An Intelligent Gear Fault Diagnosis Model Based on EMD and Evolutionary Algorithms.

S.Devendiran*1, K.Manivannan2
1,2 School of Mechanical and Building Sciences, VIT University, Vellore, Tamil Nadu, India
devendiran@vit.ac.in
manivannan.k@vit.ac.in

Abstract -- Gears are a vital element considering its applications in a variety of machine tool applications. An unpredicted failure of the gear may cause substantial economic losses. For this very reason, fault diagnosis in gears has been the subject of intensive research. An intelligent method to diagnose and predict the gear fault using vibration signal is proposed in this research work. A signal which contains useful information from various conditions of gear is extracted from an experimental rig. From the original acceleration vibration signals, statistical features are extracted after using the EMD (Empirical mode decomposition) which decomposes the signal into a finite number of stationary intrinsic mode functions (IMFs). Extracted features are recognized and classified by a novel heuristic classifier artificial bee colony (ABC) algorithm. In order to select the predominant features, traditional Genetic algorithm (GA) as well as ReliefF Genetic Algorithm (RFGA) is utilized. The fault diagnosis results are compared with support vector machine (SVM) classifier and their relative efficiency were compared based on the classification accuracy.

Index Term-- Experimental rig, fault diagnosis, Empirical mode decomposition, Artificial bee colony algorithm, Genetic algorithm, ReliefF Genetic Algorithm & Support vector machine.

I. INTRODUCTION

The introduction of new methodologies in condition monitoring and fault diagnosis of gears [1] enhances continuous production rate and prevents shutdown in manufacturing processes. Vibration signals of normal gear differ from the faulty gears, whose vibrations are processed [2] to forecast unexpected breakdowns. This is a suitable for accurate gear fault diagnosis. The conventional methods for processing vibration signals widely employed to detect gear failures are classified as frequency domain technique, time-domain technique and time-frequency domain technique. In particular time-domain analysis focuses on statistical characteristics of vibration signal and frequency domain approach uses transformation of the time-domain signal to the frequency domain usually in the form of a fast Fourier transform (FFT) algorithm [3].

Usually, the vibration signals extracted from gear box in an rotating machinery consist of non-stationary components contain rich information about faults of components, therefore, it is important to analyse the non-stationary signals. But FFT analysis is not suitable for those signals [4]. Wavelet transform has significantly expanded new aspects in signal and image processing over the last decade. It has significantly expanded the signal analyst's capabilities and offers certain benefits over the classical fourier transform (FFT). The wavelet transform was utilized to represent all possible types of transients in vibration signals generated by faults in a gearbox [5]. Machine learning algorithms like Artificial neural networks [6], C4.5[7] are reported in fault diagnosis. Effective condition monitoring requires identification of best features in terms of reduction in computation as well the accuracy in diagnosis, which is essentially data reduction strategy which has to be implemented. Principle component analysis PCA [8] an unsupervised dimension reduction method is used to make feature reduction. Based on available literatures, it could be noted that the MSVM and ABC as a classifier has still an immense interest and very little efforts have been performed for the fault diagnosis for gears. Looking into tremendous information contained in EMD an attempt in conjunction with the MSVM and ABC and the optimized feature selection process using GA & RFGA for the multi-fault classification would be a worth effort, which is lacking in available literatures.

In Section 1 we briefly reviewed the works which are strictly connected to the subject of this paper. In Section 2 Experimental setup and experimental procedure is presented. In Section 3 Theoretical background of empirical mode decomposition (EMD) and application of extracted signals were presented. In Section 4 Theoretical background of evolutionary optimization technique GA and RFGA were presented. In Section 5 Theory behind the classification algorithms ABC and MSVM is given. In Section 6 overall implementation process execution such as feature extraction, feature selection and classification, its result analysis and discussion were presented and Last section contains conclusions.
II. EXPERIMENTAL SETUP & EXPERIMENTAL PROCEDURE

Experimental set-up is shown in Fig-1. Setup consists of three phase 0.5 hp AC motor, variable frequency drive (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.

