# FAULT DIAGNOSIS OF BEVEL GEARBOX USING ARTIFICIAL INTELLIGENCE TECHNIQUE

**M.Tech. Thesis**

By

**Shivam Kumar**

1802103022

Discipline of Mechanical Engineering

**Indian Institute of Technology Indore**

June 2020

# FAULT DIAGNOSIS OF BEVEL GEARBOX USING ARTIFICIAL INTELLIGENCE TECHNIQUE

## A THESIS

*Submitted in partial fulfillment of the requirement for the award of the degree*

*Of*

**Master of Technology**

**In**

**Mechanical Engineering**

*with specialization in*

**Mechanical Systems Design**

*By*

**Shivam Kumar**

Discipline of Mechanical Engineering

**Indian Institute of Technology Indore**

June 2020

# INDIAN INSTITUTE OF TECHNOLOGY INDORE

## CANDIDATE DECLERATION

I hereby certify that the work which has been presented in the thesis entitled **FAULT DIAGNOSIS OF BEVEL GEARBOX USING ARTIFICIAL INTELLIGENCE TECHNIQUE** in the partial fulfillment of the requirements for the award of the degree of **MASTER OF TECHNOLOGY** and submitted in the **DISCIPLINE OF MECHANICAL ENGINEERING, Indian Institute of Technology Indore**, is the authentic record of my own work carried out during the time period from ***May 2019*** to ***June 2020***. Thesis submission under the guidance of Dr. Anand Parey, Professor, Dept. of Mechanical Engineering, IIT Indore.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

**(Shivam Kumar)**

---

This is to certify that the above statement made by the candidate is correct to the best of my/our knowledge.

23/06/2020

**Signature of the supervisor of M.Tech. thesis**

**Prof. ANAND PAREY**

**Professor, IIT Indore**

---

**Shivam Kumar** has successfully given his M.Tech. Oral Examination held on 23-06-2020

23/06/2020

Signature of Supervisor
of M.Tech. thesis
Date: 23/06/2020

Convener, DPGC

23/06/2020 23/06/2020

**Signature of PSPC Member**
**Dr. PAVAN KUMAR KANKAR**
**Associate professor, IIT Indore**

**Signature of PSPC member**
**Dr. SANTOSH KUMAR SAHU**
**Associate professor, IIT Indore**

---

Signature of the Chairman, Oral Examination Board with Date

---

# ACKNOWLEDGEMENT

The successful result and the outcomes of this thesis i.e., FAULT DIAGNOSIS OF BEVEL GEARBOX USING ARTIFICIAL INTELLIGENCE TECHNIQUE required a lot of pilotage and headship and I am vastly fortunate to have acquired this from my thesis supervisor. I would like to denote my felicitations to Professor ANAND PAREY to help me and direct me at every step of this project and for the useful explanations, notions and contributions throughout the learning process of this master thesis. I would also like to pay tribute to him for the supervision to rectify me at wrongly taken steps.

My gratitude is also extended towards my PSPC member associate professor Dr. PAVAN KUMAR KANKAR for his guidance and cooperation. I would like to manifest my solemn thanks to my senior Mr. Pavan Gupta for his moral support and his work. I express my deep sense of gratitude to PhD scholar Mr. Anupam Kumar for his help and suggestions. The doorway to Prof. ANAND PAREY cabin was always commodious whenever I face a misery spot or had a query about my research or writing. He sequentially allowed this paper to be my work but vectored me in the right direction while he thought I needed it. I am also obliged to the PSPC member for his unbounded support and patronage.

Thanks for your encouragement.

Shivam

**Shivam Kumar**

# ABSTRACT

The project consists of the diagnosis of bevel gear faults as fault diagnosis is the important parameter for the rotating machinery. Hence the early detection of the fault will result in avoiding the serious damage to the machinery or any harm to the operator working on it. As gearbox is a crucial component of any machinery the regular monitoring of its condition is required. The study in this thesis is focused on prediction of faults using artificial intelligence technique. The vibrational responses are obtained and analyzed for various defaults possible to occur in bevel gear inside the gear box. The bevel gear with defects like one chipped tooth and one missing tooth are used for the experiment. The various techniques used for fault assessment considering vibrational data explications includes artificial neural network (ANN), deep neural network (DNN), support vector machine (SVM) and random forest (RF). For the reduction of dimensionality in the original vibration signal, the classifiers are clubbed with first attribute evaluators alone and then with wavelet transform. The results are compared for having the combination giving us the best results. This study is the combination of two stages. In the former stage, the vibrational signals are optimized using three attribute evaluators one by one and further the data is given to classifier as input and finally the best accuracy is evaluated. In the second stage, the wavelet transform is used to decompose the signals and again other attribute evaluator is used with classifier and to achieve the best accuracy. The result also shows that this algorithm can be effectively used for other components subjected to vibrations during its functionality, mainly for rotating components.

## TABLE OF CONTENTS

# List of Figures

# List of Tables

## Nomenclature:

j: Input node

i: Output node

Xj: Input variable to ANN

Yi: output given by ANN

N: Sample Data points

Ei: error in the output

Di: Actual output corresponding to Xj

Fr: Radial Force

Fx: Axial force

Ft: Tangential force

Fn: Normal force

Wij: Initial weights

Vi: Weighted sum

H: Hyperplane

I(S): Information

E(A): Entropy

## Acronyms

AI: Artificial Intelligence

ML: Machine Learning

ANN: Artificial Neural Network

DNN: Deep Neural Network

SVM: Support Vector Machine

RF: Random Forest

IFD: Intelligent Fault Diagnosis

SLP: Single Layered Perceptron

MLP: Multi Layered Perceptron

DAQ: Data Acquisition System

MODWT: Maximal Overlap Discrete Wavelet Transform

EMD: Empirical Mode Decomposition

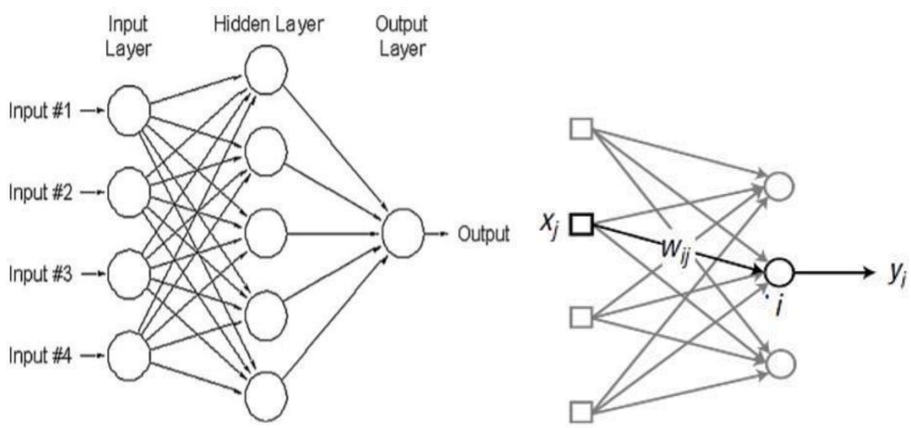
# Chapter 1: Introduction

Gearboxes are the major component in any machinery and also a critical component. It is used in industrial machinery, automotive and aerospace. As it is an unduly prerequisite component, the status and health retrospect of the gearboxes is essential to avoid the damage to the machinery or the operator working on that machine. Condition retrospection is the crucial part of predictive maintenance i.e. predicting the failure before its actual occurrence leads to save the other components from any harm. In condition monitoring we measure a parameter like temperature, pressure, noise and vibration to properly analyses the health level of machine. Here the parameter is vibration because the vibratory signal coveys the information regarding condition of the machine and hence one should analyze the signals in order to understand or judge the actual condition of a machine at a particular instance. Also, the failure of these components leads to the downtime in machine and the reduction in production which certainly results in damage in all aspects. Finding the early fault symptoms are important to reduce the losses. This can be done with various methods. Here we are putting our hands in artificial intelligence (AI). Various methods are there for intelligent fault detection (IFD).

