Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 97 (2014) 1831 – 1841

Procedia Engineering

www.elsevier.com/locate/procedia

12th GLOBAL CONGRESS ON MANUFACTURING AND MANAGEMENT, GCMM 2014

A Gear fault identification using wavelet transform, rough set based GA, ANN and C4.5 algorithm.

Rajeswari.C[a]* ,Sathiyabhama.B[b],Devendiran.S[c] & Manivannan.K[d]

[a]*School of Information Technology and Engineering ,VIT University, Vellore, Tamil Nadu, India*
[b]*Department of Computer Science and Engineering, Sona College of Technology, Salem, Tamil Nadu,India*
[c,d,]*School of Mechanical and Building Sciences ,VIT University, Vellore, Tamil Nadu, India*

Abstract

Early fault detection methodology in gear box diagnosis has been proposed to find the status of the gear based on vibration signals obtained from the experimental test rig. Signal processing categorized to time-frequency domain such as continues wavelet transform is used in the proposed work for statistical feature extraction. Feature selection method is used for selecting the extensive useful features among the extracted features to reduce the processing time. A famous optimization technique, Genetic algorithm (GA) and rough set based approach is used to select the best input features to reduce the computation burden. The efficiency of this feature selection method is evaluated based on the classification accuracy obtained from the proposed algorithms: back propagation neural network (BPNN) a famous artificial neural network algorithm and C4.5.Performance of classifiers are evaluated with the different signals acquired from the experimental test rig for different states of gears.

Selection and peer-review under responsibility of the Organizing Committee of GCMM 2014

Keywords: Vibration; Gear test rig; continues wavelet transform (CWT); feature extraction; Genetic algorithm (GA); rough set; fault diagnosis.

1. Introduction

Gears plays an vital role in a wide range of transport and industrial rotating applications. Early fault diagnosis of the gear may prevent unnecessary failures of the system .Therefore many researchers strives towards the extensive investigations on this research domain [1-2].In common vibration based time domain, frequency domain, and time–

* Corresponding author. Rajeswari.C[a]* Tel.: +91 9600376711.
E-mail address: rajeswari.c@vit.ac.in

Selection and peer-review under responsibility of the Organizing Committee of GCMM 2014
doi:10.1016/j.proeng.2014.12.337

frequency domain of techniques were used for the detection of faults in the rotating machinery components by many research [3]. In general Statistical features extraction process carried out from time domain and frequency domain can be used along with the data mining procedure for identification of fault condition of the gear. Time–frequency analysis method such as wavelet transform have been applied in this fault diagnosis to get useful information by provide local features in both time and frequency domains and it has the characteristics of multi-scale and to distinguish the abrupt components of the vibration signal [4]. In the family of wavelet, continuous wavelet transform (CWT) is considered as an effective vibration based signal processing method for gear fault detection because of its nature of giving a multi-resolution in time-frequency domain for characterize the features of non-stationary signals [5]. After signal processing, statistical characteristics of vibration signal such as mean, peak to peak, standard deviation, skewness, kurtosis, and crest factor were calculated for different states of conditions. The vast number of features within variety domains poses challenges to data mining. In order to achieve successful data mining, feature selection is an essential process [6]. Researches in this area focus on reducing computation time for the purpose of efficient research. Many researchers in this domain used principal component analysis (PCA) [7] and Genetic algorithm (GA) [8] and j48 algorithm [9] to decrease the relativity between features and decrease dimensions of features. Statistical features of different conditions either with or without feature selection or feature reduction classified through sophisticated machine learning tools such as the artificial neural network (ANN)[10], support vector machines (SVMs)[10], and fuzzy logic systems (FLS)) for its meaningful diagnostics of conditions of faults in gear components . Among them ANN and C4.5 are the two best-known and widely used algorithms presented in this research work for classification.

2. Vibration Experimental setup and experimental procedure

A. Test rig

Experimental test rig shown in Fig.1(a) comprise of AC motor, variable frequency drive (VFD) as a motor speed regulator, tri axial accelerometer as a sensor and data acquisition system (DAS-ATA0824DAQ51) integrated with computer , The frequency range of gear fault is mostly identified at middle-frequency band [$5f_0$, 1000 Hz] were f_0 is rotating frequency[7]. Therefore the sample frequency is set to 6400 Hz. The number of sampling rate is 12400 /sec. The selected constant rotating speeds are 600, 1000, 1400 rpm for the four states of gear condition such as normal, frosting, pitting and crack. For each gear condition 25 groups of sample data were acquired.

