

# Fault Diagnosis of Helical Gear Box Using Variational Mode Decomposition with Naïve Bayes and Bayes Net Classifiers through Vibration Signals

A. Kannan<sup>1\*</sup>, V. Sugumaran<sup>1</sup>, M. Amarnath<sup>2</sup>, K. P. Soman<sup>3</sup>

<sup>1</sup>School of Mechanical and Building Sciences, VIT University, Chennai Campus, India

<sup>2</sup>Department of Mechanical Engineering, Indian Institute of Information Technology Design and Manufacturing, India

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

## Abstract

*Gear is a very important machine component which finds use in most of the machines which requires some sort of power transmission or reduction. Unattended faults in gears can be catastrophic for the machine leading to halts in production and consequent economic loss. The requirement for a gear fault detection and diagnosis system is thus emphasized. Vibration signals are often used in fault diagnosis applications along with Fast Fourier transform method. However it is not effective with non-stationary signals like those obtained from gears in motion. Therefore, development of new methodologies to obtain diagnostic information from such signals is required. This paper details the use of vibration signals obtained from gears in good and simulated faulty conditions after performing preprocessing with Variational Mode Decomposition (VMD). VMD decomposes the vibration signals into various modes by identifying a compact frequency support around its central frequency so that adding all the modes reconstructs the original signal. Alternating Direction Multiplier Method (ADMM) is used in VMD to find the intrinsic mode functions. Descriptive statistical features extracted from VMD preprocessed signals, classified using Naïve Bayes and Bayes Net classifiers, and corresponding classification accuracies were calculated. The results were compared with the accuracy obtained from the statistical features extracted from the raw signal and decision tree classifier.*

**Keywords:** Fault diagnosis, naive Bayes, variational mode decomposition, vibration signal, alternating direction multiplier method, Bayes net, gear fault diagnosis.

## \*Correspondence Author

Email ID: a.kannan2011@vit.ac.in

## INTRODUCTION

Helical gearboxes are vital components of various machineries. They are used to run the machines at different gear ratios and for seamless transmission of power. Typical failures in helical gearboxes are caused by localized defects that arise when a part of the gear material with sufficient size is dislodged during operation of the gear. This happens mostly due to fatigue cracking under cyclic

contact stressing. <sup>[1]</sup> Impending failures of gears are therefore often alerted based on the early detection of localized defects. The severity of localized defects in gears cannot be measured while in operation.

Hence, various physical parameters like sound, vibration, wear debris and acoustic emission are measured for the detection and diagnosis of faults right from its early stages. The vibration signals of the defects

are often in disguise due to sound and vibrations produced by the operation of the machines. <sup>[2-4]</sup>

Transient excitations are observed in vibrations as a result of the impacts caused by localized faults in helical gears. When Fast Fourier transform (FFT) of these vibration signals are taken, it is observed that the frequency components are distributed due to overlapping of harmonics and also the presence of noise.

The frequency component itself is changed due to non-stationary nature of the signals. Due to these reasons conventional measurement methods for statistical parameters may not be effective for non-stationary signals obtained from gears in motion. <sup>[5]</sup> Therefore many researchers shifted their attention to signal processing methods to improve fault classification tools.

Pennachhi and Ricci used Empirical Mode Decomposition (EMD) to detect incipient faults in gears with intrinsic mode functions (IMF). <sup>[6]</sup> EMD lacks a mathematical theory foundation. N. Saravanan *et al.* <sup>[7]</sup> presented the effectiveness of helical gearbox fault diagnosis with wavelet based features using Proximal Support Vector Machines (PSVM) and Artificial Neural Network (ANN).

J48 algorithm was used to classify statistical features obtained from Morlet wavelet coefficients and the dominant features were given as input to train and test PSVM and ANN, and their classification accuracies were compared. Wavelets are able to represent signals in a Time-Frequency plane, but it has certain limitations. <sup>[8,9]</sup> A new processing technique was developed in order to overcome the drawbacks of other

techniques. It decomposed the signals into various modes or intrinsic mode functions using calculus variations. These modes have compact frequency support around a central frequency. Such central frequencies were identified using Alternating Direction Multiplier Method (ADMM) as an optimization tool. Various components of a signal can therefore be identified by the decomposition of a signal.

This study focuses on Variational Mode Decomposition (VMD) which is a new algorithm that extracts different modes present in the signal. An attempt is made to exploit vibration signals in order to diagnose faults in helical gearboxes. Vibration signals obtained were first preprocessed using VMD to find the modes and IMFs and then various descriptive statistical features like mean, standard deviation, kurtosis and variance were extracted.

These extracted features were classified using Naïve Bayes classifier and the obtained classification accuracy was compared with that obtained using raw vibration signals without any preprocessing.

