

## Fault Diagnosis of Helical Gear Box using Variational Mode Decomposition and Random Forest Algorithm

Akhil Muralidharan<sup>1,2</sup>, V. Sugumaran<sup>1</sup>, K.P Soman<sup>3</sup> and M. Amarnath<sup>4</sup>

**Abstract:** Gears are machine elements that transmit motion by means of successively engaging teeth. In purely scientific terms, gears are used to transmit motion. A faulty gear is a matter of serious concern as it affects the functionality of a machine to a great extent. Thus it is essential to diagnose the faults at an initial stage so as to reduce the losses that might be incurred. This necessitates the need for continuous monitoring of the gears. The vibrations produced by gears from good and simulated faulty conditions can be effectively used to detect the faults in these gears. The introduction of Variational Mode Decomposition (VMD) as a new signal pre-processing technique along with the different decision trees have provided good classification performance. VMD allows decomposition of the signal into various modes by identifying a compact frequency support around its central frequency, such that adding all the modes reconstructs the original signal. Alternating direction multiplier method (ADMM) is used by VMD to find the intrinsic mode functions on central frequencies. Meaningful statistical features can be extracted from VMD processed signals. J48 decision tree algorithm was used to identify the useful features and the selected features were used for classification using the decision trees namely, Random Forest, REP Tree and Logistic Model Tree algorithms. The performance analyses of various algorithms are discussed in detail.

**Keywords:** Gear fault diagnosis, Variational Mode Decomposition, Alternating Direction Multiplier Method, J48 tree algorithm, Random Forest algorithm, Logistic Model Tree, REP Tree.

---

<sup>1</sup> SMBS, VIT University, Chennai Campus, Vandalur-Kelambakkam Road, Chennai, India.

<sup>2</sup> Corresponding author.

<sup>3</sup> Centre for Excellence in Computational Engineering and Networking, Amrita Vishwa Vidyapeetham, Coimbatore, Tamil Nadu, India.

<sup>4</sup> Department of Mechanical Engineering, Indian Institute of Information Technology Design and Manufacturing, Jabalpur, Madhya Pradesh, India.

## 1 Introduction

Gears are one of the most important components of the modern industrial machines. All the machinery available in the industry, ranging from small scale machines to the heavy-duty industrial machines rely on common gears. Thus, it is required to make timely maintenance to ensure smooth functioning of the machines. Failure to detect the fault in the gear will lead to huge economic losses and physical damages as well. Hence it is essential to carry out an experimental study which provides a method for its proper monitoring and fault diagnosis. A transmission refers to the gearbox that uses gears and gear trains to provide speed and torque conversions from a rotating power source to another device [Uicker, Pennock, Shigley (2003); Paul (1979)]. A gear is a rotating machine part having cut teeth, which mesh with another toothed part in order to transmit torque, in most cases with teeth on the one gear being of identical shape, and often also with that shape on the other gear. Defects in gears can be classified into three categories namely; tooth breakage, cracked tooth and surface wear [Staszewski and Tomilson (1994)]. Localized faults are the most common defects which are observed in helical gear boxes. This occurs when a sizable piece of material on the contact surface is dislodged during operation, mostly by fatigue cracking under cyclic contact stressing. Vibration monitoring is a traditional method used for monitoring of gearboxes.

The study uses physical parameters such as sound, acoustic emission, vibration and wear debris for the detection and diagnosis of the inchoate faults as it is very difficult to measure the severity of the localized faults directly when the gears are running. A general review of monitoring and fault diagnosis techniques can be found in S. Nandi and H. A. Toliyat (1999) and M. E. H. Benbouzid and G. B. Kliman (2003). The recent studies suggest the application of the acoustic emission technology in research and industry [Mba and Rao (2006)]. A Lamb wave-based damage extension diagnosis method to monitor the damage on typical structures of composite aircraft is found in Dongyue Gao, Yishou Wang, Zhanjun Wu and Rahim Gorgin (2013). In relation to gearboxes a few investigators have assessed the application of Acoustic Emission technology for diagnostic and prognostic purposes [Wheitner, Houser, and Blazakis, (1993); Singh, Houser, and Vijayakar (1999); Miyachika, Oda and Koide (1995); Miyachika, Zheng, and Tsubokura (2002)]. However, it was observed that AE was more sensitive to the scale of surface damage than vibration analysis. Sound and vibrations generated by rotating machinery often mask the features of fault-related signals generated by the machine elements such as gears, bearings and cams [Wuxing, Tse, Guiicai, Shitielin (2004); Zvokelj, Zupan, and Prebil (2010); Li and Ma (1997); Tinta, Petrov, Benko, Juric, Rakar, Zele, Tavar, Rejec, and Stefanovska (2005)]. While taking fast Fourier transform of vibration signals, the harmonic and noise overlaps with frequency components.

This makes it difficult to read the actual frequency components present in the signal. The non-stationary nature of the signals makes the situation further worse by changing the frequency component itself. Hence analysis of the above signals in faulty operating conditions becomes difficult. Machine learning can be an effective tool for fault diagnosis. Amarnath M, V. Sugumaran, Deepak Jain and Hemantha kumar (2013) and V. Sugumaran, Deepak Jain, M. Amarnath and Hemantha Kumar (2013) have used decision tree algorithm in fault diagnosis of helical gear boxes which has provided significant classification accuracies.

Considering these circumstances, researchers were forced to pay their attention on signal processing methods for improving fault classification tools. Recent studies illustrate the use of Empirical Mode Decomposition (EMD) to detect incipient faults in gears. R. Ricci and P. Pennachhi (2011) used Empirical Mode Decomposition (EMD) along with intrinsic mode functions (IMF) to detect incipient faults in gears. The IMF doesn't work well with non-stationary signals. Y. G. Lei, M. J. Zuo and Z. J. He (2010) used EMD to extract features from signals for classifying the different modes and degrees of gear faults. However, EMD lacks mathematical theory foundation; the technique is faced with the difficulty of being essentially defined by an algorithm, and therefore of not admitting an analytical formulation which would allow for a theoretical analysis and performance evaluation [Gabriel Rilling, Patrick Flandrin and Paulo Goncalves (2003)]. The wavelet can represent signals in time frequency plane; however, it has some limitations [Li and Ma (1997); Huang (1998)].

The present study makes use of a new preprocessing technique to decompose the signal into various modes or IMFs using calculus variations. The modes have compact frequency support around the central frequency. ADMM was used as optimization tool to find such central frequencies concurrently. The main purpose of decomposing a signal is to identify various components of the signal. This work focuses on a new algorithm - variational mode decomposition (VMD), which extracts different modes present in the signal. In the present study, an attempt is made to exploit vibration signals for the purpose of fault diagnosis of helical gear. To extract some meaningful features, the vibration signals were preliminarily pre-processed for finding the modes and IMFs. Then, descriptive statistical features like mean, median, kurtosis etc. were extracted. With the extracted statistical features, classification was carried out using various decision tree algorithms namely; Random Forest, REP Tree and Logistic Model Tree.

## **2 Experimental setup and Procedure**

The experimental setup is shown in Fig. 1. A two stage helical gearbox with a 5 HP rating is used in the experiment. The gear box is driven by a 3-phase induction

motor which has a 5.5 HP rating and a rated speed of 1440 rpm. An inverter drive controls the speed of the motor. For the present study; the motor is operated at 80 rpm, i.e., the speed of the first stage of the gearbox is 80 rpm. A step-up ratio of 1:15 is established which generates a speed of 1200 rpm at the pinion shaft in the second stage of the gear box. Table.1 summarizes the specifications of the test rig.

