# Vibration signal based fault diagnosis of gears using ensemble empirical mode decomposition and linguistic hedges neural fuzzy classifier with selected features

**S.Devendiran[1], K.Manivannan[1], Arun tom mathew[1] & C.Rajeswari[2*]**
[1]School of Mechanical Engineering, VIT University, Vellore, Tamil Nadu, India
[2]School of Information Technology and Engineering, VIT University, Vellore, Tamil Nadu, India
***Corresponding author: E-Mail: rajeswari.c@vit.ac.in**

**ABSTRACT**

The gearbox is vital parts on most types of machinery for vary the shaft speed, torque and the power. Gear trains are considered to be among the earliest machine elements. Their operating state directly affects the machine performance, efficiency and life. Therefore, fault identification of gear has been the subject of extensive research. The vibration signals are acquired using accelerometer, under healthy and simulated faulty gear conditions from the test rig. In this study, the acquired signals are processed using EEMD (Ensemble empirical mode decomposition) and Linguistic Hedges Adaptive Neural Fuzzy Classifier with Selected Features (LHANFCSF) is presented for diagnosis of gear health monitoring. The performance evaluation of this system is estimated by using classification accuracy and k-fold cross-validation. The results indicated that the classification accuracy without feature selection was lesser when compare to after applying feature selection algorithm. The obtained classification accuracy of LHANFCSF with feature selection was very promising with regard to the other classification applications such as hidden Markov model (HMM) and back propagation neural network (BPNN) for this problem.

**KEY WORDS:** Test rig, fault diagnosis, wavelet, Ensemble empirical mode decomposition, Linguistic Hedges Adaptive Neural Fuzzy Classifier with Selected Features, hidden Markov model and back propagation neural network.

## 1. INTRODUCTION

Gearbox is one of the complex machinery and is a critical component in mechanical power transmission system. Gearboxes have wide applications in automobile, cement, petrochemical, power, paper & pulp, steel and sugar industries. The gear drives are the most effective means of transmitting power in machines due to their high degree of reliability and compactness. The gears themselves are the most important elements in the gearbox, and the degree of wear and fatigue to which they are subjected even under normal operating conditions means that they are often subject to premature failure. Mc Fadden (1986), investigated fatigue cracks in gears by amplitude and phase demodulation of meshing vibration and mention gear health condition is directly proportional to the performance of the machinery. (Meng, 1991), presented that; any real world signal can be broken down into a combination of unique sine waves. Every sine wave separated from the signal appears as a vertical line in the frequency domain. Its height represents its amplitude and its position represents the frequency. The frequency domain completely defines the vibration. Frequency domain analysis not only detects the faults in rotating machinery but also indicates the cause of the defect. (Staszewski, 1994), were applied wavelet transform to waveform data analysis in fault diagnostics of gears and carried out the fault diagnosis. (Tian, 2003), introduced an adaptive wavelet filter based on Morlet Wavelet, the parameters in the Morlet wavelet function are optimized based on the kurtosis maximization principle. The adaptive wavelet filter is found to be very effective in detection of symptoms from vibration signals of a gearbox with early fatigue tooth crack. (Elforjani, 2012), discussed about the Condition monitoring of key components in rotating machines such as gearboxes ensure reduction in costly unscheduled machine down time and explores the possibility of monitoring seeded defects on worm gears with vibration analysis. Unlike other types of gearboxes, monitoring of worm gearboxes is not widely documented. In automated decision making condition monitoring system, after the signal acquisition and extracting fault features from it, it is necessary to apply decision making process to determine the gear status. There are different algorithms for decision making. The most commonly used algorithms are artificial neural networks and fuzzy clustering. However, designing and training of these algorithms need a lot of data by Paul (2001). In some recent works, several combinations of wavelet transform, Wigner Ville Distribution (WVD) and other time-frequency methods with decision making methods such as ANN and fuzzy logic have been proposed for gear fault detection by (Yaguo, 2010; Saravanan, 2010). (Subrahmanyam, 1997), were compared the performance of a multilayer feed-forward with supervised training with that of an Adaptive Resonance Theory (ART-2) based network with an unsupervised training algorithm. A collection of features, including Kurtosis, RMS, peak values of time and high frequency domains, and peak values of autocorrelation are chosen as monitoring indices. (Wang, 2004), introduced three reference functions, based on wavelet transform, beta Kurtosis, and phase modulation for gear system monitoring. The developed neurofuzzy classifier provides a robust diagnosis for gear systems. According to the nonstationary characteristics of bearing fault vibration, a diagnosis method based on the Empirical Mode Decomposition (EMD) energy entropy, has been reported by Yu Yang (2006). An ANN, with the input features extracted from different frequency bands of the EMD, can accurately identify the localized fault pattern. ANN, support vector machine (SVM) and Fuzzy classifier are widely used as classification tool and reported in literature (Samanta, 2004). In the recent past reports of fault diagnosis of critical components such as bearings using machine learning algorithms like C4.5, SVM, PSVM are reported by Sugumaran (2006). (Saravanan, 2008), implemented decision tree for selecting best statistical features that will discriminate the fault conditions of the gear box from the signals extracted. A rule set is formed from the extracted features and fed to a fuzzy classifier. This paper also presents the usage of decision tree to generate the rules automatically from the feature set. Again (Saravanan, 2009), deal fault diagnosis of spur bevel gear box using statistical feature vectors from Morlet wavelet coefficients it is and classified using J48 algorithm and the predominant features were fed as input for training and testing multiclass proximal support vector machine the efficiency and time consumption in classifying the twenty four classes all-at-once is reported. (Sun, 2007), used decision making scheme beyond conventional fault testing. They proposed a new method based on C4.5 decision tree and principal component analysis (PCA). It was found that compared to BPNN C4.5 extracts knowledge quickly from the testing and is even superior to neural networks. (Yaguo, 2009), used a Two Stage Feature Selection and Weighting Technique (TFSWT) via Euclidian Distance Evaluation Technique (EDET) to select sensitive features and remove fault unrelated features. They used a Weighted K Nearest Neighbour (WKNN) classification algorithm to identify the gear crack levels. (Yang, 2015), presented the methodology utilise artificial bee colony algorithm is used for SVM parameter optimization of gearbox fault diagnosis, compared with genetic algorithm, the particle swarm optimization and found that the accuracy of artificial bee colony algorithm is higher. (Rajeswari, 2015), utilized ensemble empirical mode decomposition for signal processing and feature extraction, hybrid binary bat algorithm for feature selection and machine learning algorithms for classification purposes in gear fault diagnosis. In the present work, brief review is given which is strictly connected to the subject of this paper. Further relevant the experimental procedure and theoretical background of EEMD, ADAPTIVE NEURAL FUZZY CLASSIFIER BASED ON LINGUISTIC HEDGES, ANN, HMM is presented. Analysis of simulated data according to the presented procedure is also presented. Finally the results for the same are presented and Last section contains conclusions.

