International Journal of Robotics and Automation
Vol. 1: Issue 1
www.journalspub.com

Bearing Fault Diagnosis Using Vibration Signals by Variational Mode Decomposition and Naïve Bayes Classifier

A. Muralidharan[1]*, V. Sugumaran[1], K. P. Soman[2]
[1]SMBS, VIT University, Chennai Campus, Vandalur-Kelambakkam Road, Chennai, India
[2]Centre for Excellence in Computational Engineering and Networking, Amrita Vishwa Vidyapeetham, Coimbatore (Tamil Nadu), India

Abstract

A bearing is a machine element that constrains relative motion and reduces friction between moving parts in the desired motion. A faulty bearing is a serious threat to the functionality of a machine, be it big or small. Thus, it is essential to diagnose the faults in the bearings at an early stage, so as to reduce the losses that might be incurred in money and time. Inability to meet the required demand of products in the specific time due to improper functioning of the bearing is another reason of concern. Hence, there is a necessity for continuous monitoring of the bearing. The vibrations and the sounds produced by the bearings from good and simulated faulty conditions can be effectively used to detect the faults in these bearings. The use of Variational Mode Decomposition (VMD) in the study 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. VMD finds intrinsic mode functions on central frequencies using alternating direction multiplier method (ADMM). Worthwhile 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 Naïve Bayes Classifier. The performance analysis of Naïve Bayes Classifier is elaborately discussed.

Keywords: *Bearing fault diagnosis, variational mode decomposition, alternating direction multiplier method, naïve Bayes Classifier, J48 tree algorithm*

****Author for Correspondence:*** *E-mail ID: akhil.muralidharan2011@vit.ac.in*

INTRODUCTION

Bearings are vital in our modern world. Without them, many of the machines, devices and services which are now taken for granted, would not function. It is an inherently efficient product and its use and development can contribute to solutions that save energy. Timely maintenance of bearings is essential to ensure smooth functioning of the machines. Failure to detect the fault in the bearing will lead to huge economic losses and physical damages. Hence it is essential to carry out an experimental study which provides a method for its proper monitoring and fault diagnosis. A standard rolling-element bearing consists of an inner and outer raceway with a set of balls or rolling-elements placed between these two raceways and held by a cage. [1]

The bearing faults can be caused due to improper installations of the bearing onto the shaft or into the housing, misalignment of the bearing, contamination, corrosion, improper lubrication, brinelling or simply due to wear-out. [23] Researches and developments conducted in universities and in industries have yielded means for predictive condition monitoring and fault detection algorithms.

The study focussed on the use of 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 localised faults directly when the bearings are running. A general review of monitoring and fault diagnosis techniques can be found in [4, 5]. The recent studies suggest that the signature analysis of data from acceleration or velocity transducers mounted on the bearing housing or machine casing is necessary for diagnostic techniques for rolling element bearings. [6] This results in obscured vibration signals from bearings due to the other components in the associated machinery. Hence, Kim [6] experimented with an eddy current transducer for this application, however failed to discuss how the behaviour of the outer race deflection associated with bearing frequencies for a bearing with raceway defects were described through a time-domain representation. The relationship of spike signals in outer race deflection to raceway and bearing defects has been established by Yu *et al.* [7,8]. The characteristic features of the fault-related signals generated by the rotating machine elements are often disguised by sounds and vibrations. [9, 10]

While taking fast Fourier transform of vibration signals, the harmonics 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. Sugumaran *et al.* have used Bayes classifier, [11] decision tree [12] and proximal support vector machine [13] in fault diagnosis of roller bearing which has provided significant classification accuracies.

At the insight of 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 nascent faults in bearings. [14–16] The Intrinsic Mode Functions (IMF) doesn’t work well with non-stationary signals. Lei *et al*. used EMD to extract features from signals for classifying the different modes and degrees of gear faults. [17] 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. [18] The wavelet can represent signals in time frequency plane; however, it has some limitations. [19] The present study used a new pre-processing technique developed to decompose the signal into various modes or IMF’s using calculus of variations. The modes have compact frequency support around the central frequency. Alternating direction multiplier method (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 roller bearings. To extract some meaningful features, the vibration signals were preliminarily pre-processed for finding the modes and IMFs. Then, useful statistical features like median, variance, standard deviation, kurtosis etc., were extracted. With the extracted statistical features, classification was carried out using Naive Bayes classifier.