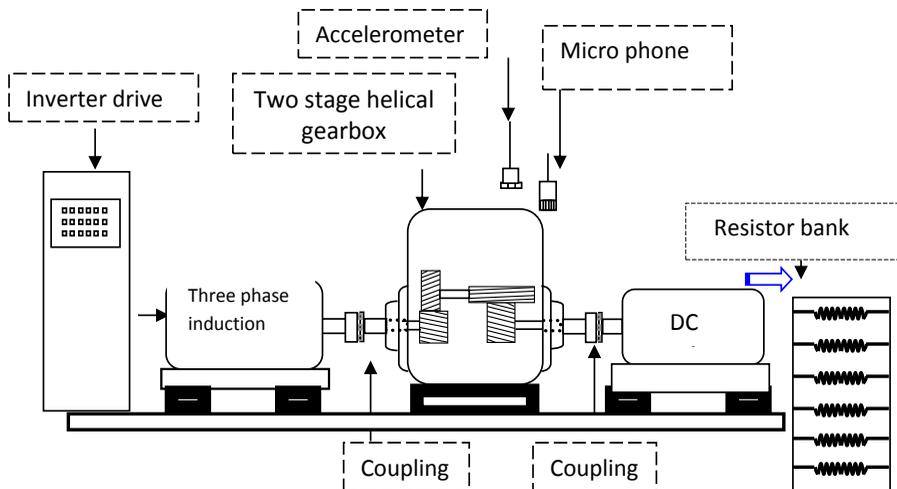


Figure 1: Experimental setup of two stage helical gearbox.

Table 1: Specifications of helical gear box.

Specifications	First Stage	Second Stage
No. of teeth	44/13	73/16
Pitch circle diameter (mm)	198/65	202/48
Pressure angle (°)	20	20
Helix angle (°)	20	15
Modules	4.5/5	2.75/3
Speed of shafts	80 rpm	1200 rpm
Mesh frequency	59 Hz	320 Hz
Step-up ratio	01:15	
Rated power	5 Hp	
Power transmitted	2.6 Hp	

The pinion is connected to a D.C motor (which is used as generator) to generate 2 kW power. The power generated is dissipated in a resistor bank. Hence, the

actual load on the gearbox is only 2.6 HP which is 52% of its rated power 5 HP. In industrial environment utilization of load varies from 50% to 100%. In the case of traditional dynamometer, torque fluctuations cause additional torsional vibrations. This can be avoided by using D.C motor and resistor bank.

Backlash can be restricted to the gears by fitting tyre couplings between the electrical machines and gear box. The generator, gear box and motor are mounted on I-beams, which are anchored to a massive foundation. Bruel & Kjær accelerometer are used to measure Vibration signals which are installed close to the test bearing. Signals are sampled at a sampling frequency of 8.2 kHz. The optimum location of the sensors is a critical issue of any successful Structural Health Monitoring System. Sensor optimization problems encompass mainly three areas of interest: system identification, damage identification and impact identification [Mallardo and Aliabadi (2013)]. The experimental setup with equipment and sensors is shown in Fig. 2.

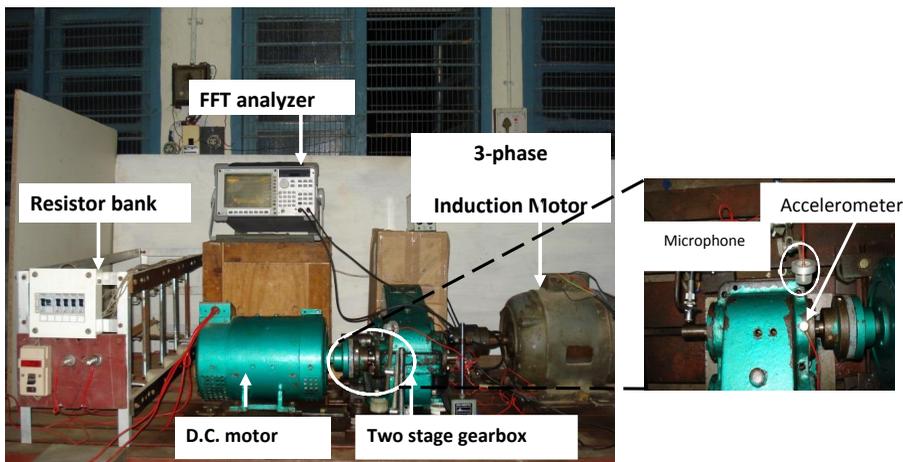


Figure 2: Photograph of experimental set up with sensors and equipments.

It is very difficult to study the fault detection procedures without seeded fault trials. Local faults in a gear box can be classified into three categories. (a) Surface wear spalling (b) cracked tooth and (c) loss of a part of tooth due to breakage of tooth at root or at a point on working tip (broken tooth or chipped tooth). There are different methods to simulate faults in gears viz. electric discharge machining (EDM), grinding and adding iron particles in gearbox lubricant and over loading the gear box i.e., accelerated test condition. The simplest approach is partial tooth removal.

### 3 Pre-Processing using Variational Mode Decomposition

Variational Mode decomposition decomposes the signal into various modes or intrinsic mode functions using calculus of variation. Each mode of the signal is assumed to have compact frequency support around a central frequency. VMD tries to find out these central frequencies and intrinsic mode functions centered on those frequencies concurrently using an optimization methodology called ADMM. The original formulation of the optimization problem is continuous in time domain.

VMD is formulated as; Minimize the sum of the bandwidths of  $k$  modes subject to the condition that sum of the  $k$  modes is equal to the original signal. The unknowns are  $k$  central frequencies and  $k$  functions centered at those frequencies. Since part of the unknowns is function, calculus of variation is applied to derive the optimal functions.

Bandwidth of an AM-FM signal primarily depends on both, with the maximum deviation of the instantaneous frequency  $\Delta f \sim \max(|\omega_k(t) - \omega_k|)$  and the rate of change of instantaneous frequency. Dragomiretskiy and Zosso proposed a function that can measure the bandwidth of an intrinsic mode function  $u_k(t)$ . At first they computed Hilbert transform of  $u_k(t)$ . Let it be  $u_k^H(t)$ . Then formed an analytic function  $(u_k(t) + ju_k^H(t))$ . The frequency spectrum of this function is one sided (exist only for positive frequency) and assumed to be centered on  $\omega_k$ . By multiplying this analytical signal with  $e^{-j\omega_k t}$ , the signal is frequency translated to be centered at origin. The integral of the square of the time derivative of this frequency translated signal is a measure of bandwidth of the intrinsic mode function  $u_k(t)$ .

$$\text{Let } u_k^M(t) = (u_k(t) + ju_k^H(t)) e^{-j\omega_k t}$$

It is a function whose spectrum is around origin (baseband). Magnitude of time derivative of this function when integrated over time is a measure of bandwidth. Hence,

$$\Delta\omega_k = \int (\partial_t (u_k^M(t))) \overline{(\partial_t (u_k^M(t)))} dt$$

$$\text{where, } \partial_t (u_k^M(t)) = \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right].$$

The integral can also expressed as a norm.

$$\Delta\omega_k = \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] \right\|_2^2$$

The sum of bandwidths of  $k$  modes is given by  $\sum_{k=1}^K \Delta\omega_k$

The resulting variational formulation is as follows:

$$\min_{u_k, \omega_k} \left\{ \sum_k \left\| \partial_t \left[ \left( \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right) e^{-j\omega_k t} \right] \right\|_2^2 \right.$$

$$s.t. \sum_k u_k = f$$

Where  $f$  is the original signal.