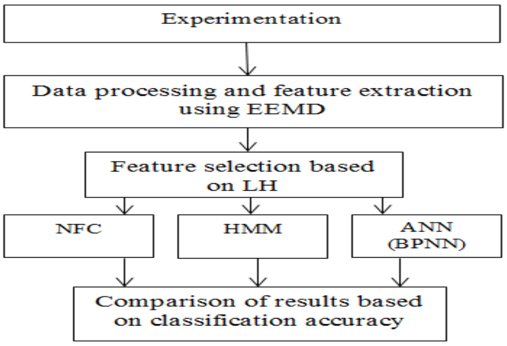



**Fig.1. Methodology**

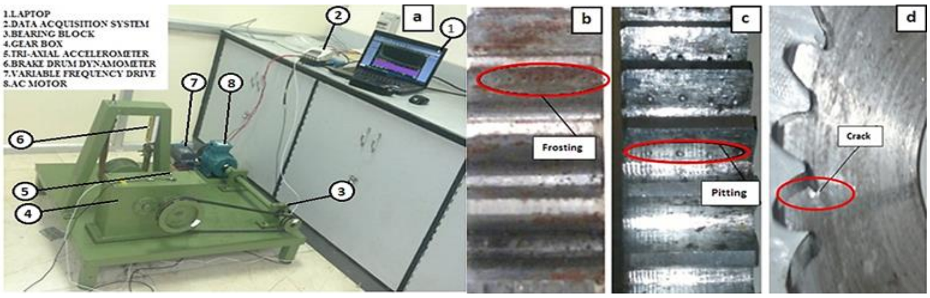



**Fig.2. Experimental setup and artificially induced gear faults**

**Experimental procedure:** Experimental set-up is shown in Figure 2 consists of three phase 0.5 hp AC motor, VFD used to control the speed of the motor, gearbox containing gear and pinion connected by means of belt drive. SAE 40 oil was used as a lubricant in the gearbox. A brake drum dynamometer setup has been connected to the gear box to control the load. The gears used in the gear box are made of 045M15 steel wherein the spur gear has 36 teeth and pinion having 24 teeth The spur gears used for this experiment had a module of 3mm and a pressure angle of 20°. Different gear condition such as normal, fault1 (frosting), fault2 (pitting), fault3 (crack) is artificially created. Tri-axial accelerometer (Vibration sensor) is fixed on gearbox to measure the signals. The accelerometer sensor is connected to data acquisition system for acquiring the data. Rotational frequency of the pinion was 28 Hz which resulted in gear meshing. Separate samples are collected for each gear condition with the sampling frequency of 12800Hz for 10 seconds and divide the total signals into small sample package which carry 6000 data points. This sample packages used for further analysis. The Fig. 1 depicts the methodology of proposed work. Vibration signals which carries the information about the condition of the component are acquired for all the four conditions of gear conditions in the present research work. Each signal is separated into data sets signals, and they are used to extract the statistical features. EEMD is employed to decompose the data into number of IMF's and the statistical features is extracted using the IMF's from both time domain and frequency domain. These features are used for further process of classification with and without feature selection.

**Theoretical background of EEMD:** EMD decompose a signal into number of characteristic method of functions called IMF's (Huang, 1998). An IMF needs to fulfil the accompanying criteria's. The quantity of extreme and zero intersection should either be equivalent to one or vary by one at most. The mean estimation of the envelope characterized by both neighbourhood maxima and nearby minima is equivalent to zero. The fundamental issue happens in EMD is that it experiences mode mixing. It is a consequence of sign intermittency. Intermittency not just aims alliancing issues in time-recurrence conveyance yet it additionally makes the importance of individual IMF indistinct. To defeat the downside of EMD, Wu (2009), presented a strategy named Ensemble exact mode deterioration which lessens the issue of mode blending. The enhanced rendition of EMD is EEMD and additionally it is a renowned instrument for a non-direct and non-stationary signal preparing.

**EEMD algorithm:** a) The number of ensemble S needs to be initialized, b) The amplitude of the added numerically generated white noise needs to be given, and i=1, c) In order to generate a new signal add a numerically generated white noise with the given amplitude to the original signal $x(t)$.

$$x_i(t) = x(t) + n_i(t) \tag{1}$$

Where $n_i(t)$ represents the i-th added white noise series, and $x_i(t)$ denotes the noise-added signal of the i$^{th}$ trial, while i=1,2,….M.

d) To decompose the newly generated signal into IMFs original EMD algorithm needs to be used.

$$x_i(t) = \sum_{s=1}^{S} c_{i,s}(t) + r_{i,s}(t) \tag{2}$$

Where S is the number of IMFs, $r_{i,s}(t)$ is the final residue, which is the mean trend of the signal, and $c_{i,s}(t)$ represents the IMFs $(c_{i,1}, c_{i,2}, ....c_{i,s}....c_s)$ which include different frequency bands ranging from high to low.

e) Repeat steps 3 and 4 S times with a different white noise series each time to obtain an ensemble of IMF

$$\{C_{1,s}\}, \{C_{2,s}(t)\}, ...., \{C_{M,s}(t)\} \tag{3}$$

f) Calculate the ensemble means of the corresponding IMFs of the decomposition as the final result:

$$c_s(t) = \frac{1}{M}\sum_{i=1}^{M} c_{i,s}(t) \tag{4}$$

Where $c_{s,t}$ is the s-th IMF decomposed by EEMD, while i=1,2,…,M, and s = 1, 2,….S.

EEMD is an improved version of original EMD and a more matured tool for a non-linear and non-stationary signal processing techniques. The principle of the EEMD is that the added white noise populates the whole time-frequency space uniformly. It facilitates a natural separation of the frequency scales, which reduces the occurrence of mode mixing. The flow chart of EEMD algorithm is given in Fig. 3 and followed by its corresponding algorithm. To the EEMD the starting information from step 1 of calculation is given. The information is the acquired raw signal information of all the four distinctive gear vibrations, in particular Normal, Spalling, Pitting and Crack. There are 4 classes to be analyzed. Arbitrarily chosen 6000 time domain data point information from every class is given as input to EEMD. The IMF is the contrast between the envelopes and the mean separation between the envelopes. In the comparable way IMF's are produced totally and deterioration of the signal is accomplished for each of the four classes. These IMF's are useful parameters to accomplish the time domain and frequency domain features. Hilbert Huang Transform system is utilized to plot the IMF's and comparably FFT is plotted for the comparing IMF's of Normal, spalling, pitting and the crack gears. The features acquired from the IMFs in time domain and frequency domain are used for further process.