EXPERIMENTAL SETUP AND PROCEDURE

A Machinery fault simulator is used for simulating faults in machine components such as bearings, gears, belts *etc.*and was used to study its behaviour through the information obtained from sensors as shown in Figure 1. A DC motor is mounted on the motor-support. The motor can take 0.5 HP power and its speed can be varied from 0 to 3000 rpm. The motor has a drive circuit. Its speed can be controlled from the panel. The motor shaft is connected to a set of bearings through a flexible coupling made of Aluminium. The diameter of the shaft is 20 mm. The flexible coupling reduces the transmission of motor vibration to other machine parts. The shaft is supported by two bearings, which are mounted on the bearing housings. The bearing housing is made of the split type so that it can be opened and bearings replaced easily. To prevent the shaft from mechanical damage during bearing studies, a sleeve is used between shaft and inner race of bearing. The outer diameter of sleeve is 30 mm. In between the bearing housings provision is made on the shaft for loading the bearings using dead weights.

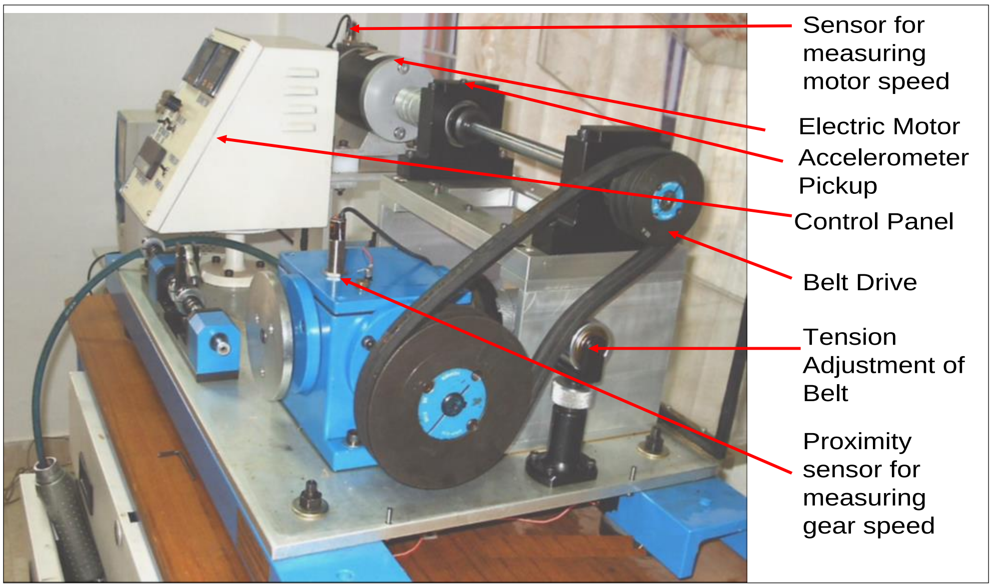

Fig.1: *The Machinery Fault Simulator.*

A small pulley is attached at the end of the shaft to transmit the rotating motion to the gearbox through belt drive. Fault simulations on belt drives can be studied using this provision. There is a belt tensioner to tighten the belt. By increasing the height of the belt tensioner, the belt can be tightened. The other side of the belt is connected to a pulley, which is in turn connected to a gearbox. The gearbox has two bevel gears perpendicular to each other. The gearbox has an electro-mechanical brake assembly to load the gear. The other side of the gear is attached to a disc, which is connected to a crank. This converts rotary motion into reciprocating motion. The crank is connected to slider shaft through slide. The slide shaft is supported by two bush housings. This mechanism is to simulate fault conditions of reciprocating mechanisms.

The control panel has a set of controls and displays as described below. There is an ON/OFF switch for DC motor. The speed can be varied by adjusting a potentiometer setting through a knob. To know the speed of the motor and speed of the gear, separate proximity pick ups (speed sensors) are fixed; both of them are connected to a single display. There is a selector to choose the speed of the motor or gear. As load increases, the current drawn by the motor increases steadily. The rate of loading can be monitored using the current. The current drawn by the motor can be read from the ammeter on the panel. There is a temperature sensor (thermocouple) to measure the temperature at the point of interest, where it is inserted. Thermocouple can be fixed at any location and the corresponding temperature can be read. There is a switch to control the brake of the gear. The whole set up is fixed on to the Aluminium base plate, which is mounted on two channels. At the bottom of the channel anti-vibration mountings are provided to reduce the transmission of vibration between setup and earth.