Artificial intelligence also referred as machine/computer intelligence is the imitation of situation or process human brain functions in machines that are programmed to think like human or to mimic the action of human. The machines comprised of artificial intelligence have the ability of learning as well as problem solving. The IFD uses the application of machine learning (ML) theories such as artificial neural network (ANN), Deep neural network (DNN), support vector machine (SVM) etc. [L. Duan, M. Xie, J. Wang, T. Bai (2018), R. Liu, B. Yang, E. Zio, X. Chen (2018)]. These applications have potential to get a rapport between the practically acquired vibrational signal and the status of the component. In the late years, many researchers have drawn their attention in developing IFD through ML algorithms to get the systems having ease in predicting the fault accurately. [S. Khan, T. Yairi (2018), R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R.X. Gao (2018)].

## 1.1 ANN and back propagation algorithm.

ANN is principally like the network of nerve cell of a biological (human or animal) nervous system. ANN is considered to be the mostly used algorithm due to its popularity. A simple ANN structure formulated of three layers: input layer, hidden layer and output layer. Hidden layer as the name pushes that the value of the hidden layer nodes is not naked to the observer. ANN is simply based on these layers and the neurons comprised by these layers [Ruonan Liu, Boyuan Yang, Enrico Zio, Xuefeng Chen (2018)]. The two types of ANN are forward neural network and recurrent neural network, classified on the basis of the direction of information travel. The basic appearance of the structure of ANN is shown in the figure 1 taken from reference [N. Saravanan, V.N.S. Kumar Siddabattuni, K.I. Ramachandran (2010)]




**(a)** (b)

Figure 1: (a) ANN architecture and (b) nomenclature for the calculation of error.

## The back-propagation algorithm:

In the figure 1 (b), j-denotes the input node, i-denotes the output node. $X_j$ is the parameter which equals input given to the corresponding neuron, $W_{ij}$ is initial weight given according to the back-propagation algorithm, $Y_i$ is the output given by $i^{th}$ neuron. Let us suppose the actual output corresponding to input $X_j$ is $D_i$. The node having no input has an output level $Y_i$ where:

The activation function $\emptyset(V_i)$ is given by:

$$\emptyset(V_i) = \frac{1}{1 - e^{-V_i}} \quad [1]$$

Where $V_i$ is the weighted sum and is formulated as:

$$Vi = \sum W_{ij} X_j \quad [2]$$

The error in the output is the odds between actual value of output and the output given by the network. i.e. error $E_i$ can be given as:

$$E_i = D_i - Y_i \quad [3]$$

Now to increase the classification accuracy by reducing the error and this can be achieved by using weight adjustment, the new weight can be calculated using the following formula:

$$W_{ij\,new} = W_{ij\,old} + \alpha E_i X_j \quad [4]$$

The parameter $\alpha$ is called the learning parameter or learning rate. If the value of alpha is high, output wanders around expected solution and if it is too low then the output doesn't converge to the expected solution. The mathematical systematic way of modifying the weight is called learning rule and for single layered network is called delta rule.

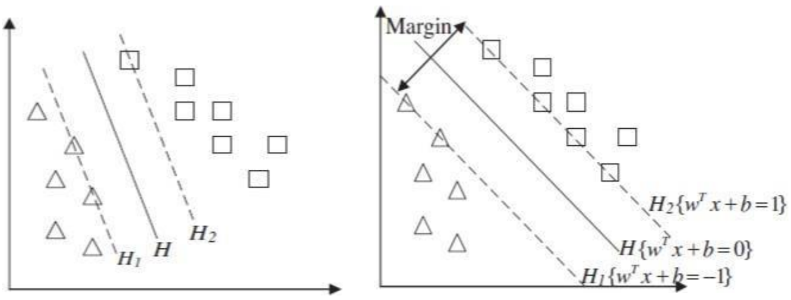
## 1.2 DNN – Deep/Dense neural network

DNN stands for deep neural network and it is similar to ANN with the only difference that the ANN consists of single hidden layer whereas DNN consists of more than one hidden layer in it. In the other way we can say that DNN is a multi- layered perceptron (MLP). ANN has been stationary from a decade. The reason behind is the consumption of time and intense computation for training of model. Multi layered perceptron reckoning of more than one number of hidden layer are used over ANN because of its good command over algorithms for optimization of network. [Manisha, Sanjeev kr. Dhull, Krishna Kant Singh (2020)].

The deep neural network has given a improved accuracy in terms of classification of the gear faults. However, the best accuracy can be achieved by varying the DNN parameters like the activation function, learning rate, number of hidden layer and the number of nodes in the hidden layer. This is called the optimization if the network [Xue Chen, Lanyong Zhang, Tong Liu a, M.M. Kamruzzaman (2019)]. Now a days DNN is expanding its ability greater than single layered perceptron (SLP). The deep learning has various applications like image recognition, speech recognition, natural language analysis etc. However, DNN has a limitation to the practical applications like which require large scale problem solving because of its high energy consumptions or high computational power. [Jaehyun Kima, Heesu Kima, Subin Huha, Jinho Lee, Kiyoung Choi, 2018].

## 1.3 Support vector machine (SVM)

Support vector machine are the machine learning models which are utilized for the classification of the different operation conditions and different health condition data with the associated algorithms and regression analysis. Feature rankings is a method of elimination of the feature giving irrelevant information. The feature having more information or having less entropy is ranked higher than other features. [Mayuri Wadkar, Fabio Di Troia, Mark Stamp, 2019]. A case which a simply a two-class problem can be considered which is shown in triangle and squares in the figure 2 taken from [P.K. Kankar, Satish C. Sharma, S.P. Harsha, 2011] in which there is a hyper plane H which is dividing or classifying the instances.




**(a)** (b)

Figure 2: -hyper plane classifying (a) with lesser margin (b) with maximum margin.

$H_1$ *and* $H_2$ are the two other planes parallel to the hyperplane H and passing through the nearest instances. The perpendicular distance between these two plans is called the margin. The SVM splits the data giving a boundary among the number of cases/classes and the margin should be taken in such a way so that it would be maximizedit leads to the negligible error in classifying the data. The instances from which the planes $H_1$ *and* $H_2$ are passed is called support vectors. [P.K. Kankar, Satish C. Sharma, S.P. Harsha, 2011].

## 1.4 Random forest classifier

As the name suggests a random forest, the general forest is what having a combination of different trees. Similarly, in a random order from a bootstrap data, the trees ate formed using a set of attributes. The random word here interprets that the chances of getting samples is equal for each tree, i.e. all the formed trees have a uniform distribution. Random forest is considered to be the most efficient technique in classifying and regression in data mining [Ch. Ravi Sekhar1, Minal 2, and E. Madhu, 2014]. Random trees are formed with the variable combination or attributes and classes and all are processed at the same time. The ultimate decision is conserved on the basis of number of similar class votes by trees. It is based on decision trees i.e. the n number of trees are formed and the decision given by each tree is considered. The final decision is dependent on the number of decision/votes given by the individual tree in the random forest. Let there be n number of decision trees and for those there will be n iteration, the steps in each iteration is: [Ch. Ravi Sekhar1, Minal 2, and E. Madhu, 2014].