**Statistical feature extraction using EEMD:** Statistical feature extraction is an important step in machine fault diagnosis. When the gear fault occurs due to non-stationary signal variations; amplitude, time domain and frequency spectrum distribution of fault gear may be different from those of normal gear and also creates the new frequency components. Ten features are listed in Table 1 and out of these ten, first five features indicate the time-domain statistical characteristics, and remaining five features indicate frequency-domain statistical characteristics. Feature Tf1– Tf5 gives the time domain vibration amplitude and energy. Feature Ff1 gives the information about frequency domain energy and convergences of the spectrum power are described by Ff2– Ff5 (Lei, 2007).

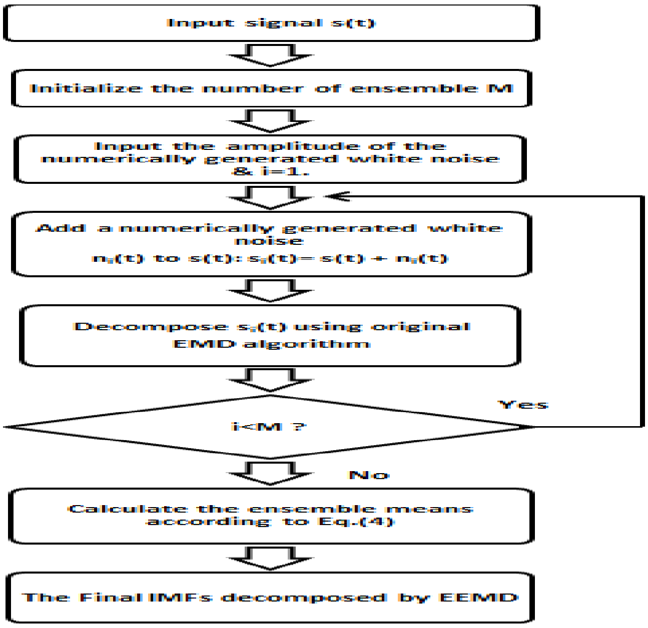



**Fig.3. Flow chart of EEMD**

**Table.1. Statistical parameters**

| Time Domain Features | Frequency Domain Features |
|---|---|
| $T\,f_1 = \frac{\sum_{n=1}^{N} y(n)}{N}$ | $F\,f_1 = \frac{\sum_{k=1}^{K} z(k)}{K}$ |
| $T\,f_2 = \sqrt{\frac{\sum_{n=1}^{N} (y(n))^2}{N}}$ | $F\,f_2 = \frac{\sum_{k=1}^{K} (z(k)-Ff_1)^2}{K-1}$ |
| $T\,f_3 = \left(\frac{\sum_{n=1}^{N} \sqrt{\lvert y(n) \rvert}}{N}\right)^2$ | $F\,f_3 = \frac{\sum_{k=1}^{K} (z(k)-Ff_1)^3}{K(\sqrt{Ff_2})^3}$ |
| $T\,f_4 = \frac{\sum_{n=1}^{N} \lvert y(n) \rvert}{N}$ | $F\,f_4 = \frac{\sum_{k=1}^{k} (z(k)-Ff_1)^4}{K(\sqrt{Ff_2})^2}$ |
| $T\,f_5 = \frac{\sum_{n=1}^{N} (y(n)-Tf_1)^3}{N}$ | $F\,f_5 = \sqrt{\frac{\sum_{k=1}^{K} (F_k-Ff_5)^2 z(k)}{K}}$ |

**Adaptive neural fuzzy classifier based on linguistic hedges:** The concept of Linguistic Hedges (LH) based on fuzzy feature selection system were exhibited by Cetisli (2010). The Liguistic hedge qualities called as LH can be utilized to demonstrate the significance level of fuzzy sets. For classification, each classes were characterized by a fuzzy classification rule. The features that contain the LHs values are connected near to focus values the features are thought to be significant and chosen. If the linguistic hedge estimations of features are near to dilation values this implies that these features are insignificant and needs to be dismisses. Hilarious highlights can be wiped out as indicated by the LHs estimation of highlights. In this procedure, if LH values of classes in any feature are greater than 0.5 and near to 1 or more noteworthy, this highlight is important, else it is unimportant. The program makes a feature determination and a dismissal rule was done by utilizing power values of features. There are chiefly two determination criteria, one is the choice of features that have the greatest hedge value for any class and the other is the selection of features that have a bigger hedge values for each class, on the grounds that any feature can't be particular for each class. For that reason, a selective function should be described from the hedge values of any feature as,

$$p_j = \prod_{i=1}^{k} p_{ij} \tag{5}$$

Where Pj denotes the selection value of the j$^{th}$ feature, and K is the number of classes. For forcing the hedge values to binary values, the initial values of hedges are taken as 0.5. After the tuning hedges, if the hedge value of any feature increases to one, the feature is selective for belonging class. If the hedge value of any feature decreases to zero, the feature is irrelevant for belonging the class. The same feature selection and classification algorithm is discussed more detailed in [23]. In adaptive neuro-fuzzy classifier models, k-means algorithm is used to initialize the fuzzy rules. Also, Gaussian membership function is only used for fuzzy set descriptions, because of its simple derivative expressions. Adaptive Neuro-Fuzzy Classifier (ANFC) with Linguistic hedges is based on fuzzy rules. Linguistic hedges are applied to the fuzzy sets of rules, and are adapted by Scaled Conjugate Gradient (SCG) algorithm. By this way, some distinctive features are emphasized by power values, and some irrelevant features are damped with power values. The power effects in any feature are generally different for different classes. The using of linguistic hedges increases the recognition rates. The pseudo code of Adaptive neuro-fuzzy classifier is given in this section. Figure 4 gives the schematic representation of adaptive neuro-fuzzy classifier (Cetisli, 2010) architecture.

| **Algorithm : ANFC** | |
|---|---|
| i. | Set the number of fuzzy rules (U) for every class .Then the total fuzzy rules are V=U.K, where U is the number of fuzzy rules. |
| ii. | Set $P_{ij=}1$,for i=1,2,…..,V and j=1,2,……,D |
| iii. | Determine the initial value of nonlinear parameters of ANFC-LH by using K=means clustering. |
| iv. | Train the ANFC-LH with $X_{new}$ Training set in training, $P_{ij}$ value should be equal to or bigger than zero for every feature and fuzzy rules ($P_{ij}\geq 0$). |
| v. | Obtain the training and testing classification results. |