The Fig 2 depicts the methodology of proposed work. Vibration signals are acquired for all the four conditions of gear conditions mentioned earlier in the paper. Each signal is separated into data sets signals, and they are used to extract the statistical features. Empirical Mode Decomposition (EMD) is employed to decompose the data into number of intrinsic mean functions (IMF’s) and there after the statistical features is extracted using the IMF’s. Both time domain and frequency domain features are extracted in similar method. The extracted features has to be reduced to improve accuracy and so genetic algorithm based feature selection technique is deployed, which minimizes the features based on their fitness. These reduced features are the next level input for classification of data. ABC and MSVM are the classifiers used to detect the misclassifications and thereby achieving better accuracy in predicting the faults. The results of ABC algorithm and MSVM are compared to decide the best approach.

III. EMD (Empirical mode decomposition)

Empirical mode decomposition (EMD) is said to be the powerful tool to improve time-frequency evaluation. EMD results a key orthogonal component termed as intrinsic mode function, the prior characteristics based on time scales are generated and decomposed into IMF’s. The natural oscillatory mode within the signals and the basic characters are identified by the IMF’s, rather than complicated pre-initiated kernels. Thus, EMD is a beneficial technique that has ability to perform on both nonlinear and dynamic processing. IMF has to satisfy two criteria: first the number of extrema and the number of zero-crossings must either equal or differ at most by one in the input data set. Secondly the average value of the envelope defined by local maxima and the envelope defined by the local minima at any point is always zero. It is a simple oscillatory mode attached within the signal. EMD method is introduced to decompose a signal into IMF components and the last component is referred as residue. The pseudo code of the EMD algorithm [9] depicted in Fig.6.

- Experimentation
- Data processing
- Feature extraction using EMD
- Feature selection using GA
- Fault classification using ABC and MSVM

Fig. 2. Methodology
The acquired raw signal data of all the four different gear vibrations, namely Normal, frosting, pitting and crack are given as input to the EMD. There are 4 classes to be examined. Randomly selected 6000 time domain data for each gear are given to the EMD. It calculates the local maxima and local minima. Interpolation of maxima and minima is found and termed as envelope and the mean distance between the envelope were calculated. The IMF is the difference between the envelope and the mean distance between the envelopes. IMF’s are generated completely and decomposition of the signals is achieved for all four classes. Each data set has numerous IMF’s imbedded into them. It generates a series of numbers which is in the form of sinusoidal waves. Each class of gear has different IMF with different waveforms. Comparing the waveforms the faults can be predicted. For effective prediction the time domain HHT plots have to be converted into FFT plots from which the peaks and troughs can be determined. The decomposition splits the data set into 8 IMF and 1 residue.

So in this case 32 IMF and 4 residues are obtained totally, which is further processed into genetic algorithm for reduction of features.

These IMF’s are helpful parameters to achieve the time domain and using FFT the frequency domain features are extracted. Hilbert Huang Transform method is used to plot the IMF’s and similarly FFT is plotted for the corresponding IMF’s. Fig.3 shows the HHT and FFT plots of normal gear. Fig.4, Fig.5, Fig.7 are the plots for frosting, pitting and the cracked gears respectively. The time domain and frequency domain features are concatenated to initiate the genetic algorithm. Only 3 IMF’s from each class is taken as they contain maximum information and it have information of the characteristics frequencies were bound on the range and their corresponding statistical features are extracted and then concatenated as a matrix input with 120 features into GA.GA then selects the features for further classification of data using classifiers.
Algorithm EMD

STEP 1 Input: \( x_0 = y(t) \), and \( i = 1 \)

STEP 2 Obtain the \( i^{th} \) IMF \( \alpha_i \)

STEP 3 \( h_{i(k-1)} = r_{i-1} \), \( k = 1 \)

STEP 4 Extract the local maxima and minima of \( h_{i(k-1)} \)

STEP 5 Interpolate the maxima and minima to form envelopes of \( h_{i(k-1)} \)

STEP 6 Calculate the mean \( m_{i(k-1)} \) of lower and upper envelopes

STEP 7 Assume \( h_i = h_{i(k-1)} - m_{i(k-1)} \)

STEP 8 If \( h_i \) is an IMF then set \( \alpha_i = h_i \), else go to step (4) with \( k = k + 1 \)

STEP 9 Define the residue \( r_{i+1} = r_{i} - \alpha_i \)

STEP 10 If \( r_{i+1} \) still has 2 or more extrema then go to step (2), else terminate.