- *Sample Data selection*: the model is trained on the basis of a sampled data by having a bootstrap data. The tree and the branches are filled by attributes in a random order.
- *Tree growth*: the growth of tree should be using splitting rules.
- *Attribute selection*: the features or attributes are randomly selected for each node. Repeating the feature for a single tree is not allowed in random forest.

- *Pruning:* the pruning is done for optimization and the pruned tree can be saved for other vibrational problem.
- *Result output*: the input is given to each tree in the formed forest and every single tree gives the output of classification which is considered as the votes for each class. The forest finally decides the class getting the higher number of votes.

## 1.5 Forces acting on bevel gears in meshing

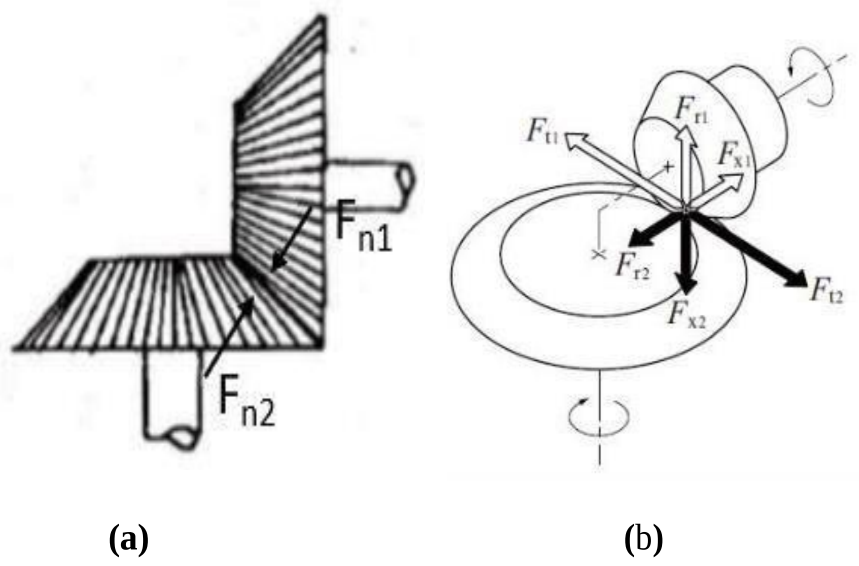



Figure 3: Representing (a) normal forces (b) radial and axial acting on bevel gears.

Figure 3 is showing the bevel gears meshed with each other at an angle of 90 degrees, force $F_{x1}$ is the axial force at the pinion, $F_{r1}$ is the radial force on the pinion and $F_{t1}$ is showing the tangential force by pinion on gear, $F_{n1}$ is the normal force which is one of the component of force acting along LOA.

Similarly, force $F_{x1}$ is the axial force at the gear, $F_{r1}$ is the radial force on the pinion and $F_{t1}$ is showing the tangential force by pinion on gear. $F_{n2}$ is the normal force which is one of the components of force acting along LOA

# Chapter 2: Review of the past work for problem formulation

## 2.1 Review of the past work

The researches have given their effort a lot on this fault diagnosis analysis. Different types of the artificial intelligence (AI) techniques and the combination can be applied to these vibration signals to classify the fault. The vibration signals are then analyzed and features are extracted from these signals which are further fed to the classifiers for training and predicting the class for new data.

The work has been done on the optimization of neural networks the constructive and the pruning method has been applied to the network in which the initial networks is made and on each training iteration where either the number of layers or the discrete number of nodes in it is a variable number. This variable number is either are increased (constructive) or decreased (pruned) for obtaining the best optimal structure of neural network [Zhiqiang Tong, Gouhei Tanaka, 2015]. Some of the researchers used the signal feature as input to different AI classifiers and compared the result depending on classification accuracy. The class prediction accuracy, if greater than the classifier is considered to be a good classifier as compared to the other. [N. Saravanan, V. N. S. Kumar Siddabattuni, K. I. Ramachandran, 2010]. extracted features from the vibrational signal (time domain signals) and fed them to ANN and SVM and the results have been compared for both classifiers [P. K. Kankar, Satish C Sharma, S.P. Harsha, 2011]. Other researches worked on fault diagnosis taking acoustic and psychoacoustic signals and features are extracted and given as an input to ANN and DNN and the results were compared and conclusion was made that psychoacoustic features were giving best accuracy and analysis on the basis of psychoacoustic features will be better as compared to vibration features. [Kane P.V., Andhare, A.B. , 2019].

Another interest has been invested on modifying the network parameter automatically. The parameter can be momentum learning rate or learning parameter. The adaptive learning algorithm is used in which the learning parameter modifies itself according to previous iteration [T. Kathirvalava kumar, S. Jeyaseeli Subavathi, 2009]. In the literature review its has been noted that the overall classification accuracy can be improved by modifying feature extraction technique i.e. algorithms like genetic algorithm [Mariela Cerrada, Grover Zurita, Diego Cabrera, René-Vinicio

Sánchez, Mariano Artés, ChanLi, 2016]. It has been adduced that genetic algorithm in annexation with another intelligent method to renovate the overall execution of fault diagnosis. A hybrid GA-BP ANN was proposed in which the feature extraction was done by discrete wavelet transform [Tyagi, S., Panigrahi, S.K., 2017]. Another way of improving accuracy is using feature selection. Feature selection is the process in which the input variables are reduced and only focused features/variables are selected which are affecting the classification output of a predictive model. To optimize text classification error minimal size of relevant features are considered. The feature selection techniques are classified as wrappers, filters, embebbed method or attribute evaluator, etc. [Gang Kou, Pei Yang, Yi Peng, Feng Xiao, Fawaz E. Alsaadi, 2020]

## 2.2 Problem formulation

Till now the work has been done on either improving the classifier or comparing of classification results of one classifier with other but there are no efforts have been put in analyzing that which feature given as input to the classifiers will be having more impact in classification. The classification of the fault in gear has been done by classifiers by giving statistical features extracted from the time domain vibration signals as input to those classifiers. The accuracy of the classification is known to be strongly dependent on the input we are giving to it. So, to give the input having less redundancy is important factor for this.

To achieve this, we can use dimensionality reduction so that we can eliminate the redundancy as well as reduce the computation power. For the above task the use of attribute evaluation is there in this paper. So, by using the evaluation technique and checking the accuracy of the model we can conclude the effect of attribute evaluation. Here the aim is to get information accurately related to the type of fault occurring in the gear. For this the classification accuracy of the fault should be as large as possible. In this paper we are going to see the time domain signal features like kurtosis, crest factor, range, etc. as input to the AI classifiers i.e. Support Vector Machine (SVM) Artificial Neural Network (ANN), Random Forest (RF)and Deep Neural Network (DNN) method. It is also important to give the correct and precise input to these classifiers to reduce the overall loss in terms of non-useful input information so, the ranker technique is used to rank the feature according to the useful information provided by them.

## 2.3 Objective to achieve

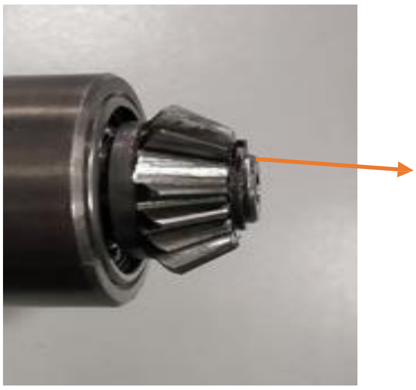
The objective here is to check the model accuracy for ANN, DNN, SVM and random forest techniques of classification and comparison of the results before and after application of attribute ranker evaluation to check that which model is giving the best accuracy and with which ranker for making a conclusion.