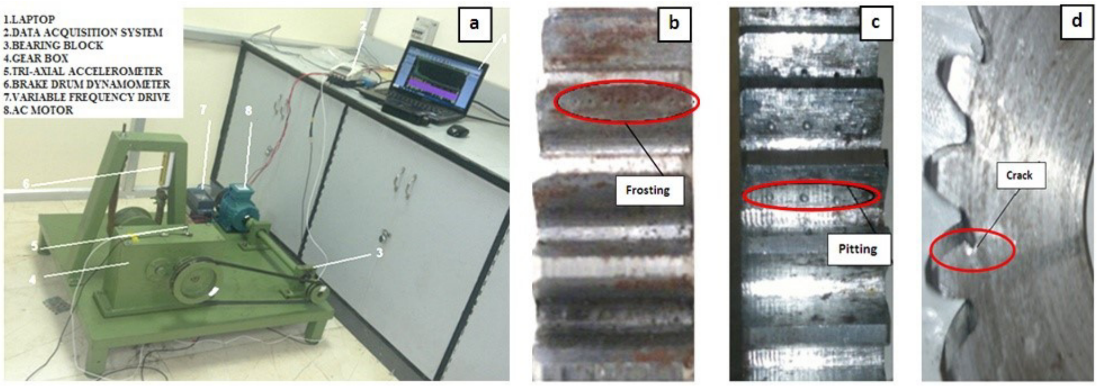

Fig.1. (a) Gear test rig (experimental set up) (b) frosting induced gear (c) pitting induced gear (d) crack induced gear

B. Experimental procedure

Experiment process of gear fault diagnosis is made on the well designed and fabricated test rig. By means of vibration data collection, pre-processing, feature extraction and feature selection training and test samples are obtained. Consequently use the training samples to train and test samples for testing. And the diagnosis result is compared with the result of neural network and C4.5 algorithm. The methodology of experiment process is shown in

Fig. 1. Gear conditions such as Normal, frosting(Fig 1(b)), pitting (Fig.1(c)) and crack (Fig.1(d)) were artificially made using electrical discharge machining (EDM) process and using these gears in the gear box the experiment was carried out in different speed and load conditions. The time domain waveforms and spectrums of vibration signal for four conditions of gear are shown in Fig.2.The methodology of proposed work is shown in the Fig.3 (a)

3. Signal Pre-processing and Feature extraction

In general machine component faults are identified through signal processing techniques such as Time domain analysis, Frequency analysis and Time-frequency analysis. Fourier analysis converts time (or space) to frequency and vice versa; an FFT (Fast Fourier Transform) rapidly computes such transformations by factorizing the DFT matrix into a product of sparse (mostly zero) factors. FFT can be used to indicate the intensity of the frequencies in the spectrum. FFT cannot find the non-stationary transient information from the samples, which serves as the reason this paper focus on wavelet transform. Versatility and Effectiveness of wavelet transform over Fourier transform is discussed elaborately in [11]. However, various kinds of factors, such as the change of the environment and the faults from machine itself often make the output signals of the running machine contain non-stationary (frequency changes with time) signals which are the chief sources of abundant information about machine fault. The features are extracted from the wavelet-transformed data. These features form a transformed space and it is used as the input of next process called feature selection and further classification. Wavelet Transform method is proposed to be used as dimensionality reduction function for the raw data. Wavelet Transform (WT) is a time-frequency decomposition of a sample signal into "wavelet" basic function. Wavelet analysis is widely used for decomposing, de-noising and signal analysis over a non-stationary signal. At high frequencies WT gives good time and poor frequency resolution, and at the same time at low frequencies it gives good frequency and poor time resolution. Investigation with wavelets proceed with breaking up a signal into shifted and scaled versions of its mother (or original) wavelet, that is obtaining one high frequency term from each level and one low frequency residual from the last level of decomposition. In other words Decomposition of signal is a process of breaking of signals into lower resolution components with respect to levels. Particularly Continuous Wavelet Transform (CWT) provides a multi-resolution in time-frequency analysis for characterizing the transitory features of non-stationary signals [12]. This locates the wavelet transform apart from the Fourier Transform, the effect were accumulation of higher frequency sine waves spread throughout the frequency axis. CWT is widely used to divide a continuous-time function into wavelets. The continuous wavelet transform of a time function z(t) is denoted as :

$$CWT(x,y) = \int_{-\infty}^{\infty} z(t)\psi^*_{(a,b)}(t)\,dt \tag{1}$$

Where $\psi^*_{(a,b)}(t)$ is a continuous function in both the time domain and the frequency domain called the mother /original wavelet and * represents operation of complex conjugate. Further expansion of $\psi^*_{(a,b)}(t)$ gives

[14] . $$\psi^*_{(x,y)} = \frac{1}{\sqrt{x}}\psi\left(\frac{t-y}{x}\right) \qquad where \qquad x, y \in R, x \neq 0 \tag{2}$$

Mother wavelet gives a source function to generate the translated and scaled version of its sibling wavelets. As given in Eqn. (2), the transform signal CWT (a, b) is defined on plane x - y, were a and b are used to change the frequency and the time location of the wavelet. Whenever high frequency resolution is required, the decrement of x will construct a high-frequency wavelet and vice versa is possible. In other side as y increases, the wavelet transverses the length of the input signal, and increases or decreases in response to changes in the local time and frequency content of the signals. Acquired signals are decomposed based on Continuous Wavelet transform, which is then used for extracting various statistical features. Transform coefficients are a measure of similarity between the raw and daughter wavelets[14]. Morlet wavelet is a cosine function which exponentially decreases at both ends (Fig. 3.b). It looks like an impulse function modulated with a cosine function. Morlet wavelet is more suitable in cases of variations found in abnormal stationary signals. The equation for Morlet Wavelet transform is given by the Eqn(3).