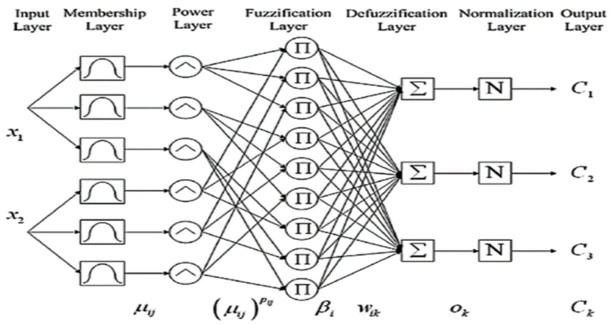



**Fig.4. Neuro-fuzzy classifier architecture. (Cetişli, 2010)**

**Theoretical background of ANN:** ANN comprises of 3 layers specifically input, hidden and output layer (Samanta, 2004). ANN is a feed forward system with sigmoid hidden and output neurons. The system utilized as a part of ANN is prepared with scaled conjugate back propagation algorithm. One of the common kind of ANN is the back propagation neural network (Hecht – Nielsen, 1989). BPNN 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 (Hornik, 1991). The network consists of an input layer with neurons equal to number of input features and one output layer with neurons equal to number of output states. 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. Procedure of BPNN algorithm is explained elaborately in Laurene (1994). The input and output layer comprise of 10 features and 4 Classes yet in BPNN it's exceptionally hard to recognize the number of neurons exhibit in the Hidden layer. The number of neurons in the hidden layer will be expanded to 25 to accomplish the minimum error value. Hidden layers depend upon on input and output data. The fundamental elements of 3 layers introduce in ANN are as follows. Input layers get signals from outer source, handling of signals is finished by a hidden layer and an output layer that sends the signals handled by a hidden layer back to the external world and the same is depicted in Fig. 5.

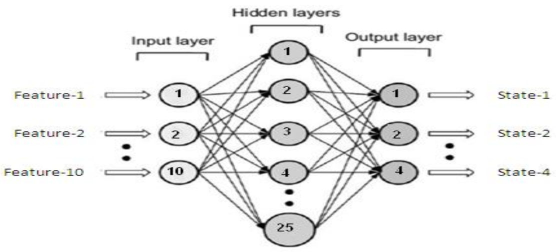



**Fig.5. Structure of ANN**

**Theoretical background of HMM:** HMM has been utilized as a part of a mixed bag of utilizations like Speech acknowledgment, Text preparing, Bio-informatics, Financial. HMM is fundamentally a machine learning technique and makes utilization of static machines. HMM is basically utilized as a part of an issue having consecutive steps. Three issues must be comprehended for HMM to be helpful in certifiable applications and they are evaluation, decoding and learning (Purushotham, 2005). Despite the fact that HMM does not give careful information about the issue to be explained, Complete displaying and learning of arrangements and it should be possible by HMM itself. In view of the probability density distribution quickly called as PDD, HMM can be isolated into two models specifically continuous and discrete. HMM comprise of taking after segments a set of states (a's), a set of possible output symbols (b's), a state transition matrix (c's), Output emission matrix (x's), Initial probability vector. HMM is spoken to by a diagram structure that comprises of N nodes, called hidden states, and arcs that represent transition between nodes.

Set of hidden states:

$$s = \{s_1, s_2, \ldots s_N\} \quad (6)$$

Where N is the number of states in HMM.

State transition probability distribution:

$$A = \{a_{ij}\}, \quad \text{Where} \quad a_{ij} = P[q_{t+1} = S]\ [q_t = S_t], \text{ for } 1 \le i, j \le \text{N}, \quad (7)$$

Set of observation symbols

$$V = \{v_1, v_2, \ldots v_m\} \quad (8)$$

Where M is the number of observation symbol per state.

Observation symbol probability distribution:

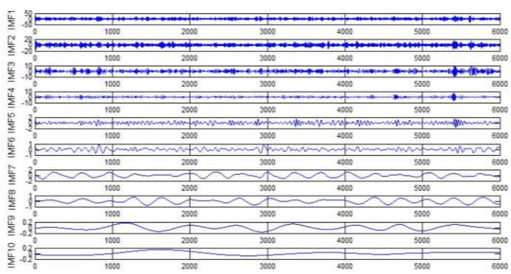
$$B = \{b_j(k)\}, \text{ where } b_j(k) = \text{P } [v_k \text{ at } t \,|\, q_t = S_j] \text{ for } 1 \le j \le N,\ 1 \le k \le M \quad (9)$$

Initial state probability distribution:

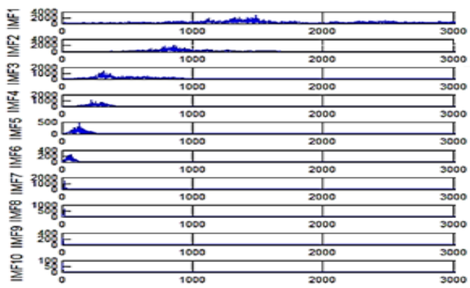
$$\pi = \{\pi_i\}, \text{ where } \pi_i = \text{ P} \,|\, q_1 = S_i], \text{ for } 1 \le i \le N \quad (10)$$

Where $q_t$ represents the hidden state at time $t$. An HMM can be represented by the compact notation $\lambda = (A, B, \pi)$. HMM modeling involves choosing the number of hidden states, N, the number of discrete symbols, M, and the specification of three probability distributions A, B, and π.

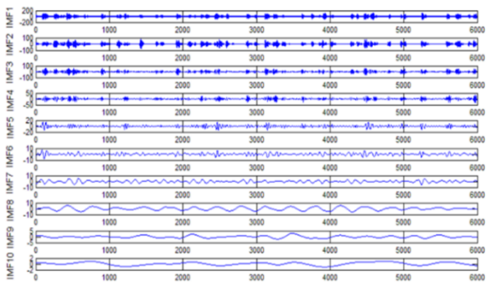
**Experimental results and discussion:** At the time of data acquisition, rotation speed of the motor is 1000 rpm (16.67 Hz), the rotation speed of the gear is 11.11 Hz, and the mesh frequency is 522.24 Hz. The gear signals are extracted in the sampling rate of 12800 Hz (6400 data points per second). For 20 seconds 128000 data points were collected through accelerometer for each condition of gear. Each condition of gear signals are split into approximately 20 samples (each sample contains 6000 data points) and there are all together 80 data samples were collected.