Fault Simulation on Bearings

Understanding why a bearing has failed is one of the best ways to prevent the same from happening again. We often come across cracks in the raceway ring and rolling elements, which is the most commonly found bearing defect. Continued use under this condition leads to larger cracks or fractures. For our experimental study, we have simulated the faults in the bearings as it is difficult to get three such defective bearings of the same type. Four cylindrical roller bearings (NU2206) were taken for study, out of which one was a brand new bearing devoid of any defects. In the other three roller bearings, defects were created using wire-cut Electric Discharge Machining (EDM), which ensured precisely defined defects.

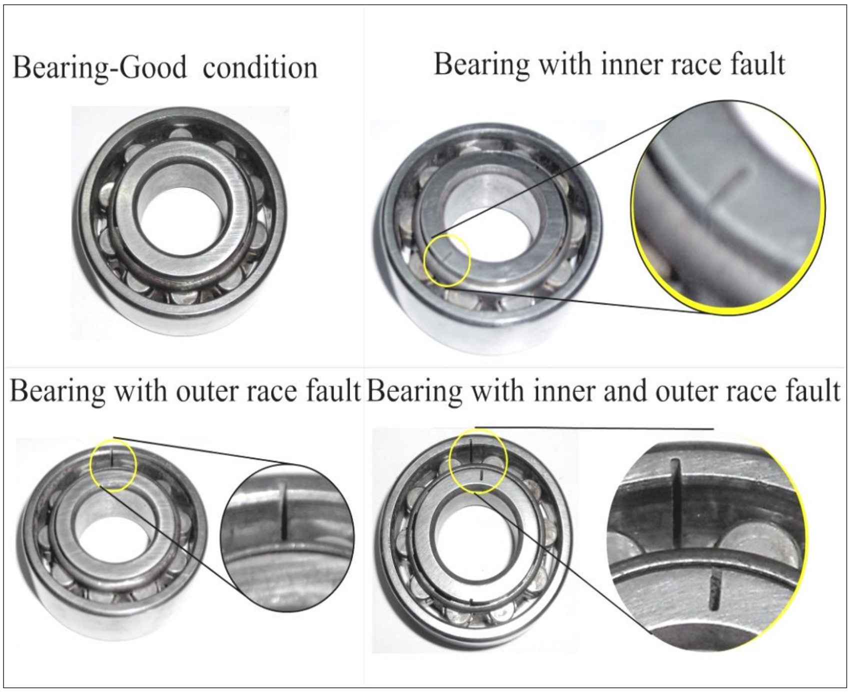

***Fig. 2:** Bearings Under Different Fault Conditions.*

The size of outer race defect is 0.0.652 mm wide and 0.981 mm deep and that of inner race defect is 0.525 mm wide and 0.827 mm deep. The images of the bearings with the simulated faults are shown in Figure 2.

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 centred 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 centred 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 $\left(u_k(t)+ju_k^H(t)\right)$. The frequency spectrum of this function is one sided (exist only for positive frequency) and assumed to be centred on ω_k. By multiplying this analytical signal with $e^{-j\omega_k t}$, the signal is frequency translated to be centred 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 $u_k(t)$.

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

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,

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

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

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:

$$L(u_k, w_k, \lambda) = \alpha \sum_k \left\| \partial_t \left[\left(\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right) e^{-j\omega_k t} \right] \right\|_2^2 + \left\| f - \sum_k u_k \right\|_2^2 + \left\langle \lambda, f - \sum_k u_k \right\rangle \qquad \text{Eq. (1)}$$

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:

<u>Update for *u* terms</u>

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

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,

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

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

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

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

Note that,

For negative ω, $\left(1 + \operatorname{sgn}(\omega) \right) \hat{u}_k(\omega) = 0$

And for positive ω, $\left(1 + \operatorname{sgn}(\omega) \right) \hat{u}_k(\omega) = 2\hat{u}_k(\omega)$