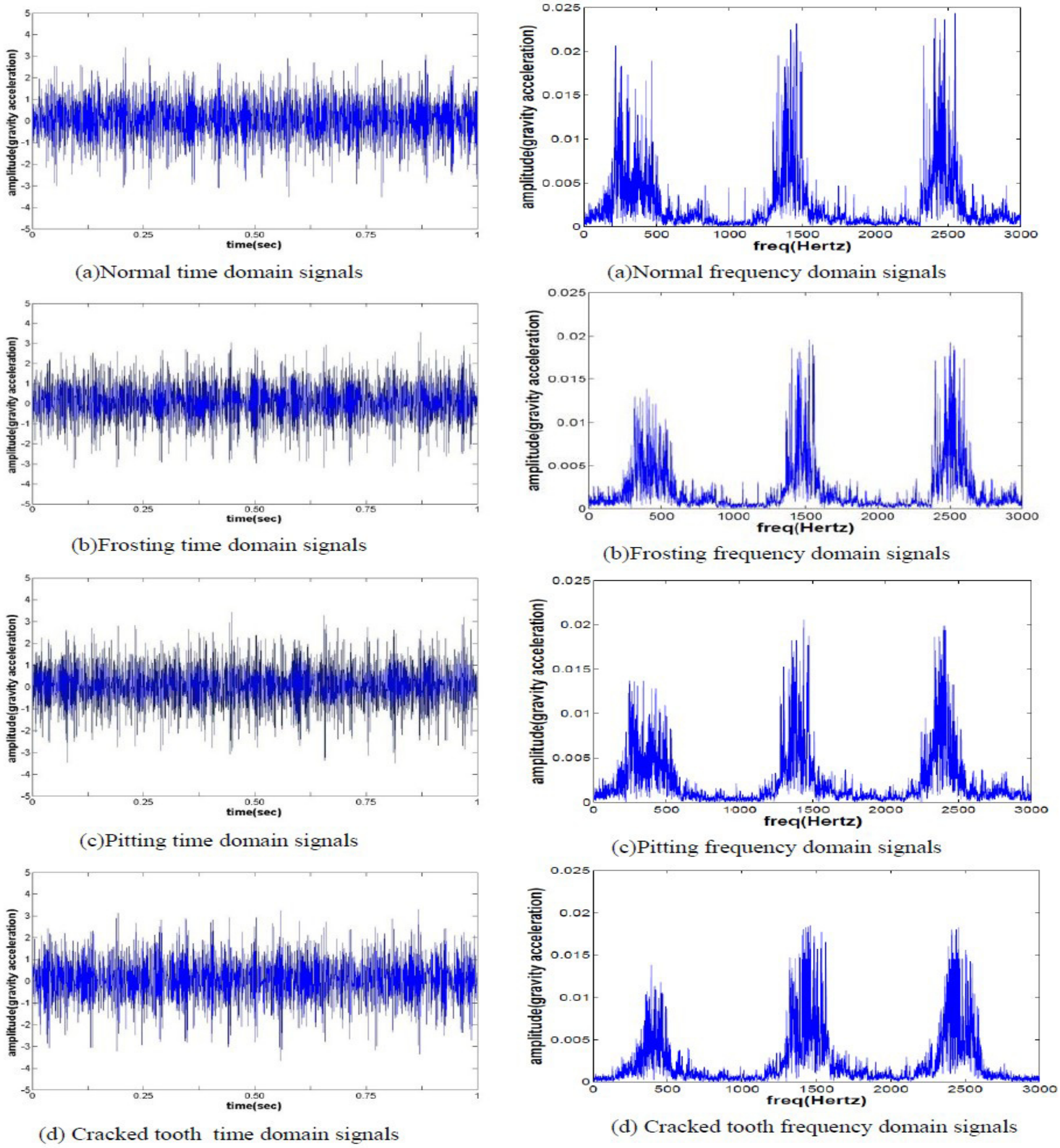

Fig.2. Time and frequency domain signals for various gear conditions.

$$\psi(t) = \frac{1}{\sqrt[4]{\pi}}\left(e^{j\omega_0 t} - e^{-\frac{\omega_0^2}{2}}\right)\cdot e^{-\frac{t^2}{2}} \quad \text{Where} \quad \omega_0\text{: Central frequency of mother wavelet} \tag{3}$$

Initially using Mat lab platform, Morlet wavelet based multilevel 1D wavelet decomposition function is chosen with 64 scales to extract the Morlet wavelet coefficients from the signals. But, for 8 level scales it gives the maximum efficiency [13] for the proposed classifiers. For further proceedings the statistical features corresponding to 8 scales were calculated and given as a input for classification algorithms. The 10 statistical features widely used [10] in gear fault diagnosis such as mean, standard deviation, root mean square, peak, Skewness , kurtosis, crest factor, clearance factor, shape factor, and impulse factor were extracted from the wavelet coefficients and feed as a input to the proposed algorithms .

