1	0.9

Table 5. Summary of overall fault classification

Sr. No.	Bearing Condition	Testing Set	Correctly classified	Mis-classified	% of Accuracy
1	HB	20	19	1	95
2	ORD	20	20	0	100
3	IRD	20	19	1	95
4	BD	20	19	1	95
5	CD	20	20	0	100
	Total	100	97	3	97

The total of five vibration signal corresponding to each bearing condition is recorded for fault classification. The length of the each signal is 10240. This 10240 data is segmented into ten so that signal length is 1024. Therefore number of vibration signal recorded for one bearing condition are 5*10=50 samples per rotating speed of rotor. Here consider five rotation speed 1000,1500,2000,2500 and 3000 rpm, hence for 50*5=250 samples iterated four times make available total 250*4=1000 samples for fault classification. Out of 1000 samples 70% of data iterated for training purpose, 20% data for validation and 10% of data is for testing purpose in the neural network for fault classification. During our training processes generally the iteration is reached first rather than mean square error. The mean square error at this time is used as a criterion for appraising the training performance of the neural network and the classification rate as the criterion for appraising each diagnosis procedure. The classification rate of bearing fault is shown in Tab. 5. Out of 20 samples for each bearing condition 19,20,19,19 and 20 are correctly classified and 1,0,1,1 and 0 misclassified for HB,ORD,IRD,BD and CD, respectively. Overall 3 samples are misclassified by proposed ANN system. The overall fault classification rate is 97%.

7.0 CONCLUSION

This paper has investigated the feasibility of applying wavelet packet decomposition for feature extraction of vibration signals. Also, highlight the significance of energy and kurtosis as main features to train and to execute neural network. To alleviate the time-invariant characteristics of the wavelet packet coefficients and to reduce the dimensionality of the input to the neural network with normalized data. The features obtained by proposed method for vibration signal yields nearly 93% correct classification when used as input to a neural network classifier with MLP.

Acknowledgment

This work was financially supported by the Department of Science and Technology, Government of India [Grant number DST/457 /MID].

References

[1] Harsha, S.P.; Nataraj, C. (2007(b)), Non-linear Dynamic Response of a Complex Rotor Bearing System, IJAVE,12(1), pp. 379 – 399.

[2] Karray, F.; Silva, C. W. de. (2004). *Soft Computing and Intelligent Systems Design, Theory, Tools and Applications*. Essex, England: 1st edn. Addison-Wesley, Canada.

[3] Kittler J. (1975): Mathematical methods of feature selection in pattern recognition, IJMMS, 7(5),pp.609-637.

[4] Li, B.; Goddu, G.; Chow, M. (1998): Detection of common bearing faults using frequency-domain vibration signals and a neural network based approach, American Control Conference, pp. 2032-2036.

[5] Lin.J.; L. Qu. (2000): Feature extraction based on morlet wavelet and its application for mechanical fault diagnosis, JSV, 234(1), pp.135-148.

[6] Liu,T.I.; Mengel, J.M. (1991): Detection of ball bearing conditions by bearing conditions by an A.I. approach. Sensors, Controls and Quality Issues in Manufacturing, ASME, PED, 55, pp.13-21.

[7] McFadden, P.D.; Smith, J.D. (1985): The vibration produced by multiple point defects in a rolling element bearing, JSV, 98 (2), pp. 263-273.

[8] Nikolaou, N.G.; Antoniadis, I.A. (2002): Demodulation of vibration signals generated by defects in rolling element bearings using complex shifted Morlet wavelets, MSSP, 16(4), pp. 677-694.

[9] Paya,B.A.; Esat,I.I; Badi, M.N.M. (1997): Artificial neural network based fault diagnosis of rotating machinery using wavelet transform as a preprocessor, MSSP, 11(5),pp.751-765.

[10] Pandya,D. H.; Upadhyaya, S.H.; Harsha,S.P. (2012): Intermittent Chaotic Behavior of High Speed Ball Bearings Due to Localized Defects, Proc. 4th ICSSD, Texas A&M University (USA) & MNIT-Jaipur, India, vol.1, pp.406-413, Excel India Publishers, New Delhi, INDIA.

[11] Rajakarunakaran, S.; Venkumar, P.; Devaraj, K.; Rao, K.S.P. (2008): Artificial Neural Network approach for Fault Detection in Rotary System, ASC, 8(1), pp. 740-748.

[12] R.Polikar, the Wavelet Tutorial (Online).Available:

[13] http://engineering.rowan.edu/polikar/WAVELETS/ WTtutorial .html

[14] Samanta, B.; Al-Balushi, K. R. (2003): Artificial Neural Network Based Fault Diagnostics of Rolling Element Bearings Using Time-Domain Features, MSSP, 17(2), pp.317–328.

[15] Sugumaran, V.; Muralidharan, V.; Ramachandran, K. I. (2007): Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. MSSP, 21,pp. 930–942.

[16] Yan, R.; Gao, R. (2004): Harmonic wavelet packet transform for on line system health diagnosis, SPIE international symposium on sensors and smart structures technologies for civil, mechanical and aerospace systems, San Diego, CA, pp.512-522.