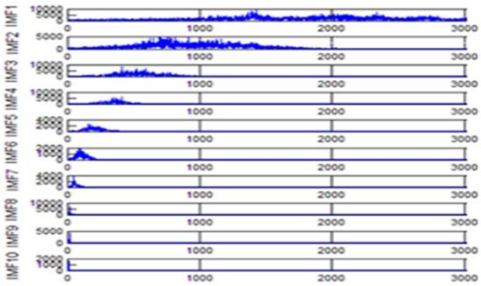
**(a) Normal gear EEMD results x axis: sample no.; y axis: amplitude**



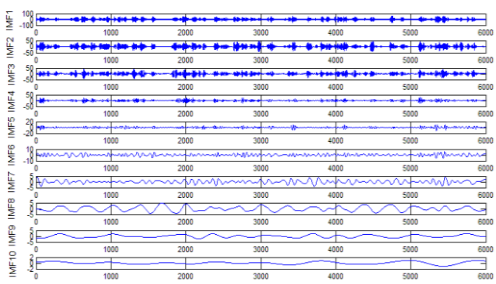
**(b) Corresponding spectrum of normal gear EEMD result, x axis: frequency (Hz).; y axis: amplitude**



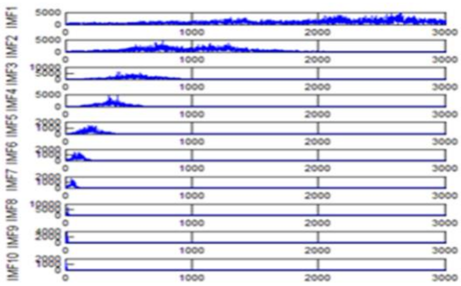
**(c) Crack gear EEMD result, x axis: sample no.; y axis: amplitude**



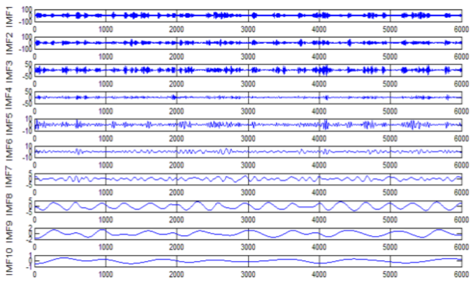
**(d) Corresponding spectrum of Crack gear EEMD result, x axis: frequency(Hz).; y axis: amplitude**



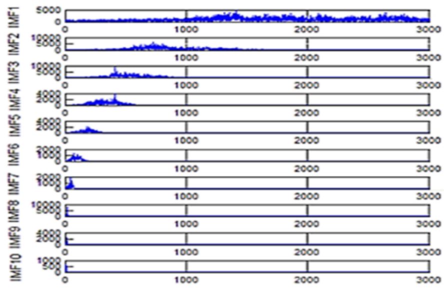
**(e) Pitting gear EEMD result, x axis: sample no.; y axis: amplitude**



**(f) Corresponding spectrum of pitting gear EEMD result, x axis: frequency(Hz).; y axis: amplitude**



**(g) Frosting gear EEMD result, x axis: sample no.; y axis: amplitude**



**(h) Corresponding spectrum of frosting gear EEMD result, x axis: frequency (Hz).; y axis: amplitude**

**Fig.6. EEMD signal and corresponding spectrum of four states of gear (Continued)**

Each sample was fed into FFT for further processing. From both the Time Domain Signals and Frequency Domain Signals, direct categorization and a difference among the four states of gears are understood and have difficulty to identify the variation among them because of the noise present in the original signal of the four running conditions of gear. By considering the present facts, each vibration sample (original signal) is decomposed by EEMD. Initially in EEMD, two important parameters has to be set; the ensemble number M and the amplitude of white noise *i*. In general, an ensemble number of a few hundred will lead to a good result, and the remaining noise would cause negligible percent of error if the added noise has the standard deviation that is a fraction of the standard deviation of the input signal. For the standard deviation of the added white noise, it is suggested to be about 20% of the standard deviation of the input signal. Hence the two parameters of EEMD are set as M=100 and *i*=20%. After setting the parameters, the signals were decomposed into 'n' IMFs and one residue according to nature of the signal. For our case, IMF component decomposition identifies eleven modes: IMF 1- IMF 10 and one residue were arrived and its corresponding spectrums also arrived for four conditions of gear and it is depicted in Fig. 6 (a) to Fig. 6 (h)

In the corresponding frequency spectrum of each IMF's say mode 1(frequency spectrum of IMF1) contains the highest signal frequencies, mode 2 the next higher frequency band and so on. The vibration change caused by a localized damage at its early stage, is usually weak and contaminated by noise, so that early fault diagnosis is more difficult and needs more complicated methods. In the time domain, a localized gear fault causes amplitude and phase modulation of the gear meshing vibration which will not be deliberately seen many times; while in the frequency domain, these modulations appeared as a series of sidebands around the gear mesh frequency and its harmonics and this procedure is followed in the past decades. In automated fault diagnosis methodology, with the help of this knowledge, the informative features are collected from the range of characteristic (gear mesh) frequencies will give more prediction accuracy. In this aspect the selection of IMFs for further processing is based on modes of decomposition. Mode 1 and mode 2 accommodate the characteristic frequency (gear mesh frequency). Modes 7 and 8 are associated with the harmonic of the rotational frequency of the input shaft. For second condition of gear is shown in Fig. 6 (d) mode 2 and mode 3 is centred between the ranges of 250 Hz to 1250 Hz, which can be clearly associated with the characteristic gear mesh frequency of the component. Similarly for third and fourth condition as shown in Fig. 6 (f) and Fig. 6 (h), mode 2 is positioned on the same range of values. From this inference it can be easily proven that the EEMD decomposes the vibration signals very effectively on an adaptive method. When compared to raw time and frequency signals, IMFs in both the domains are clearer even if it is hard to find the typical fault characteristics and it can also distinguish the four running conditions. Therefore the proposed intelligent based methodology is necessary to diagnose gear faults. Subsequently, 5 time domain and 5 frequency domain features are calculated only from IMF1 to IMF5 for each state of gear signal because of obvious characteristic and high signal energy present in first four IMFs. All the extracted features are normalized before given as input to the feature selection algorithms and subsequent classification process.