## EXPERIMENTAL SETUP AND PROCEDURE

The experimental setup consists of a 5 hp two-stage helical gearbox which is driven by a 5.5 hp 3-phase induction motor having a rated speed of 1440 rpm. For this study the speed is set at 80 rpm, i.e. the speed of the 1<sup>st</sup> stage of the motor is 80 rpm.

The speed of pinion shaft in the 2<sup>nd</sup> stage of the gearbox becomes 1200 rpm with a step up ratio of 1:15. Figure 1 illustrates the experimental setup and Table 1 gives the specifications of the helical gearbox.

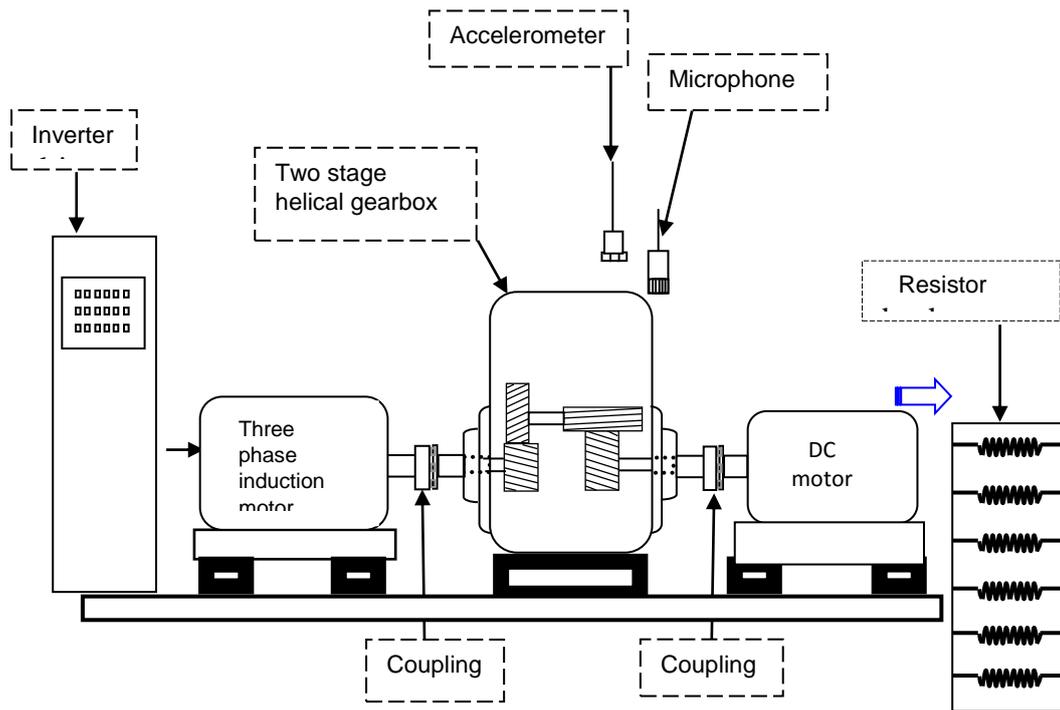


Fig. 1: Experimental Set Up.

Table 1: Specifications of Helical Gearbox.

	1 <sup>st</sup> Stage	2 <sup>nd</sup> Stage
Number 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
Shaft speed (rpm)	80 (input)	1200 (output)
Mesh frequency (Hz)	59	320
Step-up ratio	1.15	
Rated power (hp)	5	
Transmitted power (hp)	2.6	

A DC motor (used as generator) is connected to the pinion and generates 2 kW powers, which gets dissipated in a resistor bank. Therefore the actual load on the gearbox is only 2.6 hp, i.e., 52% of the rated 5 hp power. Utilization of load varies from 50% to 100% in industries. Additional torsional vibrations arising due to torque fluctuation in traditional dynamometer are avoided by the use of DC motor and resistor bank.

Backlash in the system is limited to the gears by the use of tire couplings amongst the electrical machines. Motor, generator and gearbox are mounted on I-beams which are anchored on to massive foundation. Bruel and Kjaer accelerometer installed close to the test gear measures the vibration signals. The signals are sampled at a frequency of 8.2 kHz. Figure 2 shows the experimental setup with sensors and equipment.