The augmented Lagrangian multiplier method converts this into an unconstrained optimization problem as follows:

$$L(u_k, w_k, \lambda) = \alpha \sum_k \left\| \partial_t \left[ \left( \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right) e^{-j\omega_k t} \right] \right\|_2^2$$

$$+ \left\| f - \sum_k u_k \right\|_2^2 + \left\langle \lambda, f - \sum_k u_k \right\rangle \quad (1)$$

In ADMM philosophy , one variable at a time is solved assuming all others are known.

Hence, the formula for updating  $u_k$  at the ‘ $n+1$ ’ the iteration is as follows:

Update for  $u$  terms

$$u_k^{n+1} = \arg \min_{u_k(t)} \alpha \left\| \partial_t \left[ \left( \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right) e^{-j\omega_k t} \right] \right\|_2^2$$

$$+ \left\| f - \sum_i u_i \right\|_2^2 + \left\langle \lambda, f - \sum_i u_i \right\rangle$$

By the absorbing the last inner product which is basically  $\int \lambda(t) \left( f(t) - \sum_i u_i(t) \right) dt$

in to the term  $\left\| f - \sum_i u_i \right\|_2^2 = \int \left( f(t) - \sum_i u_i(t) \right)^2 dt$ , then

$$\left\| f - \sum_i u_i \right\|_2^2 + \left\langle \lambda, f - \sum_i u_i \right\rangle = \left\| f - \sum_i u_i + \frac{\lambda}{2} \right\|_2^2$$

Therefore

$$u_k^{n+1} = \arg \min_{u_k(t)} \alpha \sum_k \left\| \partial_t \left[ \left( \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right) e^{-j\omega_k t} \right] \right\|_2^2 + \left\| f - \sum_i u_i + \frac{\lambda}{2} \right\|_2^2$$

This problem can be solved in spectral domain by noting the fact that norm in time domain is same as norm in frequency domain.

The following results are used in Fourier transform

$$u_k(t) \Leftrightarrow \hat{u}_k(\omega) \Rightarrow \partial_t (u_k(t)) \Leftrightarrow (j\omega) \hat{u}_k(\omega)$$

$$u_k(t) \Leftrightarrow \hat{u}_k(\omega) \Rightarrow \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) = u_k(t) + \frac{j}{\pi t} * u_k(t) \Leftrightarrow (1 + \text{sgn}(\omega)) \hat{u}_k(\omega)$$

Note that,

$$\text{for negative } \omega, (1 + \text{sgn}(\omega)) \hat{u}_k(\omega) = 0$$

$$\text{and for positive } \omega, (1 + \text{sgn}(\omega)) \hat{u}_k(\omega) = 2\hat{u}_k(\omega)$$

$$\begin{aligned} u_k(t) + \frac{j}{\pi t} * u_k(t) &\Leftrightarrow (1 + \text{sgn}(\omega)) \hat{u}_k(\omega) \Rightarrow \left( u_k(t) + \frac{j}{\pi t} * u_k(t) \right) e^{-j\omega_k t} \\ &\Leftrightarrow (1 + \text{sgn}(\omega + \omega_k)) \hat{u}_k(\omega + \omega_k) \end{aligned}$$

Therefore

$$u_k^{n+1} = \arg \min_{\hat{u}_k(\omega)} \alpha \|j\omega (1 + \text{sgn}(\omega + \omega_k)) \hat{u}_k(\omega + \omega_k)\|_2^2 + \left\| \hat{f} - \sum_i \hat{u}_i + \frac{\hat{\lambda}}{2} \right\|_2^2$$

Replacing  $\omega \rightarrow \omega + \omega_k$

$$u_k^{n+1} = \arg \min_{\hat{u}_k(\omega)} \alpha \|j(\omega - \omega_k) (1 + \text{sgn}(\omega)) \hat{u}_k(\omega)\|_2^2 + \left\| \hat{f} - \sum_i \hat{u}_i + \frac{\hat{\lambda}}{2} \right\|_2^2$$

In the above expression, the first term vanishes for negative frequencies

$$\begin{aligned} &\|(1 + \text{sgn}(\omega + \omega_k)) \hat{u}_k(\omega + \omega_k)\|_2^2 \\ &= \int_w (j(\omega - \omega_k) (1 + \text{sgn}(\omega)) \hat{u}_k(\omega)) \overline{(j(\omega - \omega_k) (1 + \text{sgn}(\omega)) \hat{u}_k(\omega))} d\omega \\ &= \int_0^\infty 4(\omega - \omega_k)^2 |\hat{u}_k(\omega)|^2 d\omega \end{aligned}$$

Second term is symmetric around origin, therefore

$$\begin{aligned} \left\| \hat{f}(\omega) - \sum_i \hat{u}_i + \frac{\hat{\lambda}}{2} \right\|_2^2 &= \int_{-\infty}^\infty \left( \hat{f} - \sum_i \hat{u}_i + \frac{\hat{\lambda}}{2} \right) \overline{\left( \hat{f} - \sum_i \hat{u}_i + \frac{\hat{\lambda}}{2} \right)} d\omega \\ &= 2 \int_0^\infty \left( \hat{f}(\omega) - \sum_i \hat{u}_i + \frac{\hat{\lambda}}{2} \right) \overline{\left( \hat{f} - \sum_i \hat{u}_i + \frac{\hat{\lambda}}{2} \right)} d\omega \end{aligned}$$

Also  $\left(\hat{f}(\omega) - \sum_i \hat{u}_i + \frac{\hat{\lambda}}{2}\right)$  being a complex number

$\left(\hat{f}(\omega) - \sum_i \hat{u}_i + \frac{\hat{\lambda}}{2}\right) \left(\overline{\hat{f}(\omega) - \sum_i \hat{u}_i + \frac{\hat{\lambda}}{2}}\right) = \left|\hat{f}(\omega) - \sum_i \hat{u}_i + \frac{\hat{\lambda}}{2}\right|^2$ , where  $\|\cdot\|$  represent magnitude of the complex number.