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

Therefore,

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

Replacing $\omega \to \omega + \omega_k$

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

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

$$\left\|\left(1+\operatorname{sgn}(\omega+\omega_k)\right)\hat{u}_k(\omega+\omega_k)\right\|_2^2 = \int_w \left(j(\omega-\omega_k)\left(1+\operatorname{sgn}(\omega)\right)\hat{u}_k(\omega)\right)\overline{\left(j(\omega-\omega_k)\left(1+\operatorname{sgn}(\omega)\right)\hat{u}_k(\omega)\right)} d\omega$$

$$= \int_0^\infty 4(\omega-\omega_k)^2 |\hat{u}_k(\omega)|^2 d\omega$$

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 $|\ |$ represent magnitude of the complex number.

Therefore,

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

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

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

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

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

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

Update for ω_k s;

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

$$\omega_k^{n+1} = \underset{\omega_k}{\arg\min} \left\| j\omega\left(1+\operatorname{sgn}(\omega+\omega_k)\right)\hat{u}_k(\omega+\omega_k)\right\|_2^2$$

$$\omega_k^{n+1} = \underset{\omega_k}{\arg\min} \left\| j\left(\omega - \omega_k\right)\left(1 + \operatorname{sgn}(\omega)\right)\hat{u}_k(\omega) \right\|_2^2$$

$$\omega_k^{n+1} = \underset{\omega_k}{\arg\min} \int_0^{\infty} \left(\omega - \omega_k\right)^2 |\hat{u}_k(\omega)|^2 \, d\omega$$

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

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

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

<u>Update for λ (Lamda)</u>

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

Final algorithm for VMD:

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

$$\text{repeat}$$

$$n \leftarrow n+1$$

$$\text{for } k = 1{:}K \text{ do Update } \hat{u}_k \text{ for all } \omega \geq 0$$

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

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

$$\text{end for}$$

$$\text{Dual ascent for all } \omega \geq 0\text{:}$$

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

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

Discretization of Frequency

It is first assumed that length of the mirrored signal in the time domain is one. 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 ω 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.

Feature Extraction

Descriptive statistical parameters such as mean, kurtosis, standard deviation, and variance extracted from the vibration signals are extracted from the vibrational signals 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.

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

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

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

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

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

where 's' is the sample standard deviation.

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

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

Feature Selection with 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. Thus, 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 Figure 3.

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.