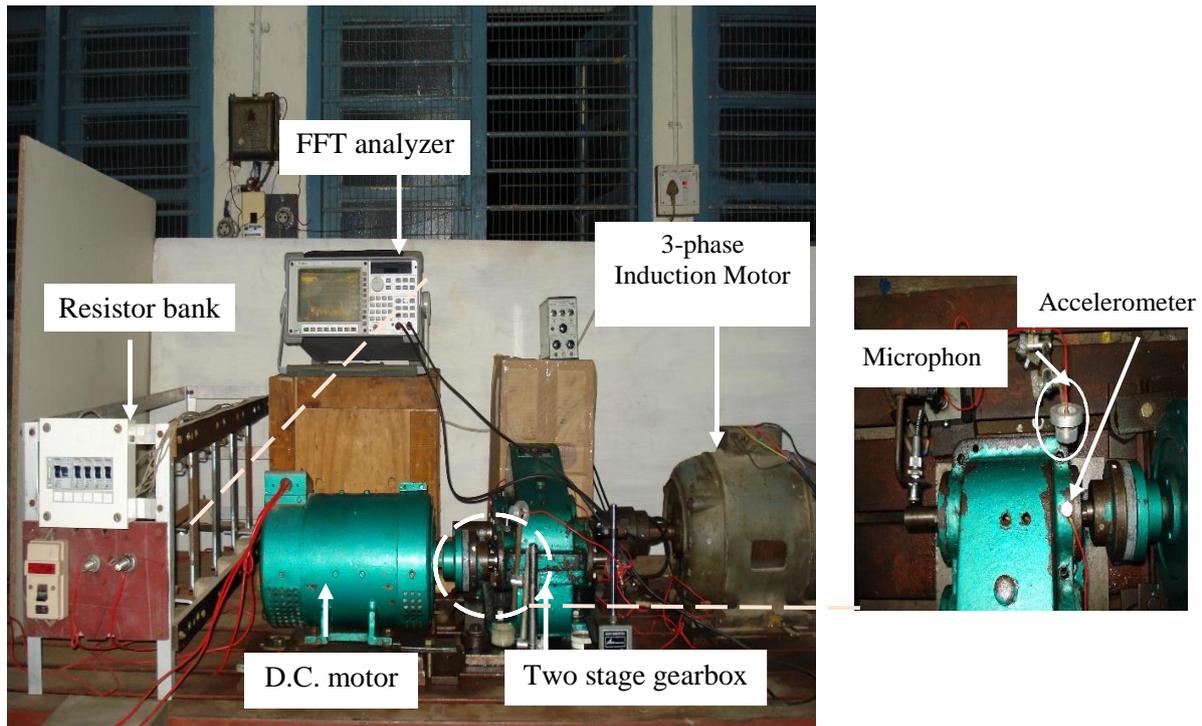


Fig. 2: Photograph: Experimental Set-Up with Equipment and Sensors.

A new gear box has an overhaul time period of more than one year. Study of fault detection procedures without seeded fault trials is very difficult. Local faults found in a gearbox can be categorized into three namely, surface wear, cracked tooth, spalling and loss of part of tooth due to breakage at a point on root or tip. Three different methods may be followed to simulate faults in gears namely, grinding, electric discharge machining (EDM) and adding iron particles in gearbox lubricant and overloading the gearbox (accelerated test condition). Partial teeth removal is the simplest method and hence was followed for this study. Partial tooth break which is common in many industrial applications is thus simulated.

**Feature Extraction**

Descriptive statistical parameters such as mean, standards deviation, kurtosis and variance were computed from the vibration signals to serve as features. These are termed as ‘statistical features’ here. Given below are brief descriptions on the extracted features.

**Mean:** It is the arithmetic average of all points in the signal.

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

I. **Standard deviation:** It is a measure of the effective energy of the vibration signal.

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

II. **Kurtosis:** It indicates the spikiness or flatness of the signal. The kurtosis value is very low for a bearing in normal condition and high for bearing in faulty condition.

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

III. **Sample variance:** It is the variance of the signal points.

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

**Variational Mode Decomposition**

VMD 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, it 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,

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

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

where,

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

$$\text{given by } \sum_{k=1}^K \Delta\omega_k$$

The resulting variational formulation is as follows:

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

$$\text{s.t. } \sum_k u_k = f$$

Where  $f$  is the original signal.

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

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

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;

$$u_k^{n+1} = \underset{u_k(t)}{\text{argmin}} \alpha \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f - \sum_i u_i \right\|_2^2 + \langle \lambda, f - \sum_i u_i \rangle$$

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;

$$\begin{aligned} \left\| f - \sum_i u_i \right\|_2^2 &= \int \left( f(t) - \sum_i u_i(t) \right)^2 dt, \text{ then} \\ \left\| f - \sum_i u_i \right\|_2^2 + \left\langle \lambda, f - \sum_i u_i \right\rangle &= \left\| f - \sum_i u_i + \frac{\lambda}{2} \right\|_2^2 \end{aligned}$$

Therefore,

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

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.

$$\begin{aligned} u_k(t) &\Leftrightarrow \hat{u}_k(\omega) \Rightarrow \partial_t(u_k(t)) \Leftrightarrow (j\omega)\hat{u}_k(\omega) \\ u_k(t) &\Leftrightarrow \hat{u}_k(\omega) \Rightarrow \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) = u_k(t) + \frac{j}{\pi t} * u_k(t) \Leftrightarrow (1 + \operatorname{sgn}(\omega))\hat{u}_k(\omega) \end{aligned}$$

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

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

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

Therefore

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

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

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

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

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

Second term is symmetric around origin, therefore;