For fault in gear we have two conditions that is gear with:

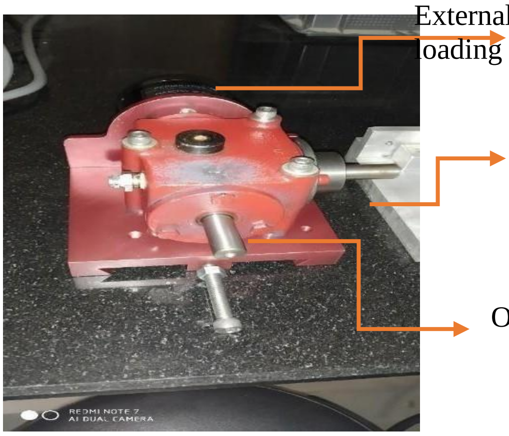
1. Chipped tooth 2. Missing tooth

**(a)**


**(b)**




**(b)**




**(d)**

Figure 4: (a) healthy gear, (b) chipped tooth gear, (c) missing tooth gear and (d) bevel gear box.

# Chapter 3: Experimental details and setup

## 3.1 Experimental details:

The figure 4 is representing that how many conditions are there in the experiment and how many experiments have been done under same conditions. There are 4 loading conditions which are no load/no torque, torque of 1 Nm, torque of 2 Nm and torque of 3 Nm and three operating frequencies i.e. 5, 10, 15 hertz frequency. There is one healthy gear and two faulty gears taken for the experiment. As mentioned before the faults given are chipped tooth and the missing tooth. Four classifiers have been used for classification namely ANN, DNN, SVM and RF. For a signal condition 5 reading have been taken which results in total 180 number of samples.

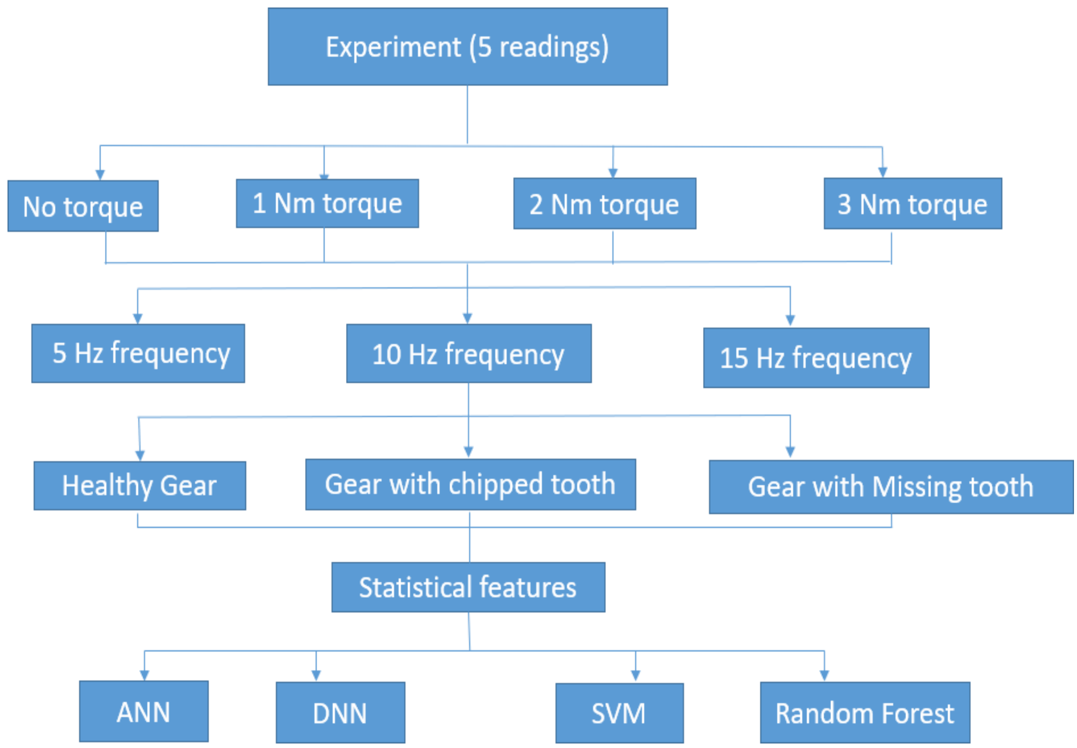
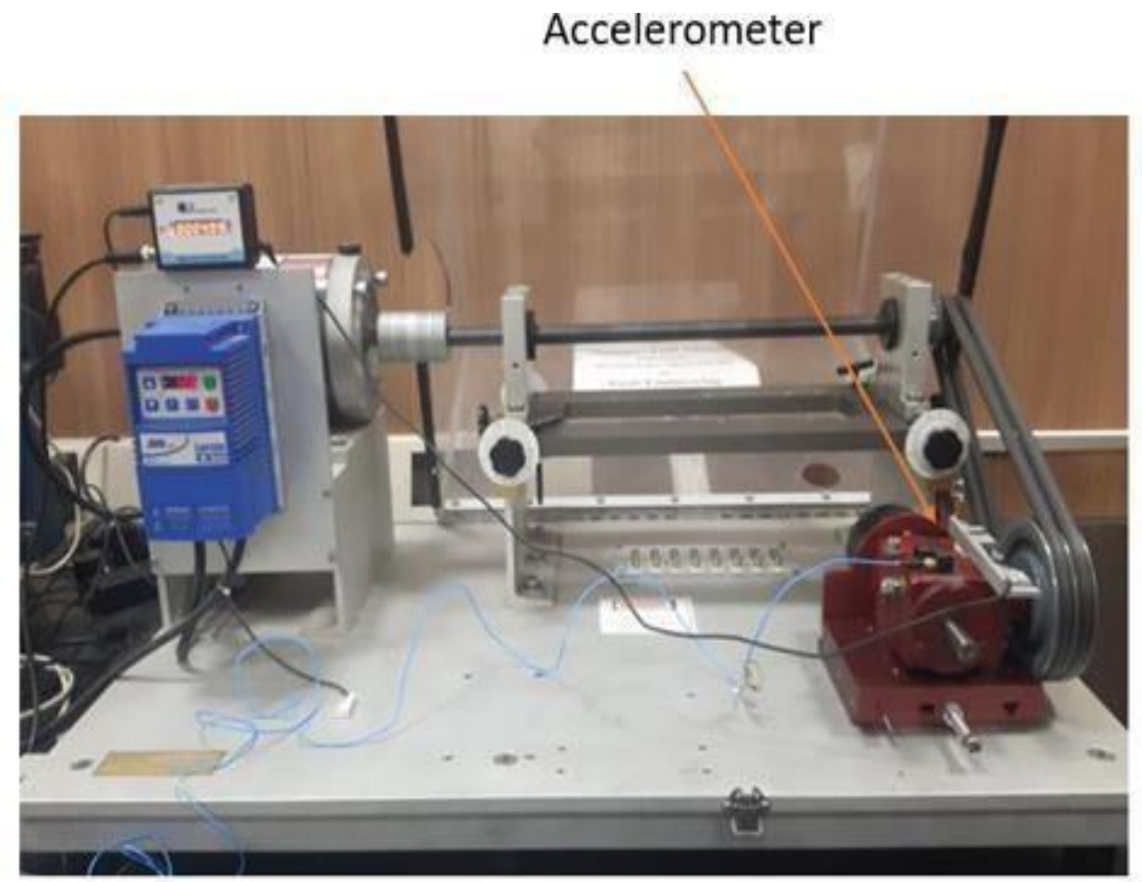