IV. RFGA (ReliefF Genetic Algorithm)

Genetic algorithm [10] is inspired by biological functioning of living beings and their characteristic of evolution from previous generations is adopted into feature selection area in our concern. Kira and Rendell [11,12] developed an algorithm called Relief, which inspired by most previous hybrid GA, the hybridizations of RFGA include initial population seeding, crossover, and mutation modification. They are able to detect conditional dependencies between features and provide a unified view on the attribute estimation in classification. However, it cannot deal with incomplete data and is limited to two class problems. Kononenko’s ReliefF algorithm Fig. 8 [13] which is an extension of Relief is more robust and can deal with incomplete and noisy data as well as multiclass problems. The candidate features are first ranked by using ReliefF, and then the feature weight information is incorporated as the heuristic to feed the GA’s initial population, crossover, and mutation operators.

The major steps of RFGA are Feature Ranking, Chromosome Decoding, Initial Population and Fitness Evaluation. ReliefF (RF) is engaged to produce the heuristic for RFGA by ranking the features with higher weight in order to decrease the chance of selecting features with lower weight. The chromosome is decoded using a binary string with the length equal to the number of a candidate feature subset. Each gene may have value “0” or “1” which indicates whether a feature is selected or not. The feature weight information derived from the Feature Ranking is used to feed the generation of initial population.
**Algorithm ReliefF**

**Input:** for each training instance a vector of feature values and the class value

**Output:** the vector W of estimations of the qualities of features

set initial weights \( W[F] = 0 \);

for \( i = 1 \) to \( m \) do

begin

randomly select an instance \( R_i \);

find \( k \) nearest hits \( H_j \);

for each class \( C = \text{class}(R_i) \) do

from class \( C \) find \( k \) nearest misses \( M_j(C) \);

for \( A = 1 \) to \( a \) do

end;

\[
W[F] = W[F] - \sum_{j=1}^{k} \frac{\text{diff}(F, R_i, H_j)}{(m.k)} + \\
\sum_{C=\text{class}(R_i)} \left[ \frac{P(C)}{1-P(\text{class}(R_i))} \sum_{j=1}^{k} \frac{\text{diff}(F, R_i, M_j(C))}{(m.k)} \right]
\]

end;

Fig. 8. Pseudo code of ReliefF algorithm

The chromosomes in RFGA are randomly generated by using only the top 10 genes according to the ranked feature list produced by RF to form the initial population. The fitness function for evaluation of the chromosome is given in equation class distance within the classes to be calculated and given by,

\[
J_c = \sum_{i=1}^{d} P_i J_i
\]

\[
J_i = \frac{1}{n_i} \sum_{k=1}^{n}(x^j_k - m_j)^T(x^j_k - m_j)
\]

\( i = 1, 2 \ldots d \), \( m_i \) is the mean vector of class \( i \), \( n_i \) is the number of samples in class \( i \), \( P_i \) is the number of samples in class \( i \).