Therefore,

$$\hat{u}_k^{n+1} = \arg \min_{\hat{u}_k(\omega), \omega > 0} \int_0^\infty \left(4\alpha(\omega - \omega_k)^2 |\hat{u}_k(\omega)|^2 + 2 \left|\hat{f}(\omega) - \sum_i \hat{u}_i + \frac{\hat{\lambda}}{2}\right|^2\right) d\omega$$

Here unknown is a function. Hence, apply Euler Lagrangian condition to obtain the solution.

$$\text{Let } F = 4(\omega - \omega_k)^2 |\hat{u}_k(\omega)|^2 + 2 \left|\hat{f}(\omega) - \sum_i \hat{u}_i + \frac{\hat{\lambda}}{2}\right|^2$$

$$\frac{dF}{d\hat{u}_k} = 0 \Rightarrow 8\alpha(\omega - \omega_k)^2 \hat{u}_k + 4 \left(\hat{f}(\omega) - \sum_i \hat{u}_i + \frac{\hat{\lambda}}{2}\right) (-1) = 0$$

$$\Rightarrow 2\alpha(\omega - \omega_k)^2 \hat{u}_k + \hat{u}_k = \left(\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i + \frac{\hat{\lambda}}{2}\right)$$

$$\Rightarrow \hat{u}_k (1 + 2\alpha(\omega - \omega_k)^2) = \left(\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i + \frac{\hat{\lambda}}{2}\right)$$

$$\hat{u}_k^{n+1} = \left(\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i + \frac{\hat{\lambda}}{2}\right) \frac{1}{(1 + 2\alpha(\omega - \omega_k)^2)}, \quad \omega \geq 0$$

Update for  $\omega_k$  s

$$\omega_k^{n+1} = \arg \min_{\omega_k} \left\| \partial_t \left[ \left( \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right) e^{-j\omega_k t} \right] \right\|_2^2$$

$$\omega_k^{n+1} = \arg \min_{\omega_k} \|j\omega (1 + \text{sgn}(\omega + \omega_k)) \hat{u}_k(\omega + \omega_k)\|_2^2$$

$$\omega_k^{n+1} = \arg \min_{\omega_k} \|j(\omega - \omega_k) (1 + \text{sgn}(\omega)) \hat{u}_k(\omega)\|_2^2$$

$$\omega_k^{n+1} = \arg \min_{\omega_k} \int_0^\infty (\omega - \omega_k)^2 |\hat{u}_k(\omega)|^2 d\omega$$

Here

$\omega_k^{n+1}$  is given by the solution of  $\int_0^\infty \frac{d}{d\omega_k} \left( (\omega - \omega_k)^2 |\hat{u}_k(\omega)|^2 \right) d\omega = 0$

$$\int_0^\infty -2(\omega - \omega_k) |\hat{u}_k(\omega)|^2 d\omega = 0$$

$$\Rightarrow \omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega}$$

Update for  $\lambda$  (Lamda)

$$\lambda^{n+1} \leftarrow \lambda^n + \tau (f - u_k^{n+1}(t))$$

Final algorithm for VMD:

initialize  $\hat{u}_k^1, \hat{\omega}_k^1, \hat{\lambda}^1, n \leftarrow 0$

repeat

$n \leftarrow n + 1$

for  $k = 1 : K$  do

Update  $\hat{u}_k$  for all  $\omega \geq 0$

$$\hat{u}_k^{n+1} \leftarrow \frac{\hat{f} - \sum_{i < k} \hat{u}_i^{n+1} - \sum_{i > k} \hat{u}_i^n + \frac{\hat{\lambda}^n}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \quad (2)$$

Update  $\omega_k$ :

$$\omega_k^{n+1} \leftarrow \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 d\omega} \quad (3)$$

end for

Dual ascent for all  $\omega \geq 0$ :

$$\hat{\lambda}^{n+1} \leftarrow \hat{\lambda}^n + \tau (\hat{f} - \sum_k \hat{u}_k^{n+1}) \quad (4)$$

until convergence:  $\sum_k \|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2 / \|\hat{u}_k^n\|_2^2 < \varepsilon$

### 3.1 Discretization of Frequency

It is first assumed that length of the mirrored signal in the time domain is 1. If total length of the mirrored signal in terms of number of discrete values is T, then sampling interval is 1/T.

The discrete frequency is assumed to vary from -0.5 to +0.5 so that it represents normalized discrete frequency. It must be noted that algorithm construct Fourier transform of different mode function values for positive frequencies only. The other half can be easily created by conjugating and reflecting on the left side.

Once all the mode functions in the frequency domain are obtained, then obtain the time domain mode functions by taking inverse Fourier transform. These mode functions correspond to mirrored signal. Then cut off the appended (reflected portions) part of the signal to obtain the desired intrinsic mode functions.

## 4 Feature Extraction

Descriptive statistical parameters such as kurtosis, mean, variance and standard deviation extracted from the vibrational signals are computed to serve as features. They are named as ‘statistical features’ here. Brief descriptions about the extracted features are given below.

**(a) Standard deviation:** This is a measure of the effective energy or power content of the vibration signal. The following formula was used for computation of standard deviation.

$$\text{Standard Deviation} = \sqrt{\frac{\sum x^2 - (\sum x)^2}{n(n-1)}}$$

**(b) Sample variance:** It is variance of the signal points and the following formula was used for computation of sample variance.

$$\text{Sample Variance} = \frac{\sum x^2 - (\sum x)^2}{n(n-1)}$$

**(c) Kurtosis:** Kurtosis indicates the flatness or the spikiness of the signal. Its value is very low for normal condition of the gear and high for faulty condition of the gear due to the spiky nature of the signal.

$$\text{Kurtosis} = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left( \frac{x_i - \bar{x}}{s} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

where ‘s’ is the sample standard deviation.

(d) **Mean:** Mean is computed as arithmetic average of all points in the signal.

$$\text{Mean} = \sum_{i=1}^n x_i$$

## 5 Feature Selection using J48 Decision tree

All the statistical features extracted from the vibrational signals do not contribute equally to the classification accuracy. It may be observed that some features are significant for the classification process, while some are purely irrelevant. The process of selecting only the relevant statistical features for the classification process so as to reduce the computational effort is known as feature selection. In the present study, the dataset is used with J48 algorithm to generate the decision tree which facilitates the feature selection process. The generated decision tree is shown in Fig.3.

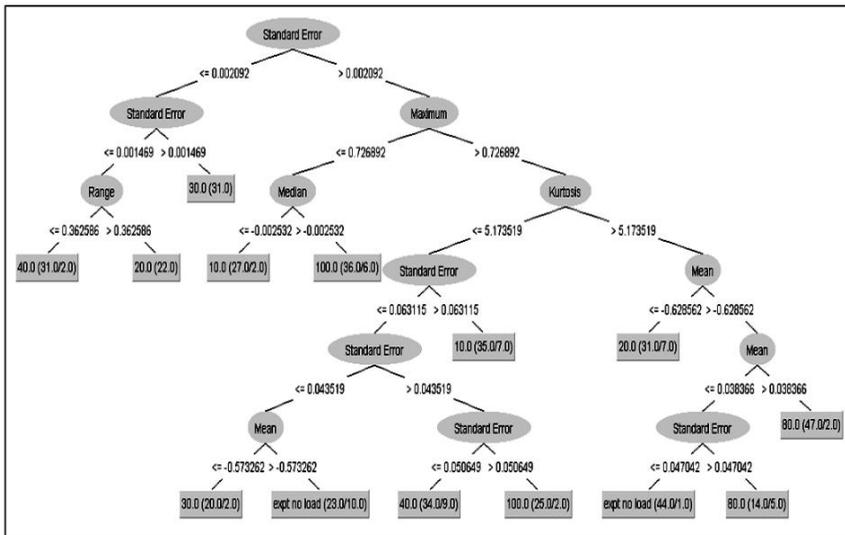


Figure 3: Decision tree for feature selection.

The features that are appearing on top of the decision tree are good for classification. The ones that do not appear are not useful for classification. The features appearing in the bottom of the tree are relatively less important ones. Hence, one can consciously choose or omit depending on the classification accuracy requirement and computational resources available.