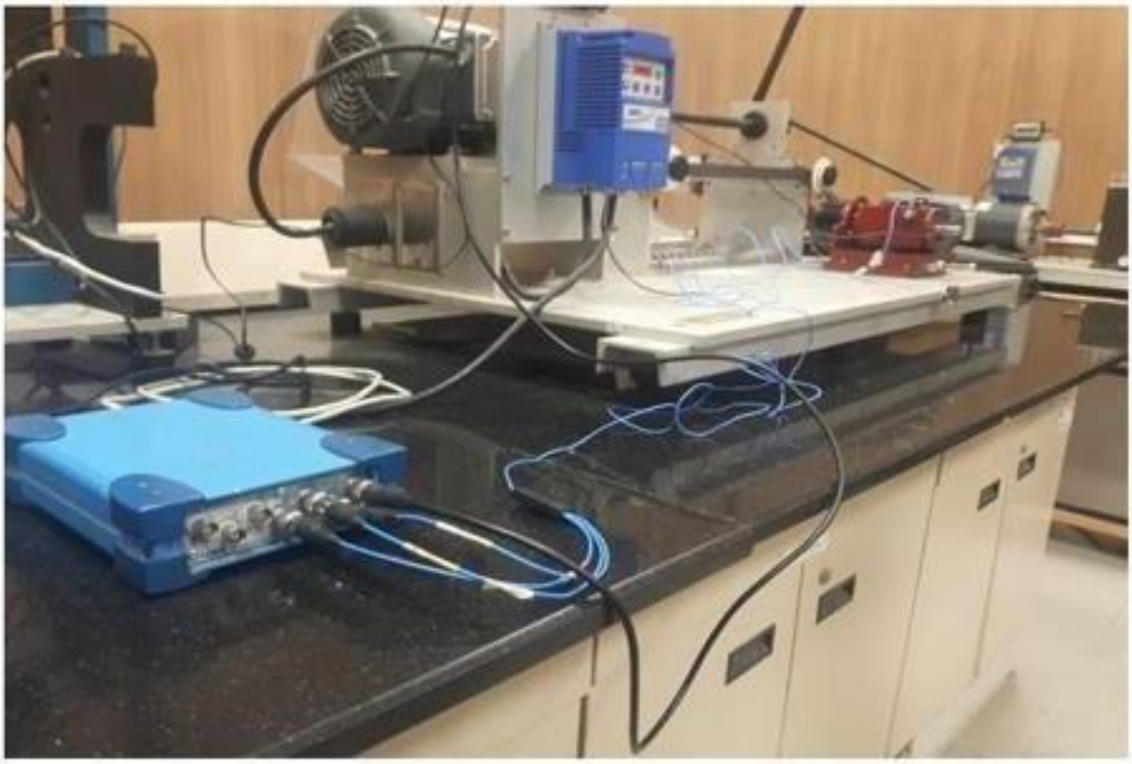

Figure 5: Represents the experimental details.

## 3.2 Experimental setup:

Setup of the experiment as shown in the figure 6 below, the shaft from the motor is connected to the input shaft of the gearbox using pulley and two V belt drive. The pulley is working as the input to the gear box and the output is connected to the loading system. The accelerometer used is a three axial piezo metric accelerometer with the sensitivity of 2.52 mV/g in x direction, 2.57 mV/g in y direction and 2.51 mV/g in z direction. The accelerometer is attached on the gearbox and connected to the DAQ system which is further connected to the monitoring device where the vibrational signals are achieved.




(a)



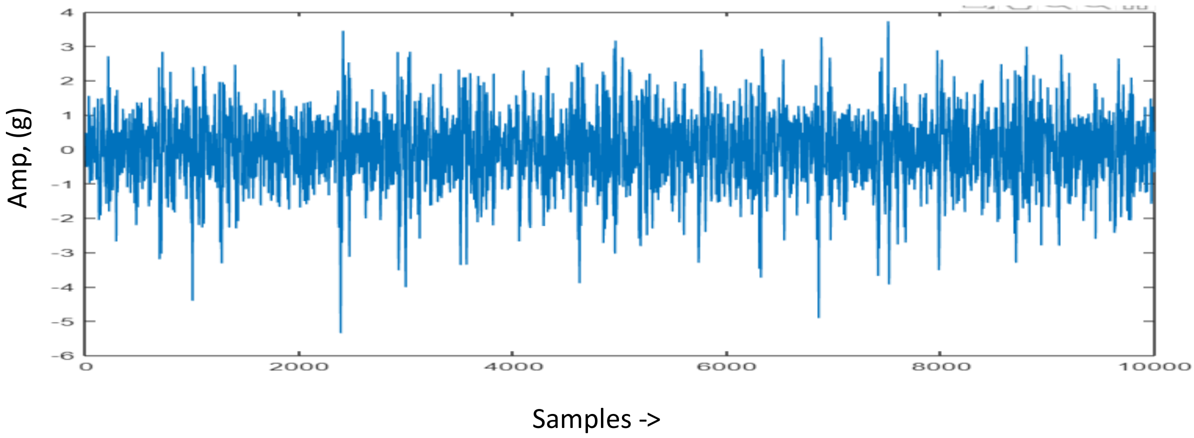
(b)

Figure 6: (a) Machine fault simulator and (b) Data acquisition system

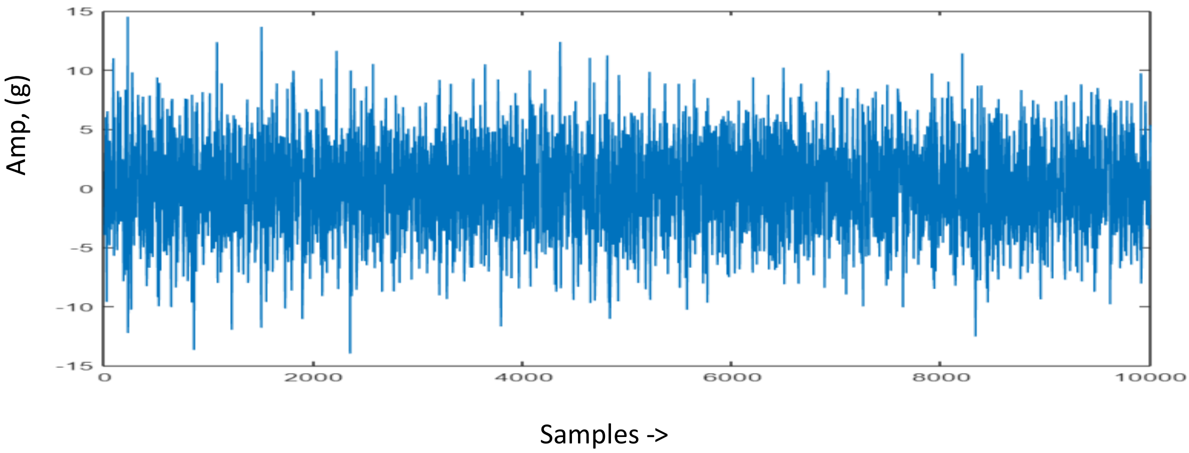
The vibration signals have been taken for all three axis and it has been analyzed that the signals were dominating in z axis (vertical) direction. The signals have been shown below for some of the cases with 10000 number of data points. The effect of the increase in operating frequency and increase in the external torque can be seen by these signals.

The amplitude is increasing by increase in operating frequencies as well as with increase of external torque. The time domain signals are extracted using NV gate software. The y axis of the graph is showing the amplitude of vibration in terms of “g” and the x axis is showing the number of data points/ samples.