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

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

$$\left( \hat{f}(\omega) - \sum_i \hat{u}_i + \frac{\hat{\lambda}}{2} \right) \left( \overline{\hat{f} - \sum_i \hat{u}_i + \frac{\hat{\lambda}}{2}} \right) = \left| \hat{f} - \sum_i \hat{u}_i + \frac{\hat{\lambda}}{2} \right|^2$$

Where  $|\cdot|$  represent magnitude of the complex number. Therefore,

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

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

$$\begin{aligned} \text{Let } F &= 4(\omega - \omega_k)^2 |\hat{u}_k(\omega)|^2 + 2 \left| \hat{f} - \sum_i \hat{u}_i + \frac{\hat{\lambda}}{2} \right|^2 \\ \frac{dF}{d\hat{u}_k} &= 0 \Rightarrow 8\alpha(\omega - \omega_k)^2 \hat{u}_k + 4 \left( \hat{f} - \sum_i \hat{u}_i + \frac{\hat{\lambda}}{2} \right) (-1) = 0 \\ \Rightarrow 2\alpha(\omega - \omega_k)^2 \hat{u}_k + \hat{u}_k &= \left( \hat{f} - \sum_{i \neq k} \hat{u}_i + \frac{\hat{\lambda}}{2} \right) \Rightarrow \hat{u}_k (1 + 2\alpha(\omega - \omega_k)^2) = \left( \hat{f} - \sum_{i \neq k} \hat{u}_i + \frac{\hat{\lambda}}{2} \right) \\ \hat{u}_k^{n+1} &= \left( \hat{f} - \sum_{i \neq k} \hat{u}_i + \frac{\hat{\lambda}}{2} \right) \frac{1}{(1 + 2(\omega - \omega_k)^2)}, \quad \omega \geq 0 \end{aligned}$$

Update for  $\omega_k$  s

$$\begin{aligned} \omega_k^{n+1} &= \underset{\omega_k}{\operatorname{argmin}} \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ \omega_k^{n+1} &= \underset{\omega_k}{\operatorname{argmin}} \left\| j\alpha(1 + \operatorname{sgn}(\omega + \omega_k)) \hat{u}_k(\omega + \omega_k) \right\|_2^2 \\ \omega_k^{n+1} &= \underset{\omega_k}{\operatorname{argmin}} \left\| j(\omega - \omega_k)(1 + \operatorname{sgn}(\omega)) \hat{u}_k(\omega) \right\|_2^2 \\ \omega_k^{n+1} &= \underset{\omega_k}{\operatorname{argmin}} \int_0^{\infty} (\omega - \omega_k)^2 |\hat{u}_k(\omega)|^2 d\omega \end{aligned}$$

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

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

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

Update for  $\lambda$  (Lamda)  
 $\hat{\lambda}^{n+1} \leftarrow \hat{\lambda}^n + \tau(\hat{f} - \sum_k \hat{u}_k^{n+1}(t))$

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  $\omega \geq 0$   $\hat{u}_k^{n+1} \leftarrow \frac{\hat{f} - \sum_{i < k} \hat{u}_i^{n+1} - \sum_{i > k} \hat{u}_i^n + \frac{\hat{\lambda}^n}{2}}{1 + 2\alpha(\omega - \hat{\omega}_k^n)^2}$   
 Update  $\omega_k$ :  
 $\hat{\omega}_k^{n+1} \leftarrow \frac{\int_0^{\hat{\lambda}} \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^{\hat{\lambda}} |\hat{u}_k^{n+1}(\omega)|^2 d\omega}$   
**end for**  
 Dual ascent for all  $\omega \geq 0$ :  
 $\hat{\lambda}^{n+1} \leftarrow \hat{\lambda}^n + \tau(\hat{f} - \sum_k \hat{u}_k^{n+1})$  **until convergence:**  $\sum_k \|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2 / \|\hat{u}_k^n\|_2^2 < \epsilon$

**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.

**Classifier**

All sample training instances are mapped into different groups by a classifier. In the present study the instances are classified as 10%, 20%, 30%, 40%, 80%, 100% faulty and normal conditions. 60 training instances are made available for each condition making a total of 420 instances. Equal number of training data ensures that the learning algorithm is no biased for any condition.