## 6 Classifier

In machine learning, classification is considered an instance of supervised learning, i.e. learning where a training set of correctly identified observations is available. A path from the root to a leaf represents the rules for classification [Mohamed, Nor Haizan, Salleh, Mohd Najib Mohd Omar, Abdul Halim (2012); Breiman, Friedman, Olshen, Stone (1984)]. An algorithm that implements classification, especially in a concrete implementation, is known as a classifier. In the present study, three classifiers are used namely, Random Forest, REP Tree and Logistical Model Tree. A brief description is given below.

### 6.1 *Random Forest*

Random Forest algorithm is an ensemble learning method for classification that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes output by individual trees. The algorithm for inducing a random forest was developed by Leo Breiman [Breiman, and Leo (2001)] and Adele Cutler [Liaw, and Andy (2012)]. The term came from random decision forests that were first proposed by Tin Kam Ho of Bell Labs in 1995. The method combines Breiman's "bagging" idea and the random selection of features, introduced independently by Ho [Ho, Tin Kam (1995, 1998)] and Amit and Geman [Amit, Yali and Geman, Donald (1997)] in order to construct a collection of decision trees with controlled variance.

### 6.2 *REP Tree (Reduced Error Pruning Tree)*

One of the simplest forms of pruning is reduced error pruning. Starting at the leaves, each node is replaced with its most popular class. If the prediction accuracy is not affected then the change is kept. While somewhat naive, reduced error pruning has the advantage of simplicity and speed.

### 6.3 *Logistic Model Tree*

A logistic model tree (LMT) is a classification model with an associated supervised training algorithm that combines logistic regression (LR) and decision tree learning [Niels Landwehr, Mark Hall, and Eibe Frank (2003)] [Landwehr, Hall, Frank, (2005)]. Logistic model trees are based on the earlier idea of a model tree: a decision tree that has linear regression models at its leaves to provide a piecewise linear regression model (where ordinary decision trees with constants at their leaves would produce a piecewise constant model) [Zvokelj, Zupan, and Prebil (2010)]

## 7 Results and Discussion

A total of 420 vibrational signals were collected for normal and abnormal conditions from a helical gear box; 60 signals from each class. The statistical features extracted from these signals were selected as features and act as input to the algorithm. 50 signals were used for training and 10 signals were used for testing. The statistical features were treated as features and act as input to the algorithm. The corresponding status or condition (10% fault, 20% fault, 30% fault, 40% fault, 80% fault, 100% fault, Good) of the classified data will be the required output of the algorithm. This input and corresponding output together forms the dataset.

### 7.1 Effect of number of features on Classification Accuracy

All the descriptive features extracted from the vibrational signals do not contribute equally to the classification accuracy. The process of reducing the number of input features for classification is known as dimensionality reduction. Table 2 and Fig. 4 illustrate the variation of classification accuracy with change in the number of features.

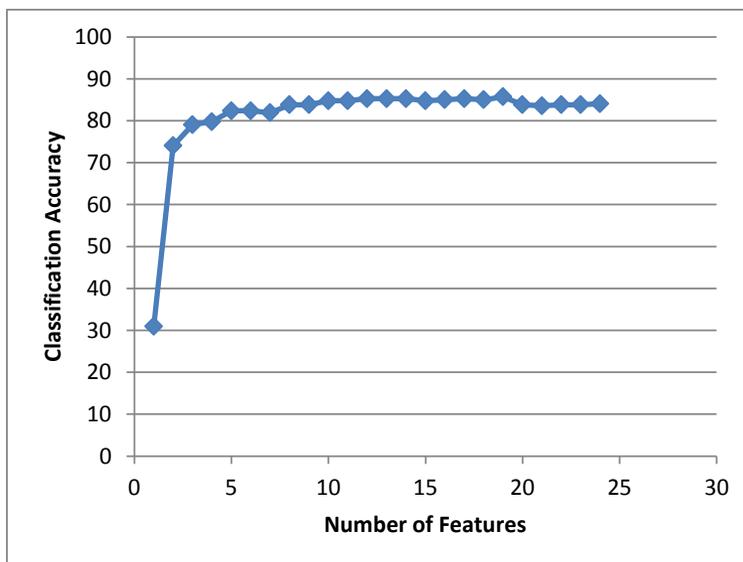


Figure 4: Effect of number of features on classification accuracy.

It is to be noted that the maximum classification accuracy is obtained when **19** features are being used instead of the total 24 features.



tends to over fit the data and when it is more the algorithm tends to generalize the model built.

Table 2: Detailed Class wise accuracy.

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.855	0.042	0.768	0.855	0.809	0.944	10% fault
0.767	0.018	0.868	0.767	0.814	0.963	20% fault
0.79	0.027	0.831	0.79	0.81	0.947	30% fault
0.857	0.021	0.871	0.857	0.864	0.975	40% fault
0.953	0.013	0.924	0.953	0.938	0.997	80% fault
0.828	0.024	0.855	0.828	0.841	0.943	100% fault
0.906	0.027	0.853	0.906	0.879	0.97	Good

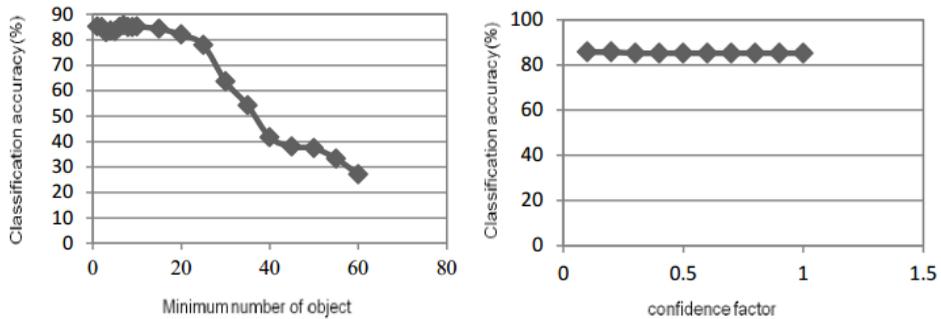


Figure 6: Minimum No of objects Vs Figure 7: Confidence factor Vs Classification accuracy.

### 7.3 Classification of VMD pre-processed signals using Logistic Model Tree Algorithm

Classification accuracy of 87.1429% was achieved using Logistic Model Tree (LMT). Fig.8 illustrates the confusion matrix obtained by this method.

The diagonal elements of the confusion matrix represent the correctly classified instances indicating an overall accuracy of **87.1429%**.

It is observed from the confusion matrix that there have been instances where good signals have been wrongly classified as having 10%, 20% and 30% faults. Though this is not harmful, it leads to wastage of time in verifying the gears. There have been other instances as well where one type of fault has been wrongly classified as other.

```

=== Confusion Matrix ===
  a  b  c  d  e  f  g  <-- classified as
45  8  0  0  3  3  1 | a = 10% fault
 7 51  1  0  1  0  0 | b = 20% fault
 0  2 53  3  0  0  2 | c = 30% fault
 0  2  4 52  0  2  0 | d = 40% fault
 3  0  0  0 57  0  0 | e = 80% fault
 1  0  2  1  0 55  1 | f = 100% fault
 0  2  2  2  1  0 53 | g = Good
    
```

Figure 8: Confusion matrix of Logical Model Tree algorithm.

In LMT, the minimum no. of instances to consider node splitting was varied from 0 to 20. Fig. 9 depicts the variation in the classification accuracy with change in number of instances. Therefore, the default value of 15 was set as the number of instances for maximum accuracy.