## 3.3 Time domain Vibrational signals

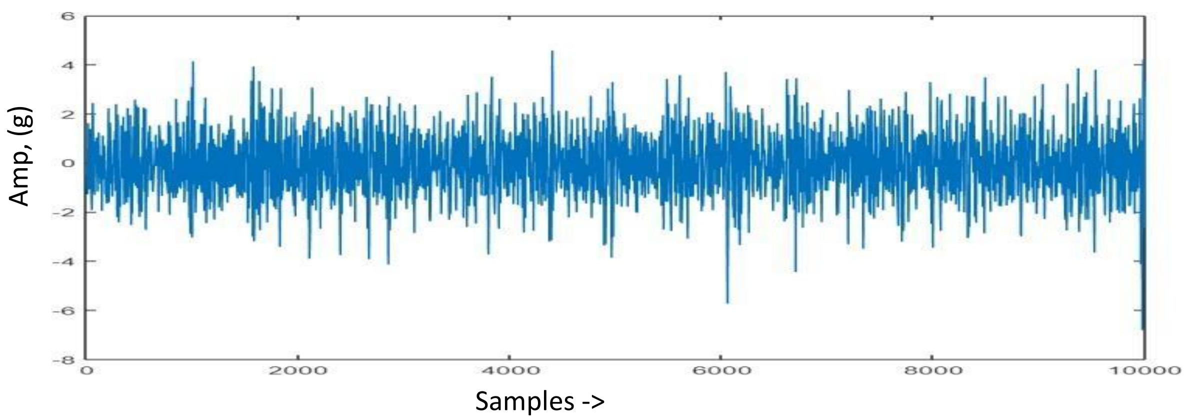



(a)

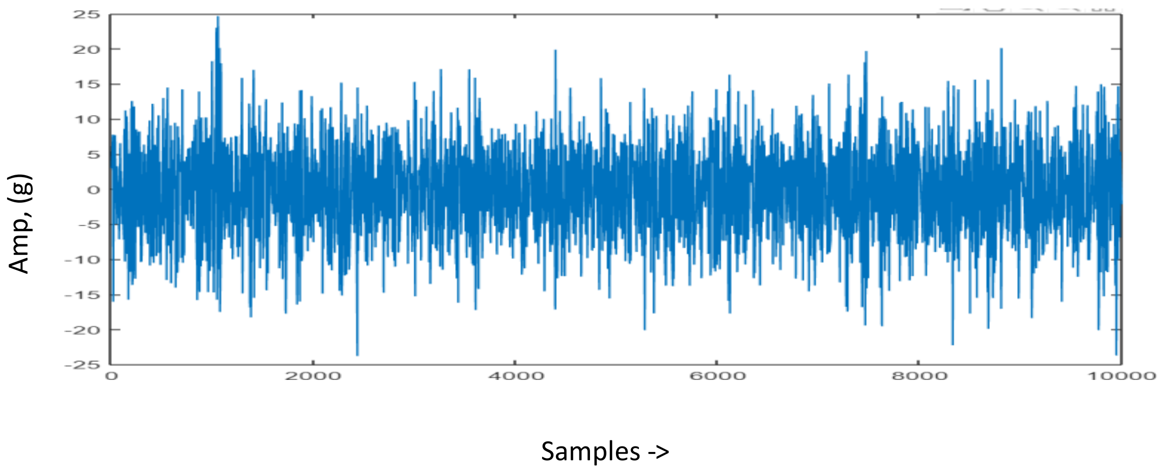



(b)

Figure 7: Represents vertical (z axis) responses for healthy gear at (a) 5 Hz and (c) 15 Hz frequencies at no torque condition.

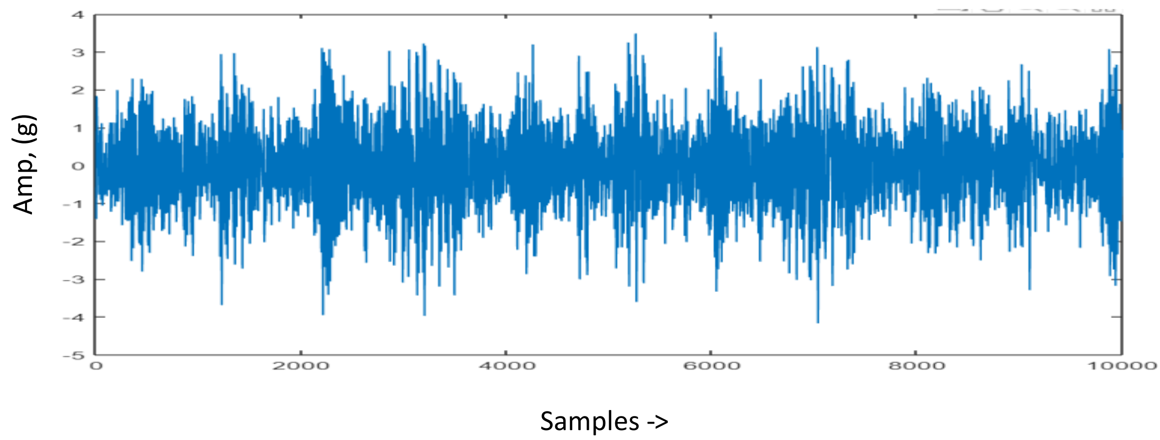



(a)

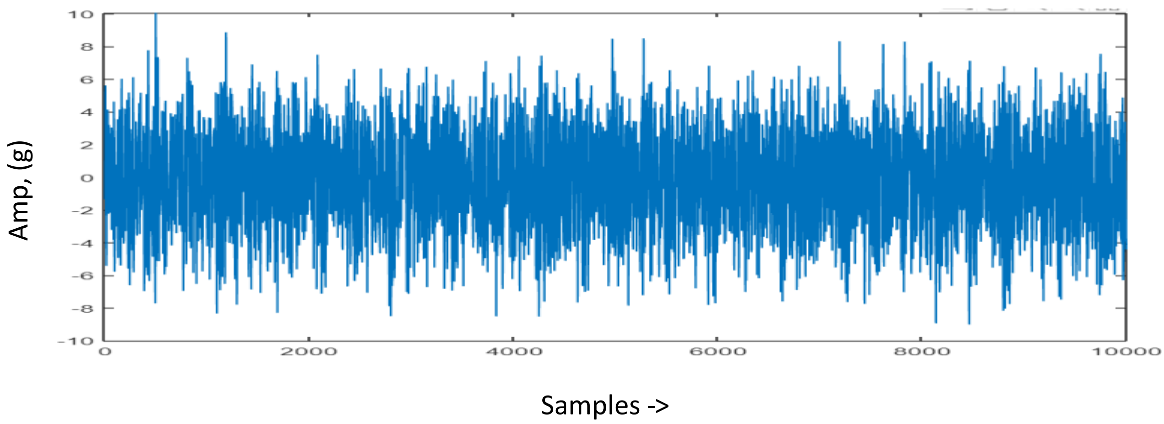



(b)

Figure 8: Represents vertical (z axis) responses for healthy gear at (a) 5 Hz and (c) 15 Hz frequencies at 3Nm torque condition.