The class distance between the classes is given by:

\[
J_b = \sum_{i=1}^{c} P_i (m_i - m)^T (m_i - m)
\]

\( m \) – Mean vector of all classes

Fitness function is defined as

\[
J = J_i + (1/J_b)
\]

Equations 1 to 4 define the work of genetic algorithm using distance evaluation technique. The features are selected based on the distance evaluated between the features as in Fig 7. The same proficiency is adopted in this case to reduce the features for classification of data.
HEORETICAL BACKGROUND OF ABC AND MSVM CLASSIFIERS

A. ABC classifier:

The artificial bee colony algorithm (ABC) is an optimization algorithm based on the intelligent foraging behaviour of honey bee swarm, proposed by Karaboga [14]. In the ABC model, the colony consists of three groups of bees: employed bees, onlookers and scouts. It is assumed that there is only one artificial employed bee for each food source. The position of a food source represents a possible solution to the optimization problem and the nectar amount of a food source corresponds to the quality (fitness) of the associated solution. The number of the employed bees is equal to the number of solutions in the population. Employee bees visit the food source already visited by it previously. At the first step, a randomly distributed initial population (food source positions) is generated. After initialization, the population is subjected to repeat the cycles of the search processes of the employed, onlooker, and scout bees, respectively. Positioning of food sources by the bees helps us to represent a desired solution to optimization problem. The sugar fluids produced by flowers are made into honey by bees called nectar.

The nectar corresponds to quality (fitness) of the associated solution and this can be calculated by the equation mentioned below.

\[ f_{i} = \left( \frac{1}{1 + f_i} \right) \]  

(5)

An artificial onlooker bee chooses a food source depending on the probability value associated with that food source, \( p_i \), calculated by the following expression

\[ P_i = \frac{\text{fit}_i}{\sum_{n=1}^{SN} \text{fit}_n} \]  

(6)

where SN is the number of food sources equal to the number of employed bees, and \( \text{fit}_i \) is the fitness of the solution given \( i \) is inversely proportional to the \( \text{fit}_i \) given in Eq.(5) where \( f_i \) is the cost function of the clustering problem.

In order to produce a candidate food position from the old one in memory, the ABC uses the following expression (7)

\[ v_{ij} = z_{ij} + \phi_j (z_j - z_{ij}) \]  

(7)

where \( k \in \{1,2,\ldots,SN\} \) and \( j \in \{1,2,\ldots,D\} \) are randomly chosen indexes. The term \( k \) is a random number between \([-1,1]\). As can be seen from Eqn. (7), as the difference between the parameters of the \( z_{ij} \) and \( z_{ij} \) decreases, the perturbation on the position \( z_{ij} \) decreases, too. It is also clear that, as the search approaches to the optimum solution in the search space, the step length is adaptively reduced.

The value of predetermined number of cycles is an important ABC algorithm’s control parameter, which is called “limit” for abandonment. Assuming that the abandoned source is \( z_{i2} \) and \( j \in \{1,2,\ldots,D\} \), then the scout discovers a new food source to be replaced with \( z_i \). This operation can be defined as in (8).

\[ Z_i^j = z_{i2}^j + \text{rand}(\delta,1)(z_{i2}^j - z_{i1}^j) \]  

(8)

A greedy selection mechanism is employed as the selection operation between the old and the candidate one. There are three control parameters in the ABC: the number of food sources which is equal to the number of employed or onlooker bees (SN), the value of limit, the maximum cycle number (MCN).
Pseudo code of ABC algorithm is presented as in Fig.10. Flow chart of ABC algorithm is illustrated in Fig 8. The reduced features using GA are given as input to this classifier which classifies the varying data. Thus accuracy can be improved by using data which are under the same cluster. These clusters are grouped and others are depicted to be the misclassified data. The results of ABC are then compared with MSVM classifier to determine the best approach.

**Pseudocode : ABC algorithm**

1. Loading of training samples  
2. Generation of initial population $z_i$, $i=1$,...,$SN$  
3. Evaluate the fitness $f_i$ of the population  
4. Set cycle to 1  
5. Repeat  
6. FOR each employed bee {  
   a. Produce new solution $v_i$ by using eqn.(7)  
   b. Calculate the value  
   c. Apply greedy selection process}  
7. Evaluate the probability values $p_i$ for the solutions $z_i$ by eqn. (6)  
8. FOR each onlooker bee {  
   d. Select a solution $z_i$ depend on $p_i$  
   e. Produce new solution $v_i$  
   f. Calculate the value $f_i$  
   g. Apply greedy selection process  
9. If there is a rejected solution for the scout then replace it with a new solution which will be randomly produced by eqn.(8)  
10. Memorize the best solution so far  
11. Cycle=cycle + 1  
12. Until cycle = maximum cycle number