In order to reduce the dimensionality of the features Linguistic Hedges feature selection process were carried out and the results of the selected feature's description and time taken for the process are shown in the Table.2. The feature selection as well classification processes are done in MATLAB platform. The input features are categorized into two types: Time and frequency domain features are extracted from IMFs of EEMD and selected features from the same are separately fed input into MATLAB embedded classifiers ANN, MATLAB coded Neural Fuzzy Classifier and Waikato Environment for Knowledge Analysis (WEKA) embedded HMM classifier to identify different states of gear through classification process. WEKA is open source software issued under General Public License used for data mining. To measure and investigate the performance of the classification algorithms 75% feature data is used for training and the remaining 25% for testing purpose. Table 3 summarizes the results based on accuracy and time taken for each simulation. The significant features are selected based on the largest linguistic hedge values as shown in Fig. 8. According to the feature selection algorithm, Tf2 (Feature 2), Tf4 (Feature 4), Tf5 (Feature 5), Ff1 (Feature 6) and Ff2 (Feature 7), are common relevant features for each class. Tf1Feature 1), Tf3 (Feature 3), Ff2 (Feature 7), Ff3 (Feature 8), Ff4 (Feature 9) and Ff5 (Feature 10) is irrelevant for each class. It can be seen from Table 2 that spiteful class is easily notable from the other class. The group of accurately measured input data is the basic requirement in order to obtain an accurate model. The classification process starts by obtaining a data set (input-output data pairs) and dividing it into a training set used to train the neuro fuzzy classifier and testing data set is used to verify the accuracy and effectiveness of the trained neuro fuzzy classifier. The choice of the suitable cross validation method to be employed in the neuro fuzzy classifier is based on a trade-off between maximizing method accuracy and stability and minimizing the operation time. K-fold cross-validation was used for better reliability of test results (Francois, 2007) and repeated k times (the 'folds'). The average of the k results gives the validation accuracy of the algorithm (Diamantidis, 2000).In the first phase, neuro fuzzy classifier is trained using all data instances without feature reduction and in second phase neuro fuzzy classifier is trained using all data instances with LH based feature reduction In this study, 60–40% partition was used for training-test of the gear fault diagnosis. The error convergence curve of NFC achieved mean RMSE values of 3.89541 e $^{-36}$ in the training phase as shown in Fig. 7. In the validation phase, 4-fold cross validation is used to compute the recognition rates.

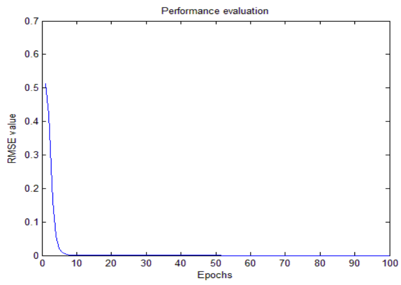



**Fig.7. Performance Evaluation of NFC without feature reduction**

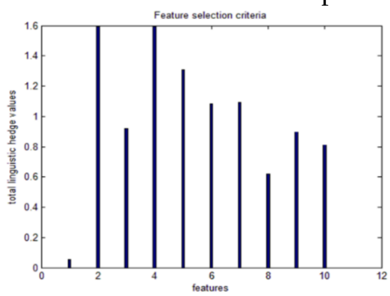



**Fig.8. Features selected for classification**

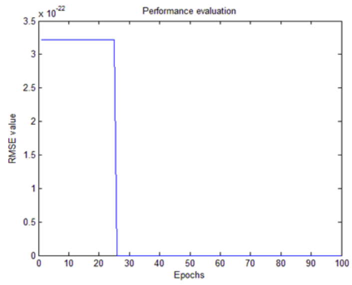



**Fig.9.Performance evaluation of NFC with feature selection**

**Table.2. The LH values of Gear dataset for every class and every feature**

| Class/ Features | $Tf_1$ | $Tf_2$ | $Tf_3$ | $Tf_4$ | $Tf_5$ | $Ff_1$ | $Ff_2$ | $Ff_3$ | $Ff_4$ | $Ff_5$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Normal | 0.047 | 1.3 | 0.353 | 1.075 | 1.118 | 1 | 1.01 | 0.405 | 0.6 | 0.519 |
| Spalling | 0.015 | 0.018 | 0.25 | 0.126 | 0.107 | 0.123 | 0.1 | 0.255 | 0.197 | 0.189 |
| Pitting | 0.011 | 0.125 | 0.109 | 0.057 | 0.125 | 0.05 | 0.025 | 0.15 | 0.035 | 0.073 |
| Crack | 0.015 | 0.157 | 0.222 | 0.242 | 0.036 | 0.007 | 0.055 | 0.001 | 0.072 | 0.024 |
| Total LH values | 0.088 | 1.6 | 0.934 | 1.5 | 1.386 | 1.18 | 1.19 | 0.811 | 0.904 | 0.805 |

The number of fuzzy rules is determined according to the number of classes. The classification rules are expressed for each class, the rules are:

**Rule 1:** IF $Tf_1$ is $A_{11}$ with $P_{11}$ = 0.047 AND $Tf_2$ is $A_{12}$ with $P_{12}$ = 1.300 AND $Tf_3$ is $A_{13}$ with $P_{13}$ = 0.353 AND $Tf_4$ is $A_{14}$ with $P_{14}$ =1.075 AND $Tf_5$ is $A_{15}$ with $P_{15}$ = 1.118 AND $Ff_1$ is $A_{16}$ with $P_{16}$ = 1.000 AND $Ff_2$ is $A_{17}$ with $P_{17}$ = 1.010 AND $Ff_3$ is $A_{18}$ with $P_{18}$ = 0.405 AND $Ff_4$ is $A_{19}$ with $P_{19}$ = 0.600 AND $Ff_5$ is $A_{110}$ with $P_{110}$ = 0.519 THEN class is NORMAL.

**Rule 2:** IF $Tf_1$ is $A_{21}$ with $P_{21}$ = 0.015 AND $Tf_2$ is $A_{22}$ with $P_{22}$ = 0.018 AND $Tf_3$ is $A_{23}$ with $P_{23}$ = 0.250 AND $Tf_4$ is $A_{24}$ with $P_{24}$ =0.126 AND $Tf_5$ is $A_{25}$ with $P_{25}$ = 0.107 AND $Ff_1$ is $A_{26}$ with $P_{26}$ = 0.123 AND $Ff_2$ is $A_{27}$ with $P_{27}$ = 0.100 AND $Ff_3$ is $A_{28}$ with $P_{28}$ = 0.255 AND $Ff_4$ is $A_{29}$ with $P_{29}$ = 0.197 AND $Ff_5$ is $A_{210}$ with $P_{210}$ = 0.189 THEN class is SPALLING.

**Rule 3:** IF $Tf_1$ is $A_{31}$ with $P_{31}$ = 0.011 AND $Tf_2$ is $A_{32}$ with $P_{32}$ = 0.125 AND $Tf_3$ is $A_{33}$ with $P_{33}$ = 0.109 AND $Tf_4$ is $A_{34}$ with $P_{34}$ =0.057 AND $Tf_5$ is $A_{35}$ with $P_{35}$ = 0.125 AND $Ff_1$ is $A_{36}$ with $P_{36}$ = 0.050 AND $Ff_2$ is $A_{37}$ with $P_{37}$ = 0.025 AND $Ff_3$ is $A_{38}$ with $P_{38}$ = 0.150 AND $Ff_4$ is $A_{39}$ with $P_{39}$ = 0.035 AND $Ff_5$ is $A_{310}$ with $P_{310}$ = 0.073 THEN class is PITTING.