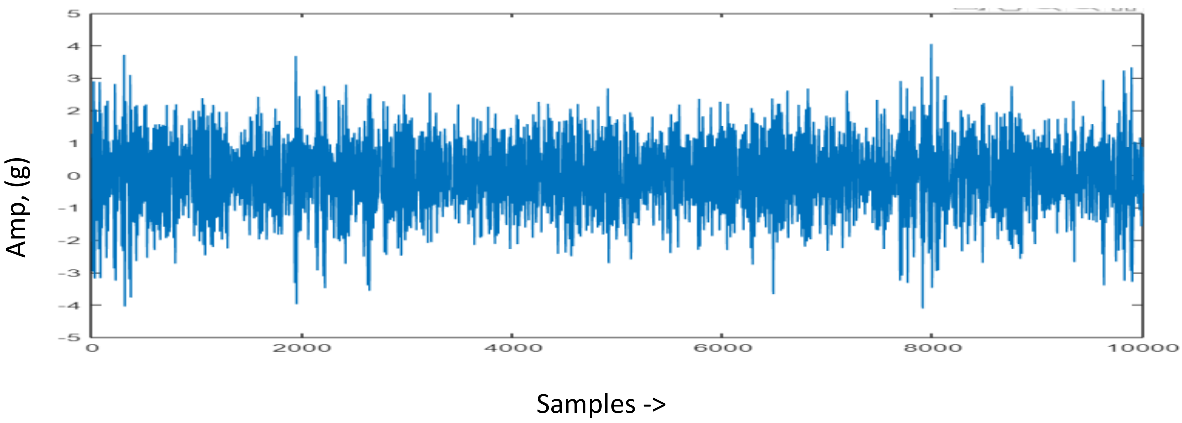



(a)

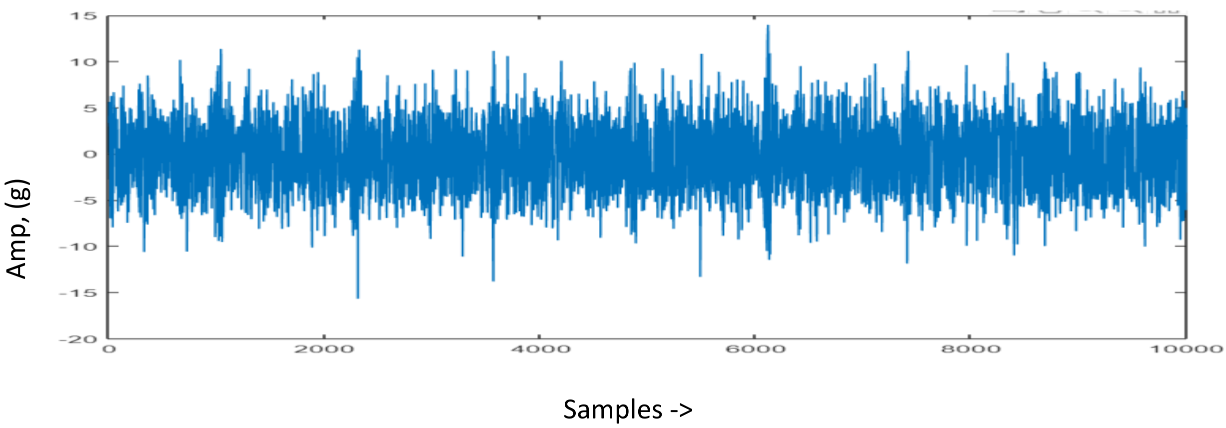



(b)

Figure 9: Represents vertical (z axis) responses for chipped tooth gear at (a) 5 Hz and (c) 15 Hz frequencies at no torque condition.

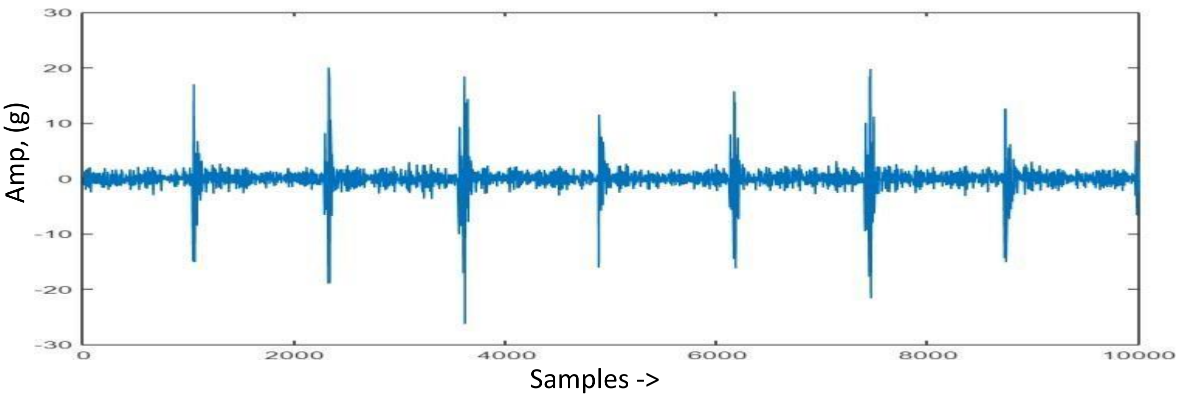



(a)

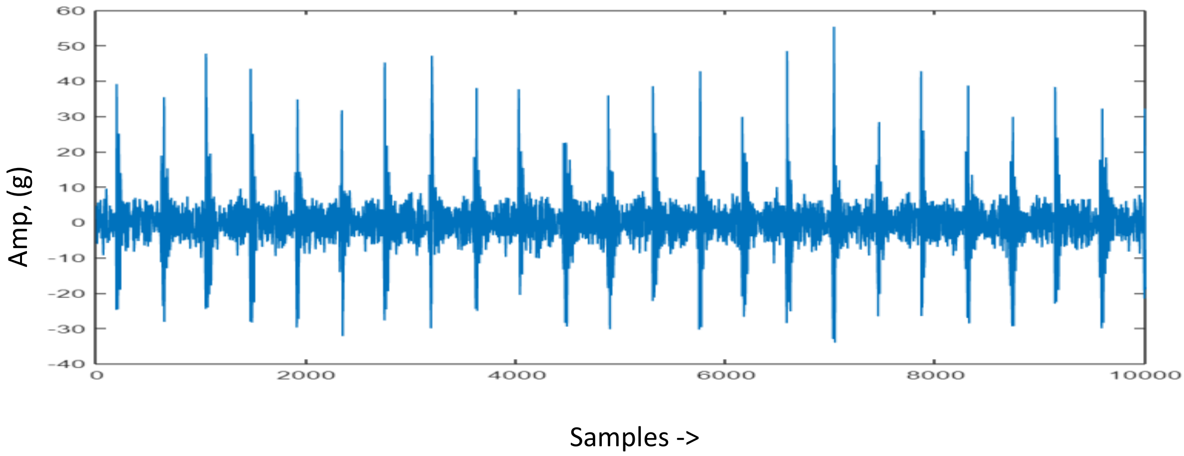



(b)

Figure 10: Represents vertical (z axis) responses for chipped tooth gear at (a) 5 Hz and (c) 15 Hz frequencies at 3 Nm torque condition

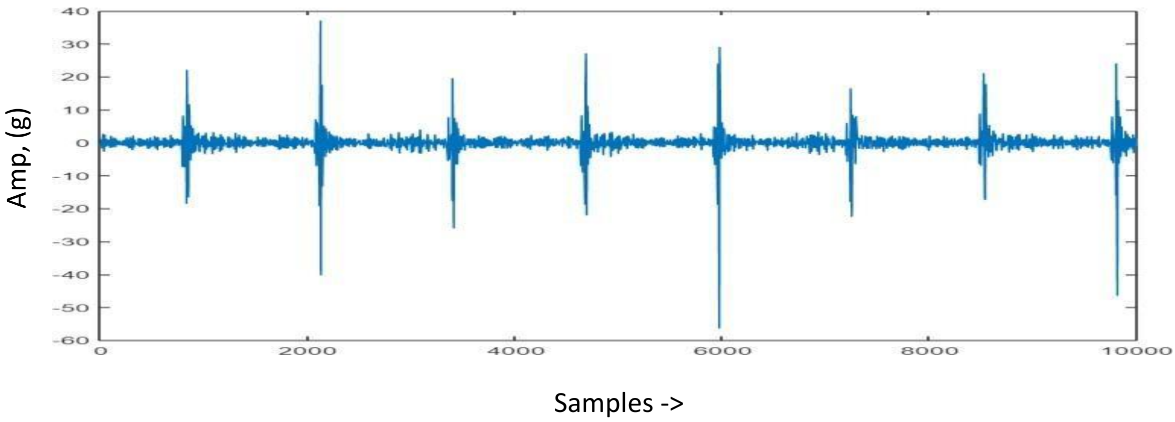



(a)