Fig. 11. Flowchart of ABC algorithm

**B. One Against – All (OAA) Multiclass Algorithm**

The method of OAA consists of KSVM models where $K$ stands for class number. The minimization function is given as,

Minimize:

$$\frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{n} \xi_i (\omega^i)^T$$

Subject to

$$(\omega^i)^T \psi(x_k) + b^i \geq 1 - \xi_i^k \ y = i$$

$$(\omega^i)^T \psi(x_k) + b^i \leq -1 + \xi_i^k \ y \neq i$$

Fig. 10 Pseudo code for ABC algorithm
\[ z_j^k \geq 0, \ j = 1, \ldots, l \]

where the training data \( x_i \) is mapped to a higher-dimensional space by function \( \phi \) and \( C \), where \( C \) is the penalty parameter., (CEP) is the percentage ratio of misclassified patterns to the test data sets. The Euclidean distances classified each pattern by assigning it to the class whose centre is closest, using the Euclidean distances, to the centre of the clusters. This assigned output (class) is compared with the desired output and if they are not exactly the same, the pattern is separated as incorrectly classified. It is calculated for all test data and the total incorrectly classified pattern number is percentage which is given by percentage ratio of misclassified sample to the size of test data set, which is given by Eq.12

\[
\text{CEP} = \frac{100 \times \text{No.of misclassified samples}}{\text{sizeofdataset}}
\]

As described above, the data is given in two pieces: the training set (the first 75%) and the test set (the last 25%). For elaborate theoretical information of SVM refer [15].

VI. EXPERIMENTAL ANALYSIS, RESULTS AND DISCUSSIONS

The gear signals are extracted in the sampling rate of 12800 Hz (6400 data points per second). For 5 seconds 32000 data points were collected through accelerometer. After obtaining the gear signals for four conditions using EMD 8+1 IMFs are derived and for each IMF, equivalent number of FFT also derived (fig 3). Using first 3 IMFs of time and frequency domain, statistical features such as Root mean square (RMS), kurtosis (KUR), Skewness (SKW), variance (VAR), Standard deviation (STD) in Table I are extracted from both time and frequency domain data.

<table>
<thead>
<tr>
<th>IMF 1</th>
<th>IMF 2</th>
<th>IMF 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std</td>
<td>Mean</td>
<td>Pk-Pk</td>
</tr>
<tr>
<td>F1-F4</td>
<td>F5-F8</td>
<td>F9-F12</td>
</tr>
<tr>
<td>F61-64</td>
<td>F65-668</td>
<td>F69-F72</td>
</tr>
</tbody>
</table>

For feature selection using GA the initial parameters taken as follows: the population is 120, the length of chromosome code is 30, number of generation is 400 and 1 crossing point. The result shows as in Fig.12, the best combination of selected features satisfying the given objective. The values greater than 0.5 is considered as a selected feature and the fitness of the process is also arrived. From the feature results Fig.12 11011000110010011001001011110 features selected. For ex: (F1,F2,F4,F5,F8,F9,F10,F13,F17,F21,F25,F27,F28,F29) are selected and the remaining are abandoned. The initial features in RFGA are similar to the features that are input to traditional GA. The process of feature reduction is based on hybridized RFGA algorithm and the features are selected. The Selected feature subsets both from GA and RFGA are used to train and test the ABC algorithm and MSVM. The total feature set calculated is split into training (70%) and testing (30%) data set. Samples for a test set are used to evaluate the ABC and MSVM classifications. The classification process is grouped into 8 schemes where-in each scheme has a combination of different classifiers and feature selection procedures. The classification as shown in Table 2 consists of the description of the Classifier, their respective schemes and the corresponding results. Among all schemes, overall average testing accuracy of 96% is higher in case of ABC-S4 schema. Thus we can infer that a combination of RFGA and ABC gives a better result.