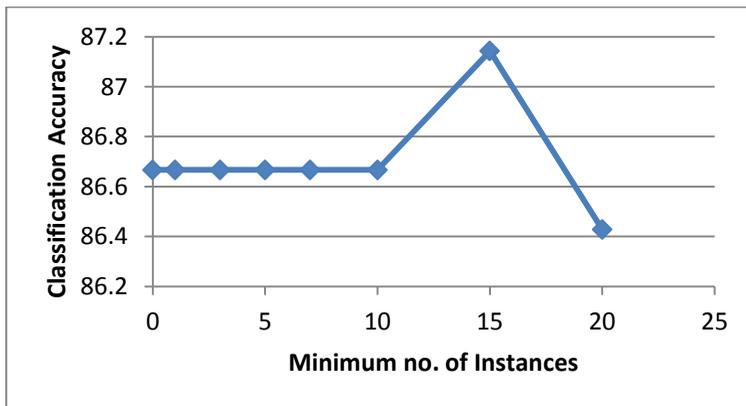


Figure 9: Minimum number of Instances vs Classification accuracy.

Figure 10 shows the variations of classification accuracy with change in the number of boosting iterations. Highest classification accuracy was achieved when the number of boosting iterations was set to 10.

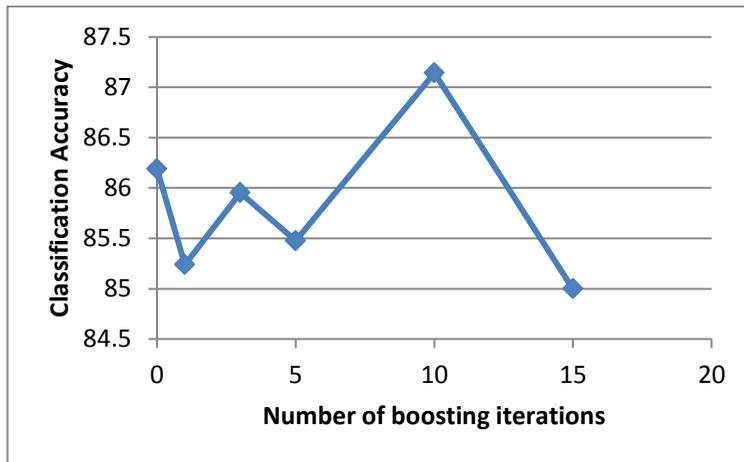


Figure 10: Number of boosting iterations vs Classification accuracy.

#### 7.4 Classification of VMD pre-processed signals using REP Tree Algorithm

This section discusses the results obtained from REP Tree Algorithm. Confusion matrix obtained by optimizing the parameters is shown in the Fig. 11. Reduced Error Pruning Tree method provides a maximum classification accuracy of **90.4762%**.

```

=== Confusion Matrix ===

  a  b  c  d  e  f  g  <-- classified as
50  4  0  1  2  3  0 | a = 10% fault
 8 52  0  0  0  0  0 | b = 20% fault
 0  2 54  1  0  2  1 | c = 30% fault
 0  1  0 57  0  2  0 | d = 40% fault
 1  0  0  0 58  0  1 | e = 80% fault
 4  0  1  3  0 52  0 | f = 100% fault
 0  0  1  0  1  1 57 | g = Good

```

Figure 11: Confusion matrix of REP Tree algorithm.

The variation of classification accuracy with change in maximum depth of tree is shown in Fig. 12. The accuracy varied from 14% to 90% when the depth of tree was changed from 0 to 10 with a maximum value attained when the depth of tree was assigned the value of 9. The maximum classification accuracy obtained is 90.4762%.

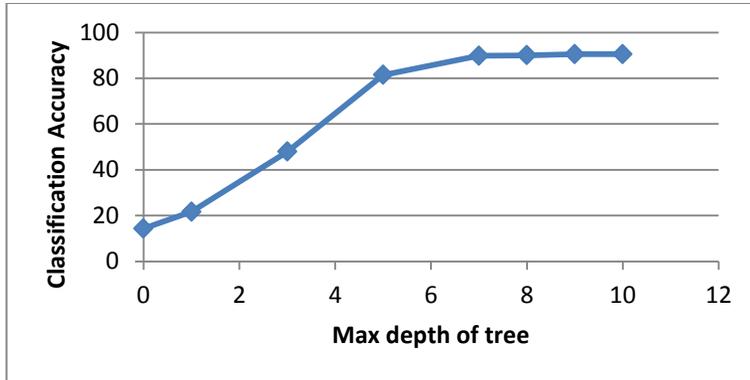


Figure 12: Max depth of tree vs Classification Accuracy.

The variation of classification accuracy with minimum number of objects is shown in Fig. 13. The accuracy varied from 89% to 80% when the number of objects was changed from 0 to 10 with a maximum value attained when the number of objects attribute was assigned the value of 0. The maximum classification accuracy obtained is 89.7619%.

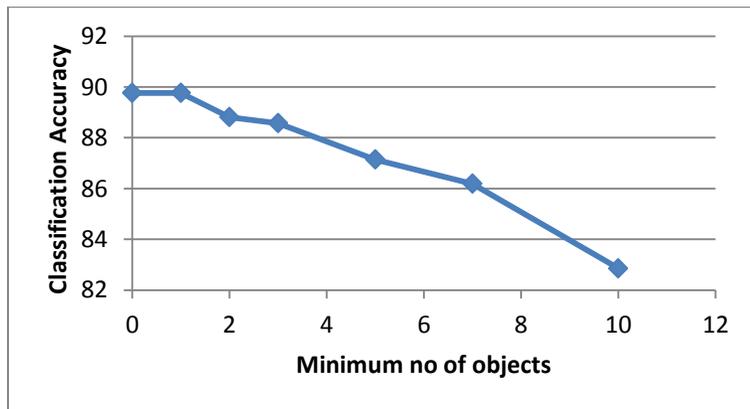


Figure 13: Minimum number of objects vs Classification Accuracy.

The variation of classification accuracy with minimum number of folds is shown in Fig. 14. The accuracy remained constant when the number of folds was changed from 2 to 10 as this attribute requires a value greater than 1. Number of folds was taken as the minimum value, ie 2 to reduce computational effort. The maximum classification accuracy obtained is 90.4762%.

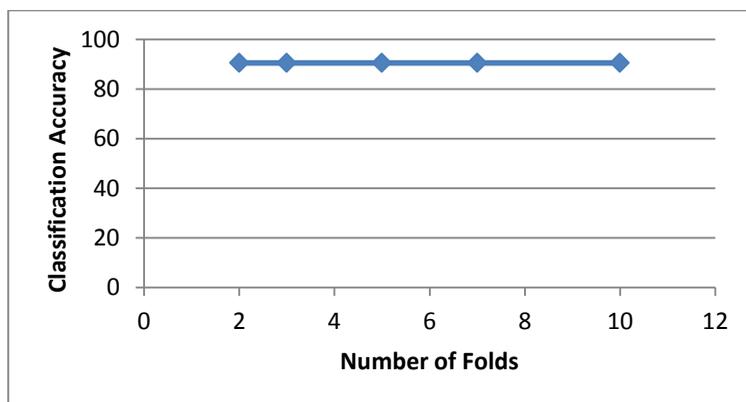


Figure 14: Number of folds vs Classification Accuracy.