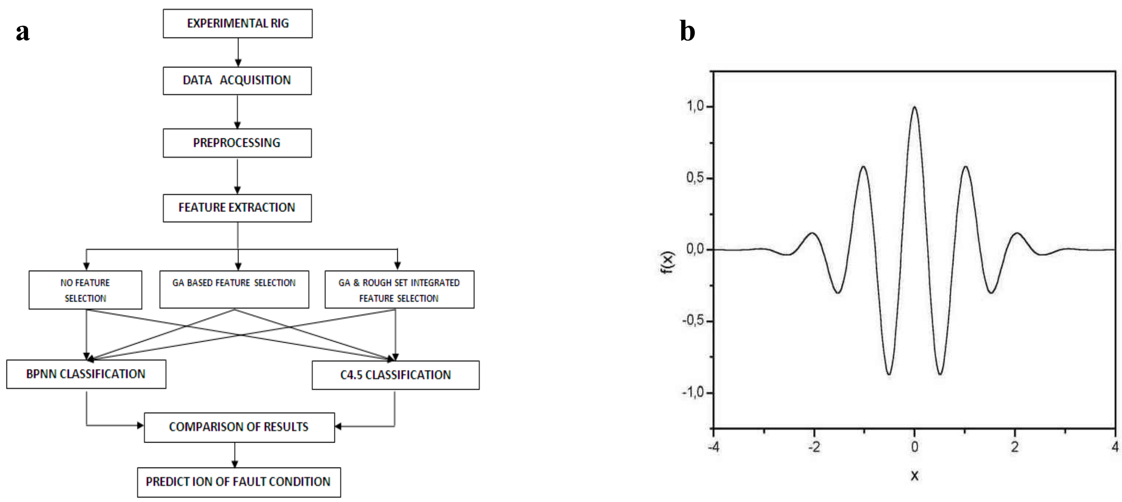

Fig.3. (a) Methodology of experiment process (b) Morlet wavelet

4. Feature selection using Genetic Algorithm

The genetic algorithm gives a result by simulating the evolutionary processes of survival of the fittest, which gives the best members of the population. The algorithm starts with a set of chromosomes called the population. By means of random, new generation of population is created from the previous population by reproduction. Then the fitness of each chromosome is calculated. The next new generation is reproduced by selecting the best pairs of parents. At the time of reproduction, crossover of the genes and random mutation occur in some of the children chromosomes. Reproduction process will proceed until the best solution is obtained. For further theoretical background of GA refer [10]. For feature selection as per [23] GA multiple objective functions were used to find the best solution. The GA chromosome represents a chain of features, in which the binary code includes 18 bits based on the number of all features taken into account features, where "1" represents a selected feature and "0" represents a discarded feature. The objective of the GA is to minimize the within-class distance and a maximize between-class distance. It is given by

$$K_c = \sum_{i=1}^{c} s_i K_i \tag{4}$$

$$K_c = (1/n_i)\left\{\sum_{i=1}^{n}\left(x_j^i - m_i\right)^T\left(x_j^i - m_i\right)\right\} \tag{5}$$

Class i =1,2,……C; m_i is the mean vector of class i ; n_i is the number of samples in class i ; s_i is the number of samples in class i.

Distance between the classes is calculated as follows

$$K_b = \left\{\sum_{i=1}^{c} s_i\left(m_i - m\right)^T\left(m_i - m\right)\right\} \tag{6}$$

m is the mean vector of all of the classes.

The GA is a method with two objectives. The first one is to determine a minimum distance in the within class, and the second is to determine a maximum in the average distance between-classes. To satisfy this purpose, the fitness function is defined as,

$$K = K_i + \left(1/K_b\right) \tag{7}$$

Thus the chromosome that minimizes the fitness function is chosen, to get optimal features. The execution parameters of genetic algorithm were set as the population size equals 100, the length of chromosome or features taken as 10, crossing point taken as 1, and number of generation is 100. The feature selection result shows the best features that satisfying the given objective function. For example; output results 1001101101 means that features F1, F4, F5, F7, F8, and F10 were selected and, remaining features were discarded. Further selected features used as a input to proposed algorithms.

5. Feature selection using Rough set based Genetic Algorithm

Rough Set (RS) theory is one of the mathematical tool that deals with feature reduction e to find a minimal subset. The basic concept is by making an upper and a lower approximation of the data set. The feature reduction is achieved by comparing equivalence relations generated by feature sets considering the dependency degree as an important measure, features are removed and the reduced feature set provides the same dependency degree as the original. For further detailed description about the rough set theoretical background refer [21]. For better understanding of feature reduction using Rough set is explained with the following example. Features have to be discretized before proceeding with feature reduction. Certain feature reduction algorithms cannot cope with features having continuous data values as the input. A potential resolution for this type of problem is to; break the continuous values in to multiple groups of data and thereby converting the same to categorical data types. This process of converting the numerical variables to categorical variables is called as discretization. Many of the research survey [22] shows that discretization techniques have been proposed based on the requirements of the data analysis. The common binning discretization method is used in this proposed work. In the proposed work, rough set based Quick Reduct (QR) algorithm is used and it calculates reduced features without totally producing all probable subsets. It initiates with a blank set and appends in sequence, one by one, individual features that effect in the maximum raise in the rough set dependency factor, until this makes its utmost probable value for the dataset. According to the algorithm, the dependency of each feature is calculated and the finest candidate is selected [24].