**Naïve Bayes**

In case of Naïve Bayes algorithm, the attributes  $A_1, A_2 \dots A_n$  are all assumed to be

conditionally independent of one another given  $B$ . Thereby the representation of  $P(A/B)$  and the problem of estimating it from the training data are simplified. We

have for  $n$  attributes of  $A$  which are conditionally independent of  $B$ ,

$$P(A_1 \dots A_n | B) = \prod_{i=1}^n P(A_i | B) \tag{Eq. (1)}$$

When  $A_i$  and  $B$  are Boolean variables we require  $2n$  parameters to define  $P(A_i = A_{ik} | B = b_j)$  for necessary  $i, j, k$ . This is a significant reduction. Then  $2(2^n - 1)$  parameters are needed to characterize  $P(A/B)$  for the case when the previously

stated assumption is not made. The main objective of this is to train a classifier so as to provide the distribution possible values of  $B$  for each instance of  $A$ . The following expression gives the probability that  $B$  will take on its  $k^{\text{th}}$  possible value.

$$P(B = b_k | A_1 \dots A_n) = \frac{P(B=b_k)P(A_1 \dots A_n | B=b_k)}{\sum_j P(B=b_j)P(A_1 \dots A_n | B=b_j)} \tag{Eq. (2)}$$

For all possible values of  $b_j$  of  $B$ , the sum is taken and on applying the above

discussed conditions, the equation for Naïve Bayes is obtained.

$$P(B = b_k | A_1 \dots A_n) = \frac{P(B=b_k) \prod_i P(A_i | B=b_k)}{\sum_j P(B=b_j) \prod_i P(A_i | B=b_j)} \tag{Eq. (3)}$$

**Bayes Net**

Bayesian Network is a probabilistic graphical model. It is a directed acyclic graph (DAG), which implies that it is a

network of random variables that make the node and the nodes are related to their conditional dependencies.

$$P(X_i | \text{parents}(X_i)) \text{ for each node } X_i$$

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | X_1, \dots, X_{n-1}) \tag{Eq. (4)}$$

$$= \prod_{i=1}^n P(X_i | \text{parents}(X_i)) \tag{Eq. (5)}$$

For  $\text{parents}(X_i) \subseteq \{X_1, \dots, X_{i-1}\}$

The joint probability distribution is computed using chain rule. Over fitting data and accommodation of missing values are key advantages of using Bayesian network.

Statistical features were considered as features and these serve as input to the algorithm. The required output of the algorithm is the classification into the above said faulty or normal conditions.

**RESULT AND DISCUSSION**

Vibration signals were collected for normal, 10%, 20%, 30%, 40%, 80% and 100% faulty conditions of helical gearbox. A total of 420 samples were recorded with 60 from each class. Among these 420 signals, 350 were used for training and remaining 70 were used for testing.

The input and resultant output together forms the dataset- Classification of non-preprocessed signals using Naïve Bayes and Bayes net classifiers:

Vibration signals without any VMD preprocessing performed, were used for this classification so that the result can be

compared with those obtained using VMD preprocessing, and Naïve Bayes and Bayes Net classifiers on the same vibration signals. M. Amarnath *et al.* [11] conducted a study to obtain maximum classification accuracy using Naïve Bayes

and Bayes Net classifiers on statistical features extracted from raw vibration signals. The confusion matrix obtained using Naïve Bayes and Bayes Net classifiers in the study are given in Table 2 and Table 3, respectively.

**Table 2: Confusion Matrix for Naïve Bayes Classifier Using Raw Vibration Signals.**

Classified actual as \	10% fault	20% fault	30% fault	40% fault	80% fault	100% fault	Good
10% fault	49	0	0	0	2	9	2
20% fault	10	46	2	2	0	0	0
30% fault	0	4	49	1	1	1	6
40% fault	1	0	3	55	0	3	1
80% fault	2	0	0	0	62	0	0
100% fault	8	0	0	7	0	49	0
Good	1	0	2	4	0	1	56

**Table 3: Confusion Matrix for Bayes Net Classifier Using Raw Vibration Signals.**

Classified actual as \	10% fault	20% fault	30% fault	40% fault	80% fault	100% fault	Good
10% fault	50	2	0	0	1	7	2
20% fault	10	46	2	2	0	0	0
30% fault	1	5	45	3	1	0	7
40% fault	1	0	4	54	0	3	1
80% fault	2	0	0	0	62	0	0
100% fault	8	1	0	8	0	46	1
Good	0	1	2	1	0	4	56

It was observed that 83.37% classification accuracy was obtained using Naïve Bayes classifier and 81.77% classification accuracy was obtained using Bayes Net classifier when statistical features were extracted from raw vibration signals without any preprocessing performed.

**Effect of Number of Features Selected from VMD Preprocessed Signals**

The vibration signals were preprocessed with VMD to find its frequency components, also called as Intrinsic Mode

Functions (IMF). The IMFs were arranged in ascending order of frequencies with every  $n^{\text{th}}$  mode having the maximum information content and highest frequency.

A total of 24 features were extracted from the vibration signals which could prove helpful in differentiating faulty and normal conditions, however some of these features may not contribute much to the

classification and can be excluded from the classification.

The 24 features are the mean, standard deviation, kurtosis and variance of six modes. This leads to two advantages, viz. reduced computation time and potentially better classification accuracy.