**Rule 4:** IF $Tf_1$ is $A_{41}$ with $P_{41}$ = 0.015 AND $Tf_2$ is $A_{42}$ with $P_{42}$ = 0.157 AND $Tf_3$ is $A_{43}$ with $P_{43}$ = 0.222 AND $Tf_4$ is $A_{44}$ with $P_{44}$ =0.242 AND $Tf_5$ is $A_{45}$ with $P_{45}$ = 0.036 AND $Ff_1$ is $A_{46}$ with $P_{46}$ = 0.007 AND $Ff_2$ is $A_{47}$ with $P_{47}$ = 0.055 AND $Ff_3$ is $A_{48}$ with $P_{48}$ = 0.001 AND $Ff_4$ is $A_{49}$ with $P_{49}$ = 0.073 AND $Ff_5$ is $A_{410}$ with $P_{410}$ = 0.024THEN class is CRACK.

**Table.3. The LH values of gear dataset for every class after selection of relevant features**

| Class/Features | $Tf_2$ | $Tf_4$ | $Tf_5$ | $Ff_1$ | $Ff_2$ |
|---|---|---|---|---|---|
| **Normal** | 1.400 | 1.100 | 1.150 | 1.100 | 1.200 |
| **Spalling** | 0.080 | 0.160 | 0.157 | 0.140 | 0.120 |
| **Pitting** | 0.150 | 0.090 | 0.150 | 0.090 | 0.050 |
| **Crack** | 0.270 | 0.400 | 0.050 | 0.020 | 0.070 |
| **Total LH values** | 1.900 | 1.750 | 1.507 | 1.350 | 1.440 |

After the classification process, it can be seen from Table 2 using one cluster for each class, some of the hedge values are bigger than 1, because the hedge values are not constrained in the classification step. As shown in Table 3, the discriminative powers of the selected features are better than all features. The classification results of the training and testing phases obtained from the neural-fuzzy classifier are depicted in Table 4. Here, each class for

LHNFCSF is intuitively defined with 4, 8, 12 and 16 fuzzy rules based on the cluster size for each class ranged from 1-4 clusters. The results indicated that the classification accuracy with feature selection, especially for cluster size 4 accuracy was 99.6215 % and 99.5948 % during training and testing phases, respectively with RMSE of 3.89541x $e^{-36}$ (shown in Fig. 9). For the same cluster size accuracy was 98.1742% and 97.6398 % during training and testing phases, respectively. The results indicated that, the selected features increase the recognition rate for test set. It means that some overlapping classes can be easily distinguished by selected features.

**Table.4. Classification results of different cluster sizes**

| Features | Cluster size for each class | Training accuracy | Testing accuracy | No of rules |
|---|---|---|---|---|
| ALL | 1 | 98.9716 | 97.5948 | 4 |
| 2,4,5,6,7 | 1 | 97.6523 | 96.5321 | 4 |
| ALL | 2 | 97.4514 | 97.3782 | 8 |
| 2,4,5,6,7 | 2 | 98.0198 | 97.9532 | 8 |
| ALL | 3 | 96.9716 | 96.5948 | 12 |
| 2,4,5,6,7 | 3 | 96.1272 | 95.8653 | 12 |
| ALL | 4 | 98.1742 | 97.6398 | 16 |
| 2,4,5,6,7 | 4 | 99.6215 | 99.5948 | 16 |

**Table.5. Performance of NFC, BPNN and HMM based on Classification accuracy and Computational time**

| Classification scheme | Training Accuracy (%) | Testing Accuracy (%) | Computational time (sec) |
|---|---|---|---|
| NFC(ALL) Cluster-1 | 98.9716 | 97.5948 | 36 |
| NFC(2,4,5,6,7) Cluster-1 | 100 | 96.5321 | 34 |
| NFC(ALL) Cluster-2 | 98.8924 | 97.5217 | 40 |
| NFC(2,4,5,6,7) Cluster-2 | 100 | 97.6532 | 38 |
| NFC(ALL) Cluster-3 | 98.9185 | 97.5537 | 44 |
| NFC(2,4,5,6,7) Cluster-3 | 100 | 98.8653 | 42 |
| NFC(ALL) Cluster-4 | 98.9815 | 97.5949 | 47 |
| NFC(2,4,5,6,7) Cluster-4 | 100 | 99.5948 | 45 |
| ANN(ALL) | 96.8234 | 97.4851 | 56 |
| ANN(2,4,5,6,7) | 97.9327 | 98.1783 | 54 |
| HMM (ALL) | 95.8102 | 96.1594 | 50 |
| HMM(2,4,5,6,7) | 96.6545 | 97.7892 | 48 |

MATLAB platform is used to execute ANN. The network will be trained with scaled conjugate gradient back propogation. For training 75% of the Samples were used. The remaining 15% of input features were used for testing or targeting the train value. Input data has 400 samples of 4 elements where 400 represents their corresponding data set 4 represents their classes. The target value of the first output node for the normal gear condition was set 1000 which indicate normal gear, 2nd neuron set to 0100 which indicate spalling fault gear, 3rd neuron set as 0010 which indicate pitting gear and 4th neuron set to 0001indicate cracked gear and remaining 15% is used for validation purpose. In training network is adjusted according to its error. Validation is mainly used to measure network generalization and to halt training when generalization stops improving. Testing has no effect on training and provides an independent measure of network performance during and after training. The number of hidden neurons used here is 25 to achieve minimum value. Training used is scaled conjugate gradient back propagation. Training automatically stops improving as indicated by an increase in Mean Square Error (MSE) of validation samples. MSE is the averaged square difference between output and targets. Resulting with lower MSE values are better. Zero output result indicates no error. Percent error indicates the fraction of samples which are misclassified. The maximum iteration number (epoch) of 100 were used in this process. Best validation performance of feature selection using ANN related to MSE was found to be 4.1249e-08 for training and MSE of 1.46487e-7, 2.47360e-8 for testing was obtained during 80th epoch. Without features selection process ANN Validation, Training and testing performance in terms of MSE. It was found to be 0.000099216, 1.4652e-07 and 9.6841e-07 respectively as shown in Fig. 10. The classifier accuracy of ANN with and without feature selection during training and testing phase is depicted in Table 5. WEKA version 3.6.2 is used to execute HMM classification process. Percentage option has been used to split the input feature data into 70% for training purpose and remaining 30% has been used for testing purpose. Table 5 shows the prediction accuracy percentage for HMM with and without feature selection during training and testing phase. The overall result of proposed methodology is depicted graphically in Fig.11.