It is also seen that the results from ABC algorithm without feature selection has given 93% accuracy whereas with feature selection, ABC produced 93.75% and 94.5% accuracy. ABC with RFGA gives a classification accuracy of 96% which is the maximum among the classifiers. Meanwhile, the MSVM classifier provides 92.5% accuracy when built without feature selection. The accuracy decreases to 91% and 91.5% with feature selections using GA even with reduced features and MSVM with RFGA gives 95% only. Thus we can infer that a combination of RFGA and ABC gives a better result.
Table II
Comparison of diagnosis results

<table>
<thead>
<tr>
<th>Classifier description</th>
<th>Scheme</th>
<th>No. of features</th>
<th>Features</th>
<th>Prediction accuracy of testing data (%)</th>
<th>Average Accuracy on testing data (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>normal</td>
<td>spalling</td>
</tr>
<tr>
<td>ABC (no feature reduction)</td>
<td>ABC-S1</td>
<td>All (30)</td>
<td>F1-F30</td>
<td>92</td>
<td>93</td>
</tr>
<tr>
<td>ABC (feature reduction with GA)</td>
<td>ABC-S2</td>
<td>14</td>
<td>F1,F2,F4,F5,F8,F9,F10,F13,F17,F21,F25,F27,F28,F29</td>
<td>92</td>
<td>95</td>
</tr>
<tr>
<td>ABC (feature reduction with GA)</td>
<td>ABC-S3</td>
<td>14</td>
<td>F2,F4,F5,F8,F10,F11,F14,F15,F18,F20,F21,F23,F24,F25</td>
<td>90</td>
<td>96</td>
</tr>
<tr>
<td>ABC (feature reduction with RFAG)</td>
<td>ABC-S4</td>
<td>10</td>
<td>F1,F4,F6,F8,F10,F11,F18,F19,F20,F24</td>
<td>95</td>
<td>95</td>
</tr>
<tr>
<td>MSVM (no feature reduction)</td>
<td>MSVM-S1</td>
<td>All (30)</td>
<td>F1-F30</td>
<td>99</td>
<td>92</td>
</tr>
<tr>
<td>MSVM (feature reduction with GA)</td>
<td>MSVM-S2</td>
<td>14</td>
<td>F1,F2,F4,F5,F8,F9,F10,F13,F17,F21,F25,F27,F28,F29</td>
<td>92</td>
<td>90</td>
</tr>
<tr>
<td>MSVM (feature reduction with GA)</td>
<td>MSVM-S3</td>
<td>14</td>
<td>F2,F4,F5,F8,F10,F11,F14,F15,F18,F20,F21,F23,F24,F25</td>
<td>90</td>
<td>95</td>
</tr>
<tr>
<td>MSVM (feature reduction with RFAG)</td>
<td>MSVM-S4</td>
<td>10</td>
<td>F1,F2,F5,F6,F8,F11,F17,F19,F20,F25</td>
<td>94</td>
<td>95</td>
</tr>
</tbody>
</table>

Fig. 12(a) GA feature selection and fitness plot for RMS
Fig. 12(b) GA feature selection and fitness plot for KURTOSIS

Fig. 13. Comparison of ABC vs MSVM
VII. CONCLUSIONS

The paper presents a new approach to diagnosing faults in gears using EMD, GA, RF-GA and ABC algorithm based artificial intelligence techniques. An ABC and One-Against-All-MSVM based procedure is presented for fault detection and identification of gear fault using statistical feature vectors such as standard deviation, mean, peak to peak, root means square and kurtosis arrived from EMD coefficients of vibration signals of various faultless and faulty conditions of a gearbox. The selection of input features and the appropriate classifier parameters have been optimized using traditional Genetic algorithm and ReliefF based Genetic Algorithm. ABC with RF-GA based feature reduction scheme gives the better classification output as well perfect accuracy and performance to identify gear faults. It will be interesting to see that the pre-conditioning of the vibration signal by EMD before the extraction of features for possibilities of better gear fault predictions.

REFERENCES