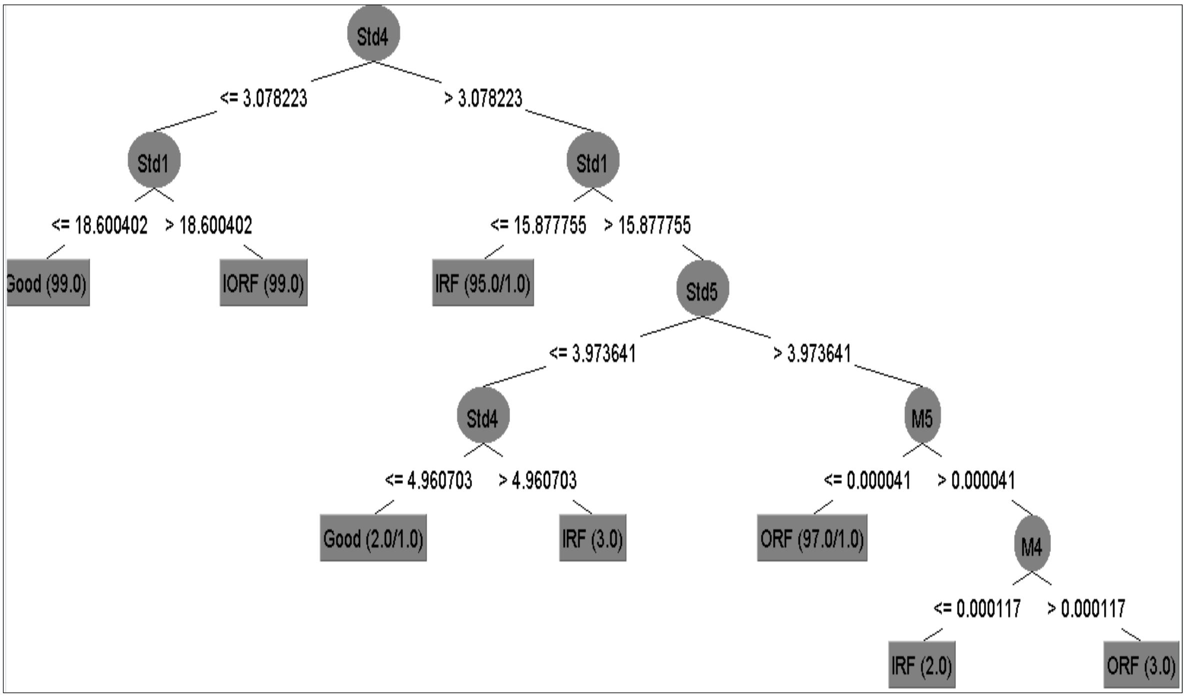

Fig. 3: *Decision Tree.*

Classifier

Classification is the problem of identifying to which of a set of categories a new observation belongs, on the basis of a training set of data containing observations whose category membership is known.

An algorithm that implements classification, especially in a concrete implementation, is known as a classifier. Naïve Bayes Classifier has been used in the present study. The Naive Bayesian classifier is based on Bayes' theorem with independence assumptions between predictors.

A Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it particularly useful for very large datasets. Despite its simplicity, the Naive Bayesian classifier often does surprisingly well and is widely used because it often outperforms more sophisticated classification methods.

The Naïve Bayes Classifier is based on Bayes rule that assumes that the attributes $A_1, A_2, A_3, \ldots\ldots A_n$, are all conditionally independent of one another given an attribute B. This assumption is very beneficial as it simplifies the representation of $P(A \mid B)$ and reduces the problem of estimating it from the training data. When A contains n attributes which are conditionally independent of one another given B, we have

$$P(A_1 \ldots \ldots . A_n \mid B) = \prod_{i=1}^{n} P(A_i \mid B)$$

When A_i and B are Boolean variables, only $2n$ parameters are needed to define $P\left(A_1 = a_{ik} \mid B = b_j\right)$ with corresponding i, j, k values. This is a substantial reduction as compared to the $2(2^n - 1)$ parameters needed to characterize $P(A \mid B)$, assuming there is no conditional independence assumption. The main focus is to train a classifier that will output the probability distribution over possible values of B, for each new instance A that needs to be classified. The expression for probability that B will take on its K^{th} possible value according to Bayes rule is

$$P(B = b_k \mid A_1 \ldots\ldots A_n) = \frac{P\big(B = b_y\big)P(A_1 \ldots\ldots\ldots A_n \mid B = b_k)}{\sum_j P\big(B = b_j\big)P\big(A_1 \ldots\ldots A_n \mid B = b_j\big)}$$

where the sum is taken over all possible values of b_j of B. Now assuming that A_i is conditionally independent given B, we can rewrite the above equation as

$$P(B = b_k \mid A_1 \ldots\ldots A_n) = \frac{P\big(B = b_y\big)\prod_i P(A_1 \mid B = b_k)}{\sum_j P\big(B = b_j\big)\prod_i P(A_1 \mid B = b_k)}$$

This is the fundamental equation for the Naïve Bayes Classifier.

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. The corresponding output together forms the dataset.

Effect of Number of Features on Classification Accuracy

As discussed earlier, out of all the statistical features extracted from the vibration signals, it is not certain that all the features contribute equally to the classification accuracy. The process of reducing the number of input features for classification is known as dimensionality reduction. Table 1 and Figure 4 illustrate the variation of classification accuracy with change in the number of features.

Table 1: *Classification Accuracy of Naïve Bayes Classifier*.

Number of Features	Classification Accuracy(in %)
1	83.50
2	97.00
3	99.25
4	99.25
5	**99.25**
6	99.00
7	92.25
8	95.50
9	97.25
10	95.00
11	94.00
12	90.50
13	89.50
14	95.25
15	96.25

It is to be noted that the maximum classification accuracy is obtained with only five features being used instead of the total 15 features.

Statistical Features with Naïve Bayes Classifier

Hemantha Kumar *et al.*[11] recorded vibrational signal samples and used it for generating the decision tree for the purpose of feature selection. The decision tree generated is shown in Figure 4. The class wise accuracy generated by this study is illustrated in Table 2. The results indicate that it generates a classification accuracy of 85% only.