5.1 Worked out example for quick reduct rough set

Table 1, illustrates a sample dataset. The set $\{p,q,r,s\}$ are defined as the conditional attributes. They can also be defined as independent attributes, as the values of these attributes does not need rely on any other attributes values. The dependent attributes are which has to dependent on other attributes values for any decision making. In Table 1, T is the class label or decision attribute.

Step 1: The dependency value of decision attributes $\{T\}$ w.r.t. to all independent attribute is calculated. Where U the universe contains non empty set of object.

$$fitness = \gamma_c(T) = \gamma_{\{p,q,r,s\}}(T) = \frac{|POS_{\{p,q,r,s\}}(T)|}{|U|} \tag{8}$$

$$U/T = |\{1,4,6,7\}\ \{2\}\ \{3,5,8\}|$$

$$U/IND(R) = |\{1\},\{2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\}|$$

$$\gamma_c(T) = \frac{|\ \{1,2,3,4,5,6,7,8\}\ |}{|\ \{1,2,3,4,5,6,7,8\}\ |} = \frac{8}{8} = 1$$

Step 2: The dependency values of each attribute w.r.t to decision attributes are calculated.

$$\gamma_{\{p\}}(T) = \frac{|POS_{\{p\}}(T)|}{|U|} = \frac{1}{8} = 0.1 \quad ; \quad \gamma_{\{q\}}(T) = \frac{|POS_{\{q\}}(T)|}{|U|} = \frac{2}{8} = 0.2 \quad ;$$

$$\gamma_{\{s\}}(T) = \frac{|POS_{\{s\}}(T)|}{|U|} = \frac{1}{8} = 0.1 \quad ; \quad \gamma_{\{r\}}(T) = \frac{|POS_{\{r\}}(T)|}{|U|} = \frac{0}{8} = 0 \quad ;$$

Step 3: As $\{T\}$ has the highest dependency with $q\{p,q\}\{r,q\}\{s,q\}$ will be calculated.

$$\gamma_{\{s,q\}}(T) = \frac{|POS_{\{s,q\}}(T)|}{|U|} = \frac{8}{8} = 1$$

$\gamma_{\{p,q\}}(T) = \frac{|POS_{\{p,q\}}(T)|}{|U|} = \frac{6}{8} = 0.6$; $\gamma_{\{r,q\}}(T) = \frac{|POS_{\{r,q\}}(T)|}{|U|} = \frac{6}{8} = 0.6$; Step 4: As {T} has the highest dependency with $\{s,q\}$ therefore $\{p,s,q\}$ $\{r,s,q\}$ will be calculated.

$$\gamma_{\{p,q,s\}}(T) = \frac{|POS_{\{p,q,s\}}(T)|}{|U|} = \frac{8}{8} = 1$$

As the dependency of $\{p,s,q\} = 1$, therefore, dependency calculation for $\{r,s,q\}$ is not required. Hence, $\{p,s,q\}$ is the reduced dataset i.e., the reduct with same predictive potential as $\{p,q,r,s\}$

i.e. , $\gamma_{\{R\}}(D) = \gamma_{\{C\}}(D)$ (9)

As the stopping criteria for dependency estimation is equal to 1, therefore the dependency estimation for other combination of attributes is not necessary. Hence, {p, q} is considered as the reduced dataset.

Algorithm-1 : GA and rough set based quick reduct

Algorithm: $HGA(x,c,m)$

Input: S , the set of all independent feature ; T ,is the dependent feature

Output: reduced feature X

Step 1: Initialize the generation, $g := 0; P_g :=$ population of x randomly generated individuals;

Step 2: Compute $fitness(i)$, $i \in P_g$;

Evaluate the dependency factor of (S) with (T) i.e., $\gamma_s(T)$

Evaluate the dependency factor of each independent attributes (Si) with T i.e., $\gamma_{\{Si\}}(T)$

Compare $\gamma_{\{si\}}(T)$ with $\gamma_{\{s\}}(T)$

If

$\gamma_{\{Si\}}(T) = \gamma_S$

Then

put (Si) in the null reduced set X

$terminate(\ )$

Else

Put (Si) in giving highest dependency (S_H)

Evaluate the dependency factor of T with all possible attribute combination i.e., $S_{H(i-H)}$

Compare with $\gamma_{\{S\}}(T)$

If

$S_{H(i-H)} = \gamma_{\{S\}}(T)$

then

put $S_{H(i-H)}$ in the null reduced set X

$terminate(\ )$

Else

Repeat steps viii. to x. Until $\gamma_{\{S\}}(T) = \gamma_{\{X\}}(T)$,

Where S is the set of independent attributes and X is the subset of independent attributes i.e., reduced set