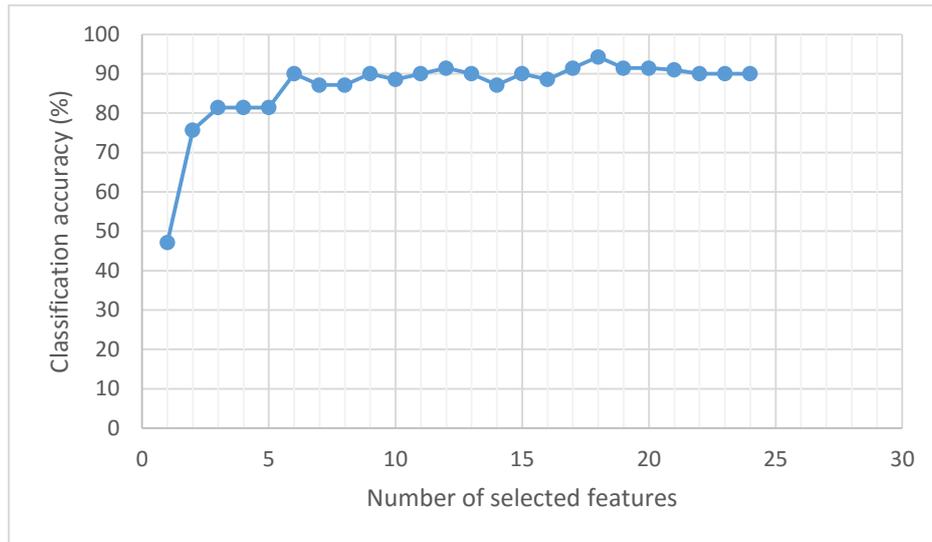
Sugumaran *et al.* [10] have elaborated the use of J48 decision tree to determine the rank of importance of the statistical features by visualizing the tree by depth first technique.

Implementing the above mentioned technique, the following features were obtained in descending order of priority.

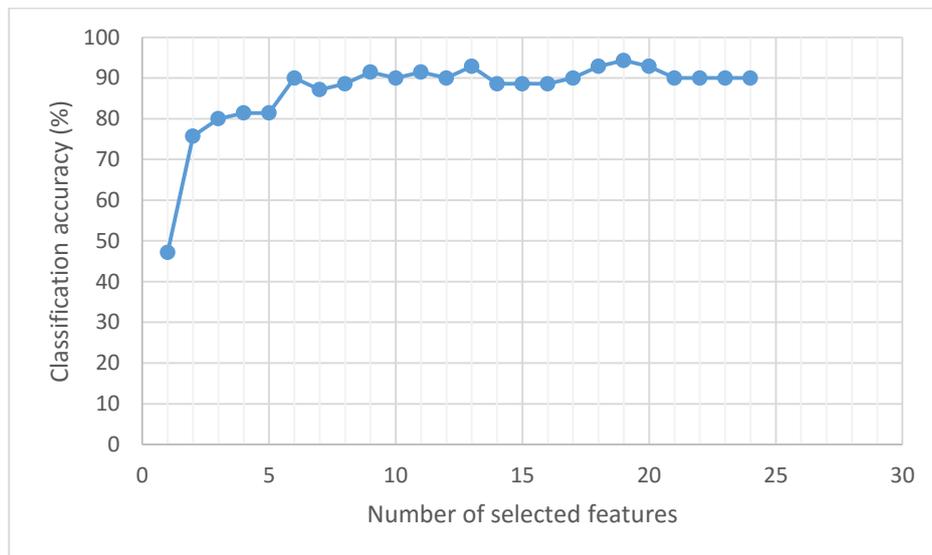
1. Variance of 6<sup>th</sup> mode.
2. Variance of 3<sup>rd</sup> mode.
3. Mean of 3<sup>rd</sup> mode.
4. Mean of 2<sup>nd</sup> mode.
5. Mean of 6<sup>th</sup> mode.
6. Standard deviation of 4<sup>th</sup> mode.
7. Standard deviation of 2<sup>nd</sup> mode.
8. Variance of 4<sup>th</sup> mode.
9. Variance of 1<sup>st</sup> mode.
10. Mean of 5<sup>th</sup> mode.
11. Mean of 1<sup>st</sup> mode.
12. Kurtosis of 6<sup>th</sup> mode.
13. Variance of 2<sup>nd</sup> mode.
14. Standard deviation of 1<sup>st</sup> mode.
15. Kurtosis of 3<sup>rd</sup> mode.
16. Kurtosis of 1<sup>st</sup> mode.
17. Kurtosis of 2<sup>nd</sup> mode.
18. Kurtosis of 4<sup>th</sup> mode.