Table 2: *Class Wise Accuracy of Naïve Bayes Classifier.*

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.833	0.067	0.862	0.833	0.847	0.968	Healthy
	0.867	0.083	0.839	0.867	0.852	0.968	IRF
	1	0	1	1	1	1	ORF
Weighted Avg.	0.9	0.05	0.9	0.9	0.9	0.979	

Note: IRF: inner race fault; ORF: outer race fault

Variational Mode Decomposition with Naïve Bayes Classifier

The descriptive statistical features after pre-processing were used with Naïve Bayes Classifier and the results are discussed here. The parameters considered for the classifier are shown in Table 3. These parameters play a vital role as altering these parameters can result in a significant change in the classification accuracy.

Table 3: *Classifier Parameters for Naïve Bayes Classifier.*

Parameters for Evaluation	**Values**
Model performance evaluation	10-fold stratified cross validation
Model building time	0.15 seconds
Display model in old format	True
Estimator	Kernel
Use supervised discretization	False

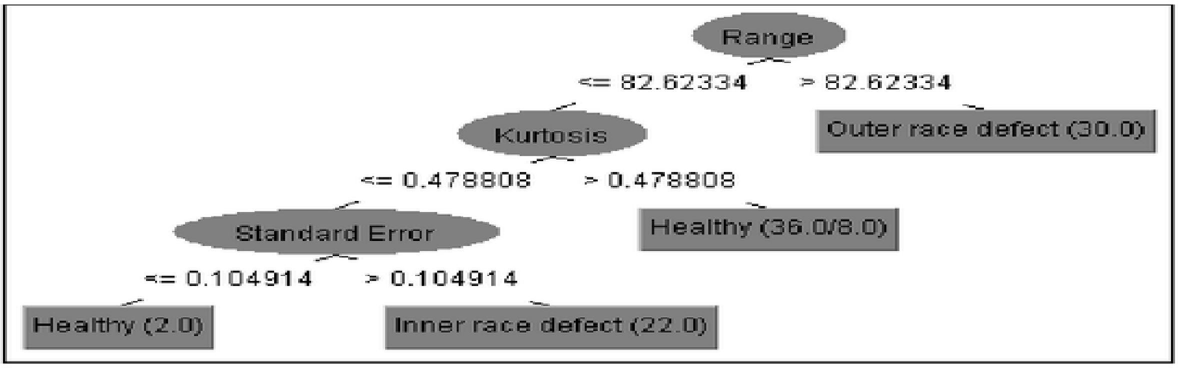

Fig. 4: *Decision tree.*

Table 4: *Improvement in Classification Accuracy on Using VMD Pre-Processing.*

Sl No.	Classification Algorithm	Classification Accuracy (%)		Improvement (%)
		Without VMD Pre-Processing	With VMD Pre-Processing	
1	Naïve Bayes	85	99.25	14.25

The classification accuracy of the Naïve Bayes classifier using these optimised parameters was found out to be 99.25%. The details of misclassification are presented in the form of a confusion matrix in Table 5.

Table 5: *Confusion Matrix for Naïve Bayes Classifier.*

Testing	Good	IRF	ORF	IORF	Normal
Good	99	0	1	0	0
IRF	0	99	1	0	0
ORF	0	0	100	0	0
IORF	0	0	1	99	0
Normal	0	0	0	0	100

Note: *Good: signals from the bearing which is flawless; IORF: inner and outer race fault.*

Summarizing the observations from the Confusion matrix, one can conclude that the classifier is performing efficiently in fault identification (100%) ensuring that a false diagnosis will not take place. The diagonal of the confusion matrix represents the correctly classified signals and other elements are the misclassified ones. In condition monitoring, the priority is fault identification and fault classification comes second. Referring to Table 5, it can be seen that misclassifications between the faulty and the good signals are almost zero. However there are some misclassifications among the faulty signals. Thus the classifier's performance is highly encouraging seeing 99.25% classification accuracy.