Step 3: Perform genetic operations

Do

{

$variation(sel, x, S, m, c)$

$O =$ New generated population of size S

$reoder(sel, x, S)$

For

$i = 1$ to S

$O[i] = mutate(sel[i], x, m);\backslash$

$O[i+1] = mutate(sel[i+1], x, m);$

If $(rand() < c)$

$crossover(O[i], O[i+1], x);$

$Compute\ fitness(i)\ for\ each\ i \in P_g; g := g+1$

}

While fitness of individuals in P_g not satisfied;

Return the individual from S ;

This proposed hybrid algorithm computes a reduced set without generating possible subsets of features. It initiates with a null set and subsequently it adds one at a time, in a cycle. A new population of chromosomes is constructed in a random manner. Fitness function for each chromosome is evaluated using the Eqn. (8) and further procedure is elaborately explained in Algorithm-1. By following this procedure, feature with highest fitness value is identified. All possible combinations of features related to the selected feature were also calculated. This process is an iterative optimization process until the best fitness is arrived; the process will be continued, updated and stored in set X . This reduced feature subset arrived based on the dependency of each feature subset calculated with respect to the dependency on decision feature.

6. Theoretical background of artificial neural network (ANN)

ANN is an information processing system alike biological neural systems. It resembles the mathematical models of the human brain. A typical ANN has three layers, namely, input layer, hidden layer, and output layer. There can be one or more hidden layers depending upon the number of dimensions of the training samples [15]. The structure and each element present in the structure of ANN can be learned from many Refs. [18-19]. One of the familiar types of ANN is the back propagation neural network (BPNN) algorithm. The proposed algorithm consists of training and testing process. Training process consist of three stages (1) Feed forward input (2) Arithmetic operations and error calculation in back propagation mode and (3) weight modification[16]. Procedure of BPNN algorithm is explained through following steps [17].

i. Initialize the weights with small random values.

ii. Transfer of input signal x_i to each input unit (Xi, i=1...n) and propagated to the units of hidden layers.

iii. ummation of each hidden unit (h_j, j=1...p) of its weighted input signals,

$$h\_in_j = V_{oj} + \sum_{i=1}^{n} x_i v_{ij}$$

iv. Comparing its output signal applying its activation function and it sends the signals to output unit layers

$$h_j = f(h\_in_j)$$

v. Summation of weighted input signals to each output unit (O_k, 1...m), and applies its activation function to compute its output signal.

$$o\_in_k = w_{ok} + \sum_{j=1}^{n} h_j w_{jk}$$

$$o_k = f(o\_in_k)$$

vi. Computation of its error information term, as each output unit (O_k, k=1...m) receives a target pattern corresponding to the input training pattern

$$\delta_k = (t_k - o_k) f'(o\_in_k)$$

vii. Calculation of its weight correction term (updated later) and sends to the next layers

$$\Delta w_{jk} = \alpha \delta_{khj}$$

$$\Delta w_{ok} = \alpha \delta_k$$

viii. Each hidden unit (Zj, j=1...p) sums its δ inputs,

$$\delta\_in_j = \sum_{k=1}^{m} \delta_k w_{jk}$$

ix. Eqn.(viii) is multiplied by the derivative of its activation function to calculate its error information term,

$$\delta_j = \delta\_in_j f'(h\_in_j)$$

x. Calculates its weight correction term, and calculation of its bias correction term,

$$\Delta v_{ij} = \alpha \delta_j x_i \qquad \Delta v_{oj} = \alpha \delta_j$$

xi. Weight Adjustment process has been done by updating of each output unit (Y_k, k=1...m)in its bias and weights (j=0...p):

$$w_{jk}(new) = w_{jk}(old) + \Delta w_{jk}$$

xii. Updates its bias and weights (i=0...n) each hidden unit (Zj, j=1...p) :

$$v = v_{ij}(old) + \Delta v_{ij}$$

xiii. Stopping the test condition

xiv. Repeat the above process again if test condition fails

7. Implementation of neural network and C4.5 algorithm

To improve the accuracy and efficiency of neural networks the data were scaled to a particular range, such as [0.0, 1.0]. Min-max normalization [18] which performs a linear transformation on the original data is used for this purpose. Main advantage of this normalization method is, it preserves the relationships among the original values. MATLAB R 2013 version software is used to implement the BPNN algorithm. In the proposed work the network consist one input layer with neurons equal to number of input features(Ten features) and one output layer with neurons equal to number of output states (4). There is some difficulty to decide the number of hidden neurons. It may depend on the number of input nodes, output nodes and the transfer function. Initial neuron number at hidden layer was calculated as 8 [19].