**Table 4:** Number of Selected Features and Classification Accuracy with Naïve Bayes Classifier.

No. of selected features	Classification accuracy (%)
1	47.15
2	75.71
3	81.43
4	81.43
5	81.43
6	90.00
7	87.14
8	87.14
9	90.00
10	88.57
11	90.00
12	91.43
13	90.00
14	87.14
15	90.00
16	88.57
17	91.43
18	94.29
19	91.43
20	91.43
21	91.00
22	90.00
23	90.00
24	90.00



**Fig. 3:** Classification Accuracy vs. Number of Features selected for Classification Using Naïve Bayes.



**Fig. 4:** Classification Accuracy vs. Number of features Selected for Classification Using Bayes Net.

The remaining six features were not used in J48 classifier and hence can be used in random order. A study was conducted on the effect of number of features selected on the classification accuracy obtained with Naïve Bayes classifier and the results are given in Table 4.

The same was conducted with Bayes Net classifier and the results are given in Table 5.

It is evident from Figure 3 that on using Naïve Bayes classifier, maximum classification accuracy of 94.29% was obtained when the number of selected features is taken as 18, and on using Bayes Net classifier, highest accuracy of 94.29% was obtained when 19 features were selected (Figure 4). These 18 features were taken in the order previously mentioned and the 19<sup>th</sup> feature was chosen randomly.

**Table 5: Number of Selected Features and Classification Accuracy with Bayes Net Classifier.**

No. of selected features	Classification accuracy (%)
1	47.15
2	75.71
3	80.00
4	81.43
5	81.43
6	90.00
7	87.14
8	88.57
9	91.43
10	90.00
11	91.43
12	90.00
13	92.86
14	88.57
15	88.57
16	88.57
17	90.00
18	92.86
19	94.29
20	92.86
21	90.00
22	90.00
23	90.00
24	90.00

Classification of VMD preprocessed signals using Naïve Bayes classifier.

The statistical features extracted and selected were classified using Naïve Bayes classifier and the results are presented in Table 6. A maximum classification accuracy of 94.29% was obtained when supervised discretion was used to convert numeric attributes to nominal ones. Using a kernel estimator for numeric attributes rather than a normal distribution increased the classification accuracy from 38.57% to 42.86%. Table 7 shows the confusion matrix for Naïve Bayes classification when supervised discretion was used.

**Table 6: Classification Accuracy Obtained with Naïve Bayes Classifier.**

Use kernel estimator	Use supervised discretization	Classification accuracy (%)
False	False	38.5715
False	True	94.2857
True	False	42.8571

**Table 7: Confusion Matrix for Naïve Bayes and Bayes Net Classifiers.**

Classified actual as \	10% fault	20% fault	30% fault	40% fault	80% fault	100% fault	Good
10% fault	11	0	0	0	0	0	1
20% fault	0	11	0	0	0	0	0
30% fault	0	1	9	0	0	0	1
40% fault	0	0	0	9	0	0	0
80% fault	0	0	0	0	6	0	0
100% fault	0	0	0	0	0	8	1
Good	0	0	0	0	0	0	12

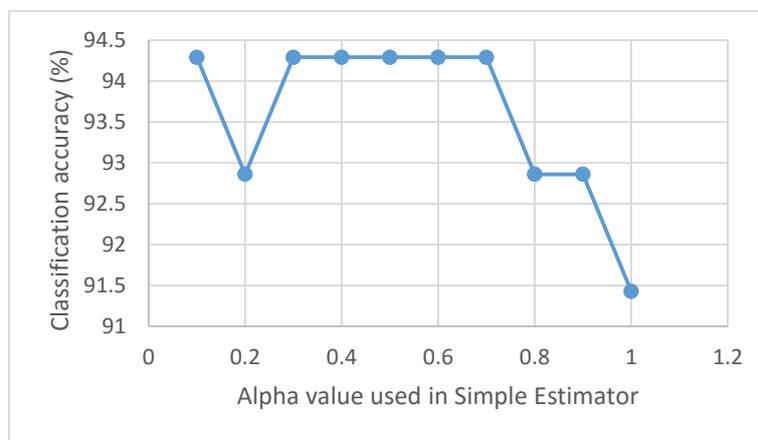
As evident from the confusion matrix in Table 4, there were four misclassifications out of the 70 test samples giving a classification accuracy of 94.29%. This is a major improvement over the 83.37% accuracy obtained with Naïve Bayes

classifier and raw signals. Thus the advantage of using VMD preprocessing is evident from this study. Classification of VMD preprocessed signals using Bayes net classifier. The statistical features extracted and selected were classified

using Bayes Net classifier after optimization of its parameters. The corresponding confusion matrix for Bayes Net classifier was observed to be same as that obtained using Naïve Bayes classifier (Table 5). A maximum classification accuracy of 94.29% was obtained. The parameters used for Bayes Net classifier are given in Table 8.