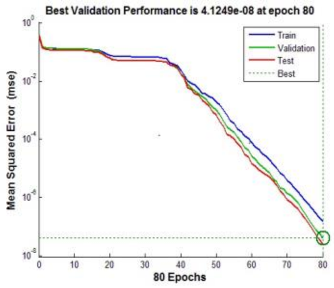



**Fig.10. Epoch Vs Mean square error using ANN**

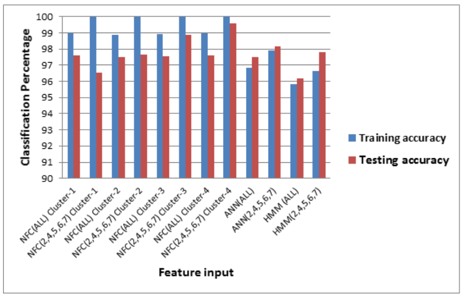



**Fig.11. Comparison of Classification accuracy of different classifiers**

## 2. CONCLUSION

A NFC, ANN, HMM based strategy was introduced for predicting gear faults by utilizing measurable statistical feature vectors from EEMD coefficients of vibration signals of different states of a gear conditions.

- The selection of input features and the suitable classifier input parameters have been upgraded utilizing Linguistic Hedges Based on Neural Fuzzy Classification Process.
- The selected features are alone then classified using NFC, ANN and HMM.
- In this research work, WEKA based Classifier performance comparison was made with and without feature selection using the extracted features.
- It was observed that EEMD feature extraction followed by Linguistic hedges feature selection classified using Neural fuzzy classification with a cluster size of 4 given the best fault diagnosis with a training and testing accuracy with a decent Computational time .

## REFERENCES


Cetisli B, The effect of linguistic hedges on feature selection: Part 2, Expert Systems with Applications, 37 (8), 2010, 6102-6108.

Elforjani M, Mba D, Muhammad A and Sire A, Condition monitoring of worm gears. Applied Acoustics, 73 (8), 2012, 859-863.

Hecht-Nielsen R, Theory of the back propagation neural network, In Neural Networks, IJCNN, International Joint Conference, 1989, 593-605.

Hornik K, Approximation capabilities of multilayer feed forward networks, Neural Networks, 4 (2), 1991, 251-257.

Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q and Liu HH, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, In Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 454 (1971), 1998, 903-995.

Laurene Fausett, Fundamentals of neural networks: architectures, algorithms, and applications, Prentice-Hall, Inc, 1994.

Lei Y, Z He, Y Zi, and Q Hu, Fault diagnosis of rotating machinery based on multiple ANFIS combination with Gas, Mechanical Systems and Signal Processing, 21 (5), 2007, 2280-2294.

Mc Fadden PD, Detecting fatigue cracks in gears by amplitude and phase demodulation of the meshing vibration, ASME J. Vib. Acoust., 108, 1986, 165–170.

Meng Q & L Qu, Rotating machinery fault diagnosis using Wigner distribution, Mechanical Systems and Signal Processing, 5 (3), 1991, 155-166.

Paul D Samuel, Darryll J Pines, Classifying Helicopter Gearbox Faults Using a Normalized Energy Metric, Institute of Physics Publishing, Smart Materials and Structures, 10, 2001, 145–153.

Purushotham V, Narayanan S and Prasad SA, Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition, NDT and E International, 38 (8), 2005, 654-664.

Rajeswari C, Sathiyabhama B, Devendiran S and Manivannan K, Diagnostics of gear faults using ensemble empirical mode decomposition, hybrid binary bat algorithm and machine learning algorithms, Journal of Vibro Engineering, 17 (3), 2015, 1169-1187.

Samanta B, Gear fault detection using artificial neural networks and support vector machines with genetic algorithms, Mechanical Systems and Signal Processing, 18 (3), 2004, 625-644.

Saravanan N, Cholairajan S and Ramachandran KI, Vibration-based fault diagnosis of spur bevel gear box using fuzzy technique, Expert systems with applications, 36 (2), 2009, 3119-3135.

Saravanan N, Kumar VNS, Siddabattuni KI, 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.

Saravanan N, Siddabattuni VK and Ramachandran KI, A comparative study on classification of features by SVM and PSVM extracted using Morlet wavelet for fault diagnosis of spur bevel gear box, Expert systems with applications, 35 (3), 2008, 1351-1366.

Staszewski WJ and Tomlinson GR, Application of the Wavelet Transform to Fault Detection in a Spur Gear, Mechanical Systems and Signal Processing, 8, 1994, 289-307.

Subrahmanyam M and Sujatha C, Using neural networks for the diagnosis of localized defects in ball bearings, Tribology International, 30 (10), 1997, 739-752.

Sugumaran V, Muralidharan and Ramachandran KI, Feature selection using Decision Tree and classification through Proximal Support Vector Machine for fault diagnostics of roller bearing, Mechanical Systems and Signal Processing, 21, 2006, 930–942.

Sun W, Chen J and Li J, Decision tree and PCA-based fault diagnosis of rotating machinery, Mechanical Systems and Signal Processing, 21 (3), 2007, 1300-1317.

Tian X, Lin J, Fyfe KR and Zuo MJ, Gearbox fault diagnosis using independent component analysis in the frequency domain and wavelet filtering. In Acoustics, Speech, and Signal Processing, Proceedings (ICASSP'03), IEEE International Conference, 2, 2003, 245.

Wang W, Ismail F and Golnaraghi F, A Neuro-Fuzzy approach to gear system monitoring, IEEE Transactions on Fuzzy Systems, 12 (5), 2004, 710-723.

Wu Z and Huang NE, Ensemble empirical mode decomposition: a noise-assisted data analysis method, Advances in adaptive data analysis, 1 (1), 2009, 1-41.

Yaguo Lei, Ming J. Zuo, Gear Crack Level Identification Based on Weighted K Nearest Neighbor Classification Algorithm, Mechanical Systems and Signal Processing, 23, 2009, 1535-1547.

Yaguo Lei, Ming J. Zuo, Zhengjia He, Yanyang Zi, A Multidimensional Hybrid Intelligent Method for Gear Fault Diagnosis, Expert Systems with Applications, 37, 2010, 1419-1430.

Yang D, Liu Y, Li S, Li X and Ma L, Gear fault diagnosis based on support vector machine optimized by artificial bee colony algorithm, Mechanism and Machine Theory, 90, 2015, 219-229.

Yu Yang, Dejie Y and Junsheng C, A roller bearing fault diagnosis method based on EMD energy entropy and ANN, J. Sound and Vibration, 294, 2006, 269- 277.