Levenberg -Marquardt algorithm is used to train the network. At Initial stage weights are randomly generated and had 10-8-4 topology for training, further numbers of hidden neurons were increased from 8 to 25 in the hidden layer to achieve the minimum error. Sigmoid function is used as the transfer function. The training data set carries 60% of total samples, in particular 4/5 of 60% is used for training and rest 1/5 of 60% is used for validation. The testing set comprises the other 40% of samples. For comparing the efficiency and accuracy of the data mining results decision tree based C4.5algorithm is used. During initialization of the decision tree; a training data set is divided based on certain decision rules until one subset match with particular state of class. Quinlan, (1993) [20] proposed the methodology of test feature selection criterion of decision tree had a significant importance and how to use information entropy evaluation function calculation based on the information theory. Readers are referred to the theoretical background and tutorials on C4.5 [7, 20] for details. J48 algorithms available in WEKA accomplishment of C4.5 Algorithm were used to perform classification process. Similar to ANN procedure the features selected from the previous step in the methodology divided into two parts as training and testing. Randomly about 60% were selected for training the model and the remaining 40% samples are used as testing set. 10-fold cross-validation process is employed to evaluate the classification accuracy.

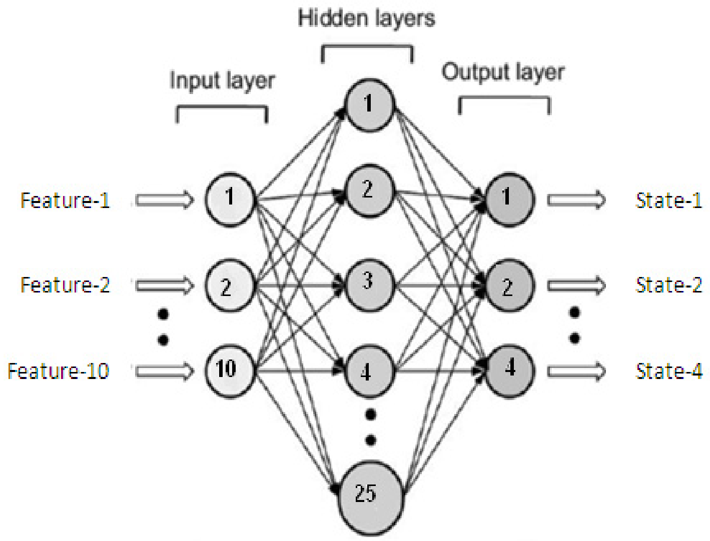

Fig.4. Artificial neural network structure.

Table 1. Discretized sample data set

Elements	p	q	r	s	T
1	-1	-1	1	0	1
2	1	0	-1	-1	0
3	0	1	1	1	-1
4	-1	0	0	1	1
5	1	1	0	-1	-1
6	1	0	-1	0	1
7	-1	-1	0	1	1
8	-1	1	1	1	-1

8. Conclusions

This paper deals with wavelet feature extraction as well introduced the hybridized approach i.e., Genetic algorithm (GA) and rough set (RS) to make an optimum selection of features in gear fault diagnosis. In general, classification accuracy measures are used to evaluate the performance of the classifiers. The experimental result (Table-2 & Fig 4) of the proposed Hybrid methodology shows significant increase in the predictive classification accuracy in the fault monitoring system. When comparing the classification results of hybrid BPNN and hybrid C4.5, The hybrid BPNN approach gives more classification accuracy (98.75%). But in terms of computation time hybrid C4.5 shows better results(97.5%). Hence the results of the classification accuracy have proved the effectiveness of the proposed method for fault identification of the gear over the existing method.

Table 2. Performance comparison of classifiers with and without feature selection

Classifier	Prediction accuracy of testing data (%)					Average Accuracy on testing data (%)	Average computing time for training (sec)
	NO.. OF INPUT FEATURES	STATE-1	STATE-2	STATE-3	STATE-4		
BPNN(no feature reduction)	10	100	90	85	92.5	97.5	280.78
BPNN(feature reduction with GA)	5	100	95	100	93	97	187.42
BPNN(feature reduction with rough set based GA)	4	100	100	95	100	98.75	131.37
C4.5(no feature reduction)	10	100	92	85	95	93	0.072
C4.5(feature reduction with GA)	5	100	90	100	97	96.75	0.0031
C4.5(feature reduction with rough set based GA)	4	100	90	100	100	97.5	0.0012

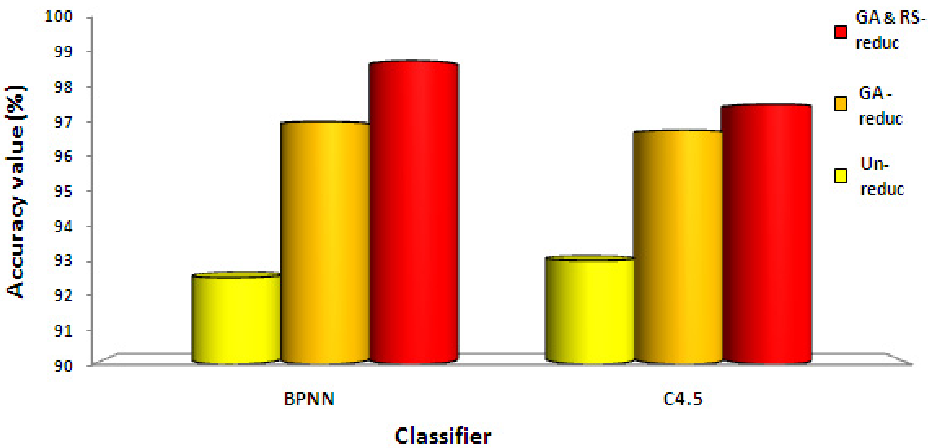

Fig.5. Classification accuracy value (%) of various classifiers

As a future work this experimental diagnosis in feature selection process will be extended to hybridize other evolutionary algorithms such as Ant colony optimization, artificial bee colony optimization (ABC) with rough set techniques.