**Table 8: Parameters used for Bayes Net Classifier.**

Parameter	Set value
Estimator	Simple Estimator
Search algorithm	K2
K2 score type	Bayes
Initial network for structure learning	Bayes network



**Fig. 5: Classification Accuracy vs. Alpha Value Used in Simple Estimator.**

**Table 9: Max No. of Parents used in K2 Search Algorithm and Corresponding Classification Accuracy Achieved.**

Maximum no. of parents	Classification accuracy (%)
1	94.29
2	91.43
3	91.43
100000*	91.43

Figure 5 illustrates the variation in classification accuracy when the value of alpha used in Simple Estimator is varied from 0.1 to 1.0. Alpha is used for estimating the probability tables and can be interpreted as the initial count on each value. As evident from Figure 5 maximum classification accuracy of 94.29% was attained when the alpha was set at 0.1 or from 0.3 to 0.7.

Hence, an alpha value of 0.5 (default value) was set for this study. Table 9 shows the variation in classification accuracy for different values of maximum number of parents a node in the Bayes net can have. When initialized as Naive Bayes, setting this parameter to one result in a Naive Bayes classifier. When set to two, a Tree Augmented Bayes Network (TAN) is learned, and when set >2, a Bayes Net Augmented Bayes Network (BAN) is learned.

By setting it to a value much larger than the number of nodes in the network (the default of 100000 guarantees this), no restriction on the number of parents is enforced. It is observed that maximum accuracy is achieved by setting the number of parents to one.

**SUMMARY OF RESULTS**

*Table 10: Comparison of Classification Accuracy Attained with and Without VMD Preprocessing of Signals.*

Classifier	Classification accuracy (%)	
	Without preprocessing	With VMD preprocessing
Naïve Bayes	83.77	94.29
Bayes Net	81.77	94.29

Table 10 compares the classification accuracy obtained on classifying raw signals and signals which have undergone VMD preprocessing using Naïve Bayes and Bayes Net classifiers. It is observed that there is an improvement of up to 12.52% in classification accuracy when VMD preprocessing is performed on the same signals. This establishes the superiority of the use VMD preprocessing fault diagnosis tools for helical gearbox.

**CONCLUSION**

A new signal processing technique called VMD along with Naïve Bayes and Bayes net classifiers were presented in this paper. In order to benchmark the new VMD preprocessed features in classification, statistical features extracted from raw signals without preprocessing along with Naïve Bayes and Bayes Net classifiers were used. The superiority of VMD preprocessed features in classification was established on comparing the previously obtained classification accuracies with those attained using Naïve Bayes and Bayes Net classifiers on the statistical features extracted from VMD preprocessed signals.

From the observed significant increase in classification accuracy, it can be concluded that VMD preprocessed signals with Naïve Bayes and Bayes Net classifiers perform with greater classification accuracy in helical gearbox fault diagnosis.

**ACKNOWLEDGEMENT**

This research received no specific grant from any funding agency in the public, commercial, or Not-for-profit sectors.

**REFERENCES**

1. Asi O. Fatigue failure of a helical gear in a gearbox. *Engineering Failure Analysis*. 2006; 13(7): 1116–25p.
2. Dalpiaz G., Rivola A., Rubini R. Effectiveness and Sensitivity of Vibration Processing Techniques for Local Fault Detection in Gears. *Mechanical Systems and Signal Processing*. 2000; 14(3): 387–412p.
3. Toutountzakis T., Mba D. Observations of acoustic emission activity during gear defect diagnosis. *NDT & E International*. 2003; 36(7): 471–77p.
4. Ebersbach S., Peng Z., Kessissoglou N.J. The investigation of the condition and faults of a spur gearbox using vibration and wear debris analysis techniques. *Wear*. 2006; 260(1–2): 16–24p.
5. Rai V.K., Mohanty A.R. Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert-Huang transform. *Mechanical Systems and Signal Processing*. 2007; 21(6): 2607–15p.
6. Ricci R., Pennacchi P. Diagnostics of gear faults based on EMD and automatic selection of intrinsic mode functions. *Mechanical systems and Signal Processing*. 2011; 25: 821–38p.

7. Saravanan N., Siddabattuni V.N.S., Ramachandran KI, Fault diagnosis of spur bevel gear box using artificial neural network, and proximal support vector machine. *Applied Soft Computing*. 2010; 10(1): 344–60p.
8. Bin G.F., Gao J.J., Li X.J., et al. Early fault diagnosis of rotating machinery based on wavelet packets-Empirical mode decomposition feature extraction and neural network. *Mechanical Systems and Signal Processing*. 2012; 27: 696–711p.
9. Yang Q., An D., EMD and Wavelet Transform Based Fault Diagnosis for Wind Turbine Gear Box. *Advances in Mechanical Engineering*. 2013; 212836: 9.
10. Sugumaran V., Ramachandran K.I. Effect of number of features on classification of roller bearing faults using SVM and PSVM. *Expert Systems with Applications*. 2011; 38.
11. Amarnath M., Jain D., Sugumaran V., et al. Fault diagnosis of helical gear box using Naïve Bayes and Bayes net. *International Journal of Decision Support Systems*. 2014.