### 7.5 Classification of VMD pre-processed signals using Random Forest Algorithm

This section discusses the results obtained from Random Forest Algorithm. Confusion matrix obtained by optimizing the parameters is shown in the Fig. 15. The diagonal elements of the confusion matrix represent the correctly classified instances indicating an overall accuracy of **91.4286%**. The Random Forest algorithm also provided an improvement over the other two classifiers used and hence gives the best result among the lot.

Figure 16 shows the variation in classification accuracy with change in the maximum allowed depth of trees (0 for unlimited depth). There was a decline in accuracy when the maximum depth was changed to values 1 and 2. For all other values an accuracy of 91.4286% was observed.

Similarly the number of attributes to be used in random selection was varied from 5 to 10 and the corresponding change in classification accuracy was observed (refer figure 17).

Figure 18 depicts the variation in classification accuracy on changing the number of trees from the minimum value of 1 to 15. It is observed that the accuracy in-

```

=== Confusion Matrix ===
  a  b  c  d  e  f  g  <-- classified as
50  6  0  0  2  1  1 | a = 10% fault
 6 54  0  0  0  0  0 | b = 20% fault
 0  2 53  1  0  0  4 | c = 30% fault
 0  0  0 56  0  3  1 | d = 40% fault
 1  0  0  0 59  0  0 | e = 80% fault
 0  0  2  3  0 55  0 | f = 100% fault
 1  1  1  0  0  0 57 | g = Good
    
```

Figure 15: Confusion Matrix of Random Forest algorithm.

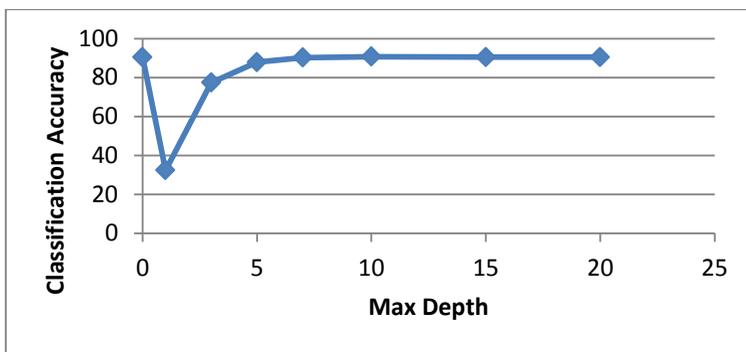


Figure 16: Max depth vs Classification Accuracy.

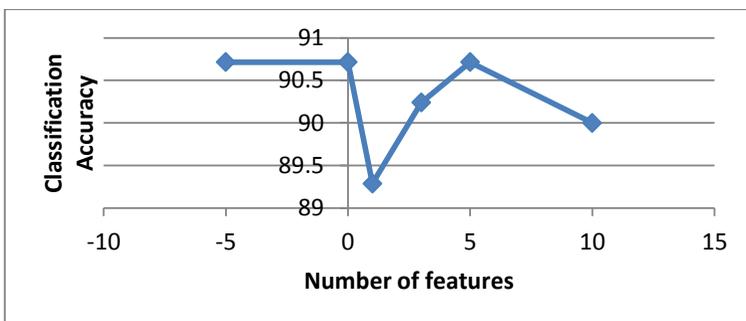


Figure 17: Number of features vs Classification Accuracy.

creased from 82% to 90% when the number of trees was increased from 1 to 10 and remained constant after that.

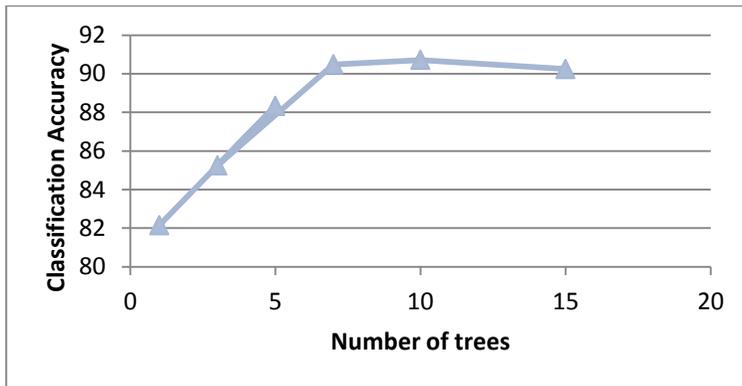


Figure 18: Number of trees vs Classification Accuracy.

The variation in classification accuracy on changing the random number seed to be used from 0 to 10 is shown in Fig. 19. Random variations were observed for seed values from 0 to 5 and on further increase in seed value the accuracy increased. Therefore the seed value was set to 0 to obtain the maximum accuracy of 91.4286%.

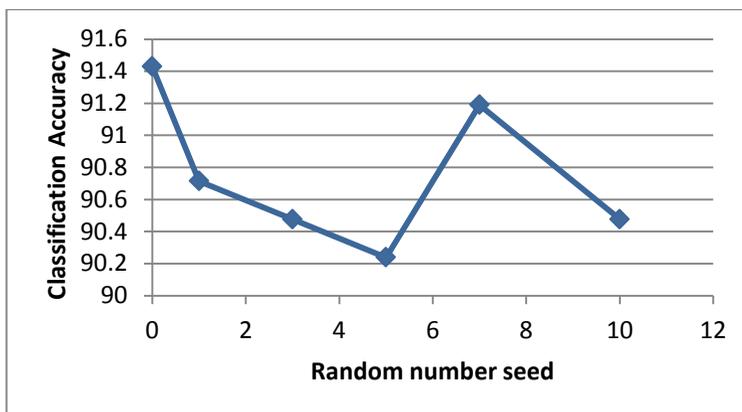


Figure 19: Random number seed vs Classification Accuracy.

## 7.6 Summary of Results

The different algorithms used for classifications of VMD preprocessed signals and the maximum classification accuracy obtained with each algorithm is listed in table 3 in descending order of accuracy.

Table 3: Classification accuracy obtained with VMD preprocessed signals using various decision tree classifiers.

SI No.	Algorithm	Classification Accuracy ( in % )
1	Random Forest	91.4286
2	REP Tree	90.4762
3	Logistic Model Tree	87.1429

Classifications performed using Random Forest Algorithm was able to provide maximum classification accuracy of 91.4286%.

## 8 Conclusion

All the machinery available in the industry, ranging from small scale machines to the heavy-duty industrial machines rely on common gears. Hence, this necessitates the need of monitoring the condition of gears. Faulty gears can affect the functionality of a machine to a large extent. Here, Machine Learning was used as a simple but powerful tool for fault diagnosis. The introduction of Variational Mode Decomposition (VMD) as a new signal pre-processing technique along with the different decision trees have provided outstanding performance characteristics with a classification accuracy reaching **91.4286%**. For bench marking the new features and classifier, statistical features extracted from raw signal (without VMD pre-processing) and various decision tree algorithms have been taken up. The accuracy achieved by VMD pre-processed vibration signals is far superior to that generated using the signals which were not VMD pre-processed (86%). From the results and discussions, one can conclude that VMD pre-processed signals with decision tree perform impeccably in fault diagnosis of helical gear box. Its ability to distinguish between good and faulty signals with more than 90% accuracy motivates its use in the industry.