References

[1] P.D. McFadden (1986) Detection fatigue cracks in gears by amplitude and phase demodulation of meshing vibration, J. of Vib and Acoustics 108 (4) 165–170.
[2] J.I. Taylor, Back to the basics of rotating machinery vibration analysis, Sound Vib. 29 (2) (1995) 12–16.
[3] J.F.A. Leblanc, J.R.F. Dube and B. Devereux, (1990)Helicopter gearbox vibration analysis in the Canadian forces — applications and lessons, Proc. of the first int. conf., gearbox noise and vibration, Cambridge, UK, , pp. 173–177.
[4] J. Lin, M.J. Zuo, (2003) Gearbox fault diagnosis using adaptive wavelet filter, Mech.Systems and Signal Processing 17 (6) 1259–1269.
[5] Tse, P. W., Wen-xian, and Tam, H. Y., (2004). Machine fault diagnosis through an effective exact wavelet analysis . J. of Sound and Vibration, (277), pp. 1005–1024.
[6] H. Liu, H. Motoda., (2007) Computational Methods of FeatureSelection, Chapman and Hall/CRC Press, USA.
[7] Weixiang Sun_, Jin Chen, Jiaqing Li-Decision tree and PCA-based fault diagnosis of rotating machinery- Mech. Systems and Signal Processing 21 (2007) 1300–1317.
[8] Ngoc-Tu Nguyen, Hong-Hee Lee* and Jeong-Min Kwon, Optimal feature selection using genetic algorithm for mechanical fault detection of induction motor J. of Mech. Sci and Tech 22 (2008) 490~496.
[9] V. Sugumaran, V. Muralidharan and K. I. Ramachandran, Feature selection using Decision Tree and classification through Proximal Support Vector Machine for fault diagnostics of roller bearing, Mech. Systems and Signal Processing 21 (2007)930-942.

[10] Samanta, B., 2004. “Gear fault detection using artificial neural networks and support vector machines with genetic algorithms”. Mech. Systems andSignal Processing (18),pp. 625–644.
[11] N. Tandon and BC.Nakra, "Detection of defects in rolling element bearings by vibration monitoring," J. Instn of Engrs (India) — Mech Eng 73,1993, p.271–82.
[12] R.A. Collacott, Mechanical Fault Diagnosis and Condition Monitoring, Chapman & Hall, 1977.
[13] N. Saravanan, V.N.S. Kumar Siddabattuni, K.I. Ramachandran, Fault diagnosis of spur bevel gear box using artificial neural network (ANN), and proximal support vector machine (PSVM), applied soft computing 10 {2010} 344-360.
[14] Mallat, A Wavelet Tour of Signal Processing, Academic Press, 1998.
[15]M.T. Hagan, H.B. Demuth, M.H. Beale, Neural Network Design, PWS Publishing Company, 1996.
[16] K. Hornik, Approximation capabilities of multilayer feed forward networks, Neural Networks 4 (2) (1991) 251–257.
[17] Fausett L. Fundamentals of Neural Networks Architectures, Algorithms and Applications , Pearson Education: India, 2004.
[18] Hosseini , Abbas , “Neural network approach for estimation of hole-diameter in thin plates perforated by spherical projectiles”, Thin-Walled Structures, Vol 46, Issue 6, June, pp. 592-601(2008).
[19] L.M. Zhang, Models and Applications of Artificial Neural Networks, Fudan University Press, 1992.
[20]Quinlan, J. R. (1993). C4.5: Programs for machine learning. San Francisco, CA: Morgan Kaufman.
[21]N. Rami, N. Khushaba, A. Al-Ani, A. Al-Jumaily, Feature subset selection using differential evolution and a statistical repair mechanism, in: Expert Systems with Applications, Elsevier, 2011, pp. 11515–11526.
[22]G. Zhang, L. Hu, W. Jin, Discretization of continuous attributes in rough set theory and its application, Computational and Information Science Lecture Notes inComputer Science 3314 (2005) 1020–1026.
[23]Ngoc-Tu Nguyen, Hong-Hee Lee* and Jeong-Min Kwon , Optimal feature selection using genetic algorithm for mechanical fault detection of induction motor, Journal of Mechanical Science and Technology 22 (2008) 490~496.
[24]C. Velayutham, K. Thangavel, Unsupervised quick reduct algorithm using rough set theory, J. of Elect Sci and Tech. 9 (2011) 193–201.