CONCLUSION

Faults in the bearings possess a major threat to the machinery in terms of its performance and can also lead to economic loss as well as the physical damages. Thus it is essential for the continuous monitoring of the bearing. From this study, it is apparent that Machine learning is a simple but powerful tool for fault diagnosis. The introduction of VMD as a new signal pre-processing technique along with Naïve Bayes Classifier has provided remarkable performance characteristics with a classification accuracy reaching 99.25%. For bench marking the new features and classier, statistical features extracted from raw signal (without VMD pre-processing) and Naïve Bayes Classifier respectively 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 (85%). From the results and discussions, we can conclude that VMD pre-processed signals with Naïve Bayes Classifier perform flawlessly in fault diagnosis of bearings. Its ability to distinguish between good and faulty signals with 100% accuracy motivates its use in the industry.

REFERENCES

1. Changsen W. *Analysis of rolling element bearings*. London: Mechanical Engineering Pub. Limited; 1991.
2. Benbouzid M, Kliman GB. What stator current processing-based technique to use for induction motor rotor faults diagnosis? *IEEE Trans. on Energy Conversion*. 2003; 18(2): 238–44p.
3. Li B, Chow MY, Tipsuwan Y, *et al.* Neural-network-based motor rolling bearing fault diagnosis. *IEEE Trans. on Industrial Electronics*. 2000; 47(5): 1060–69p.
4. Nandi S, Toliyat HA. Condition monitoring and fault diagnosis of electrical machine, *Proceedings of IEEE IAS Annual Meeting Conference*. 1999; 197–204p.
5. Mathew J, Alfredson RJ. The Condition Monitoring of Rolling Element Bearings using Vibration

Analysis. *Journal of Vibration, Acoustics, Stress and Reliability in Design*.1984; 106:447–53p.
6. Kim PY. A Review of Rolling Element Bearing Monitoring (III): Preliminary Test Results on Eddy Current Transducer Technique. *Proceedings of Third International Conference on Vibrations in Rotating Machinery*. The Institution of Mechanical Engineers, York, England; 1984; 119–25p.
7. Yu JJ, Bently DE, Goldman P, *et al*. Rolling Element Bearing Defect Detection and Diagnostics Using Displacement Transducers. *ASME*; Paper 2001-GT-28.
8. Detection of Rolling Element Bearing Defects in Time Base. *Bently Rotor Dynamics Research Corporation*. 2000.
9. Zvokelj M, Zupan S, Prebil I. 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*. 2010; 24: 1049–67p.
10. Li CJ, Ma J. Wavelet decomposition of vibrations for detection of bearing localized defects. *NDT&E International*. 1997; 30 (3): 143–9p.
11. Kumar H, Kumar TAR, Sugumaran V, *et al*. Fault Diagnosis of Bearings through Vibration Signal Using Bayes Classifiers. *International Journal of Computer Aided Engineering and Technology*. 2012.
12. Amarnath HM, Sugumaran V. Exploiting Sound Signals for Fault Diagnosis of Bearings Using Decision Tree, Measurements. 2013; 46(3): 1250–56p.
13. Sugumaran V, Muralidharan V, Ramachandran KI. Feature selection using Decision Tree and classification through Proximal Support Vector Machine for fault diagnostics of roller bearing. *Mechanical Systems and Signal Processing*. 2007; 21(2): 930–42p.
14. Loutridis SJ. Damage detection in gear systems using Empirical mode decomposition *Engineering Structures*. 2004; 26(12): 1833–41p.
15. Yu D, Cheng J, Yang Y. Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings. *Mechanical systems and signal processing*. 2005; 19(2): 259–70p.
16. Gao Q, Duan C, Fan H, *et al*. Rotating machine fault diagnosis using Empirical mode decomposition. *Mechanical systems and signal processing*.2008; 22(5): 1072–81p.
17. Lei YG, Zuo MJ, He ZJ, *et al*. A multidimensional hybrid intelligent method for gear fault diagnosis. *Expert systems with applications*. 2010; 37:1419–30p.
18. Rilling G, Flandrinand P, Goncalves P. on Emperical Mode Decomposition and its Algorithms. *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP-03*; 2003. Grado (I)
19. Huang N, Shen Z, Long SR, *et al*. The empirical mode decomposition and Hilbert spectrum for non-linear non-stationary time series analysis. *Proceedings of Royal Society of London*; 1998; A 454: 903–95p.