## References

**Amarnath, M. V.; Sugumaran, D. J.; Kumar, H.** (2013): Fault Diagnosis of Helical Gear Box Using Decision Tree and Best-First Tree, *International Journal of Research in Mechanical Engineering*, vol. 1, Issue 1, pp. 22-33.

- Amit, Y.; Geman, D.** (1997): Shape quantization and recognition with randomized trees. *Neural Computation*, vol. 9, no. 7, pp. 1545–1588.
- Benbouzid, M. E. H.; Kliman, G. B.** (2003): What stator current processing-based technique to use for induction motor faults diagnosis? *IEEE trans. Energy Conversion*, vol. 18, no. 2, pp. 238-244.
- Breiman, L.; Friedman, J.; Stone, C. J.; Olshen, R. A.** (1984): Classification and regression trees. *Monterey, CA: Wadsworth*.
- Breiman, L.** (2001): Random Forests. *Machine Learning*, vol. 45, no. 1, pp. 5–32.
- Gao, Q.; Duan, C.; Fan, H.; Meng, Q.** (2008): Rotating machine fault diagnosis using Empirical mode decomposition. *Mechanical systems and signal processing*, vol. 22, no. 5, pp. 1072-81.
- Gao, D.; Wang, Y.; Wu, Z.; Gorgin, R.** (2013): Damage Extension Diagnosis Method for Typical Structures of Composite Aircraft Based on Lamb Waves. *SDHM: Structural Durability & Health Monitoring*, vol. 9, no. 3, pp. 233-252.
- Huang, N. E. et al.** (1998): The empirical mode decomposition and Hilbert spectrum for non-linear non-stationary time series analysis. *Proceedings of Royal Society of London*, vol. 454, no. 1971, pp. 903-995.
- Ho, T. K.** (1995): Random Decision Forest. *Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC*, 14–16 August, pp. 278–282.
- Ho, T. K.** (1998): The Random Subspace Method for Constructing Decision Forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844.
- Landwehr, N.; Hall, M.; Frank, E.** (2003): Logistic model trees. *ECML PKDD*.
- Landwehr, N.; Hall, M.; Frank, E.** (2005): Logistic Model Trees. *Machine Learning*, vol. 59, no. 1-2, pp. 161-205.
- Lei, Y.; Zuo, M. J.; He, Z.; Zi, Y.** (2010): A multidimensional hybrid intelligent method for gear fault diagnosis. *Expert systems with applications*, vol. 37, no. 2, pp. 1419-1430.
- Li, C. J.; Ma, J.** (1997): Wavelet decomposition of vibrations for detection of bearing localized defects. *NDT&E International*, vol. 30, no. 3, pp. 143-149.
- Li, C. J.; Ma, J.** (1997): Wavelet decomposition of vibrations for detection of bearing localized defects. *NDT&E International*, vol. 30, no. 3, pp. 143-149.
- Liaw, A.** (2012): Documentation for R package random Forest. Retrieved 15 March 2013.
- Loutridis, S. J.** (2004): Damage detection in gear systems using Empirical mode

decomposition. *Engineering Structures*, vol. 26, no. 12, pp. 1833-41.

**Mallardo, V.; Aliabadi, M. H.** (2013): Optimal Sensor Placement for Structural, Damage and Impact Identification: A Review. *SDHM: Structural Durability & Health Monitoring*, vol. 9, no. 4, pp. 287-323.

**Mba, D.; Rao, R. B. K. N.** (2006): Development of Acoustic Emission Technology for Condition Monitoring and Diagnosis of Rotating Machines: Bearings, Pumps, Gearboxes, Engines, and Rotating Structures. *Shock and Vibration Digest*, vol. 38, no. 1, pp. 3-16.

**Miyachika, K.; Oda, S.; Koide, T.** (1995): Acoustic Emission of Bending Fatigue Process of Spur Gear Teeth. *Journal of Acoustic Emission*, vol. 13, no. 1-2, pp. S47-S53.

**Miyachika, K.; Zheng, Y.; Tsubokura, K.** (2002): Acoustic Emission of bending fatigue process of supercarburized spur gear teeth, Progress in Acoustic Emission XI, Anonymous. *The Japanese Society for NDI*, pp. 304-310.

**Mohamed, W. N. H. W.; Salleh, M. N. M.; Omar, A. H.** (2012): A comparative study of Reduced Error Pruning method in decision tree algorithms. *Control System, Computing and Engineering (ICCSCE)*, 2012.

**Nandi, S.; Toliyat, H. A.** (1999): Condition monitoring and fault diagnosis of electrical machine, in Proc. *IEEE IAS Annual Meeting Conference*, vol. 1, pp. 197-204.

**Paul, B.** (1979): *Kinematics and Dynamics of Planar Machinery*. Prentice Hall.

**Quinlan, J. R.** (1986): Induction of Decision Trees. *Machine Learning*, vol. 1, pp. 81-106.

**Ricci, R.; Pennacchi, P.** (2011): Diagnostics of gear faults based on EMD and automatic selection of intrinsic mode functions. *Mechanical systems and Signal Processing*, vol. 25, pp. 821-838.

**Singh, A.; Houser, D. R.; Vijayakar, S.** (1999): Detecting Gear Tooth Breakage using Acoustic Emission: A Feasibility and Sensor Placement Study. *Journal of Mechanical Design, Transactions of the ASME*, vol. 121, no.4, pp. 587-593.

**Rilling, G.; Flandrin, P.; Goncalves, P.** (2003): On Empirical Mode Decomposition and its Algorithms. *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, vol. 3, NSIP-03, 2003, Grado (I).

**Staszewski, W. J.; Tomilson, G. R.** (1994): Applications of wavelet transform to fault detection in spur gear. *Mechanical Systems and Signal Processing*, vol. 8, no. 3, pp. 289-307.

**Sugumaran, V.; Jain, D.; Amarnath, M.; Kumar, H.** (2013): Fault Diagnosis of Helical Gear Box using Decision Tree through Vibration Signals. *International*

*Journal of Performability Engineering*, vol. 9, no. 2, pp. 221.

**Tinta, D.; Petrovčič, J.; Benko, U. et al.** (2005): Fault diagnosis of vacuum cleaner motors. *Control Engineering Practice*, vol. 13, pp. 177–187.

**Uicker, J. J.; Pennock, G. R.; Shigley, J. E.** (2003): *Theory of Machines and Mechanisms (3rd ed.)*. New York: Oxford University Press. ISBN 9780195155983.

**Wheitner, J.; Houser, D.; Blazakis, C.** (1993): Gear tooth bending fatigue crack detection by acoustic emission and tooth compliance. *ASME*, 93FTM9.

**Wuxing, L.; Peter, W. T.; Guicai, Z.; Tielin, S.** (2004): Classification of gear faults using cumulant and the radial basis function. *Mechanical Systems and Signal Processing*, vol. 18, pp. 381-389.

**Žvokelj, M.; Zupan, S.; Prebil, I.** (2010): Multivariate and multiscale monitoring of large-size low-speed bearings using Ensemble Empirical Mode Decomposition method combined with Principal Component Analysis. *Mechanical Systems and Signal Processing*, vol. 24, pp. 1049-1067.

**Žvokelj, M.; Zupan, S.; Prebil, I.** (2010): Multivariate and multiscale monitoring of large-size low-speed bearings using Ensemble Empirical Mode Decomposition method combined with Principal Component Analysis. *Mechanical Systems and Signal Processing* vol. 24, pp. 1049-1067.