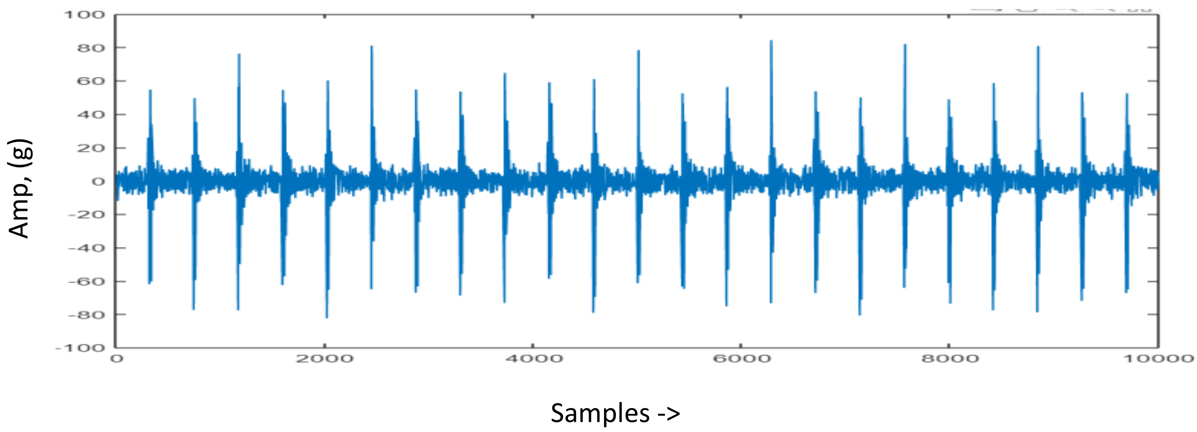



(b)

Figure 11: Represents vertical (z axis) responses for missing tooth gear at (a) 5 Hz and (c) 15 Hz frequencies at no torque condition.




(a)




(b)

Figure 12: Represents vertical (z axis) responses for missing tooth gear at (a) 5 Hz and (c) 15 Hz frequencies at 3Nm torque condition.

# Chapter 4: Methodology

## 4.1 : Feature extraction

The extracted features selected for the input before ranking are:

**1.*Skewness***: skewness is related to the symmetry of the signal i.e. the distribution of the data is either symmetrical or unsymmetrical from the mean position

**2.*Variance***: variance states the odds between instantaneous squared sum (normalized)and mean value.

**3.*Kurtosis***: it is a measure of whether the data gives the peak relative to the normal distribution data or gives the flatness relative to it.

**4.*Standard Deviation***: every signal comprises of energy and standard deviationis what gives the information about the energy contained in signal

**5.*Mean***: ratio of sum of amplitude at all ample points to the number of samples.

**6.*Range***: the difference between max (+ve) and max (-ve) of the amplitude for the given samples.

**7.*Crest factor***: The crest factor is the ratio of the peak amplitude in the whole signal and its root mean square value.

**8.*Torque***: 0, 1, 2, 3 in Nm and Frequency: 5, 10, 15 in Hz

| S. No. | FEATURE | Formulae |
|---|---|---|
| 1. | Skewness | $\frac{\sum_{n=1}^{N}(x(n)-\mu)^3}{N\sigma^3}$ |
| 2. | Variance | $\frac{\sum_{n=1}^{N}(x(n)-\mu)^2}{N}$ |
| 3. | Kurtosis | $\frac{\sum_{n=1}^{N}(x(n)-\mu)^4}{N\sigma^4}$ |
| 4. | Standard Deviation | $\sqrt{\frac{\sum_{i=1}^{N}(x_i-\overline{x})^2}{N-1}}$ |
| 5. | Mean | $\frac{\sum_{n=1}^{N}(x(n))}{N}$ |
| 6. | Range | $\max[x]-\min[x]$ |
| 7. | Crest Factor | $\frac{Peak\ Value}{RMS\ Value}$ |

Table 1: formulation of the extracted features

From the feature in shown in the table 1, it is known that there will be redundancy in the information given by the extracted features. There can be many reasons for that like if one feature is dependent on other than this will create a redundancy etc. so, three types of ranking algorithms (attribute evaluator) are used for ranking the features and the first four ranked features along with load and frequency will be used as input to the classifiers. There is no constraint of using only first four ranked features but to set the same parameter for comparison of the results after applying these attribute evaluators we have to choose same number of inputs.

## 4.2 : Attribute evaluator

**Info gain attribute evaluator:**

Information gain generally describes the useful information given by each attribute by calculating the total information and subtracting the entropy associated with that information for corresponding attribute. The value of gain will always be between (0-1).

Let us suppose a data set S to be having s number of samples with m variable values of classes. The relevant information which is sufficiently reliable to categorize the data in classes is given by M.A. Jayaram et al. [Asha Gowda Karegowda, A. S. Manjunath & M.A. Jayaram, 2010]

Gain = Information (I(S)) – entropy E(A)

$$I(s) = -\sum_{i=1}^{m} p_i \log_2 p_i \quad [5]$$

Here we know that $p_i$ is the probability of a variable sample that it belongs to the class Ci and is mathematically formulated by $s_i$/s.

If we select attribute A having v different values then suppose $s_{ij}$ to be giving the samples present in a subset $S_j$ of class Ci. $S_j$ is a subset having only samples with values $a_j$ of A. The relevant information or the loss of information is dependent on the separation into subsets by A, is given by:

$$E(A) = -\sum_{i=1}^{m} I(S)\,\frac{s_{1i}+s_{2i}+s_{3i}+\cdots+s_{mi}}{s} \qquad [6]$$

**Gain Ratio attribute evaluator:**

$$Splitinfo(S) = -\sum_{i=1}^{m} \frac{|s_i|}{|s|}\log_2\frac{|s_i|}{|s|} \qquad [7]$$

The value $Splitinfo(S)$ represents the information loss or addition by partitioning of the data selected for training into v partitions respectively for v outcomes for attribute A while testing of data.

The definition of gain ratio is:

Gain Ratio(A) = Gain (A)/ Split Info$_A$(S)

The attribute which is giving the highest value for the above ratio is chosen for splitting the attributes.

**One R attribute evaluator:**

One R i.e. "One Rule", is a simple technique but gives the accurate results in which we check the classification accuracy giving single attribute at a time and the highest accuracy giving attribute is ranked first and so on. This technique will give specify ranking taking one attribute at a time that is why it is called One R rule.

For every attribute the classification is as follows;

1. The counting of that how often the same value is appearing for target class.
2. Finding the highly frequent class i.e. for which the value appears maximum.
3. The calculation of the error for each attribute.
4. Arranging the attribute in order of increasing error for ranking.
5. Selecting the highly ranked attributes for further classification

## 4.3 Ranker model

The experimental details are shown in figure 4. There are total 9 attributes shown in table 1 including torque and operating frequency and 180 number of samples. The samples are classified using four classifiers namely ANN, DNN, SVM and random forest before and after applying ranker algorithm. Weka tool is used for the classification and the results are shown below.

The ranker algorithm is applied to the extracted features for feature selection which results in the specifying the rank of the corresponding features the top ranked features are selected on the basis of forward check method i.e. checking the frequency by giving firstly the top one feature then the top two features and so on.

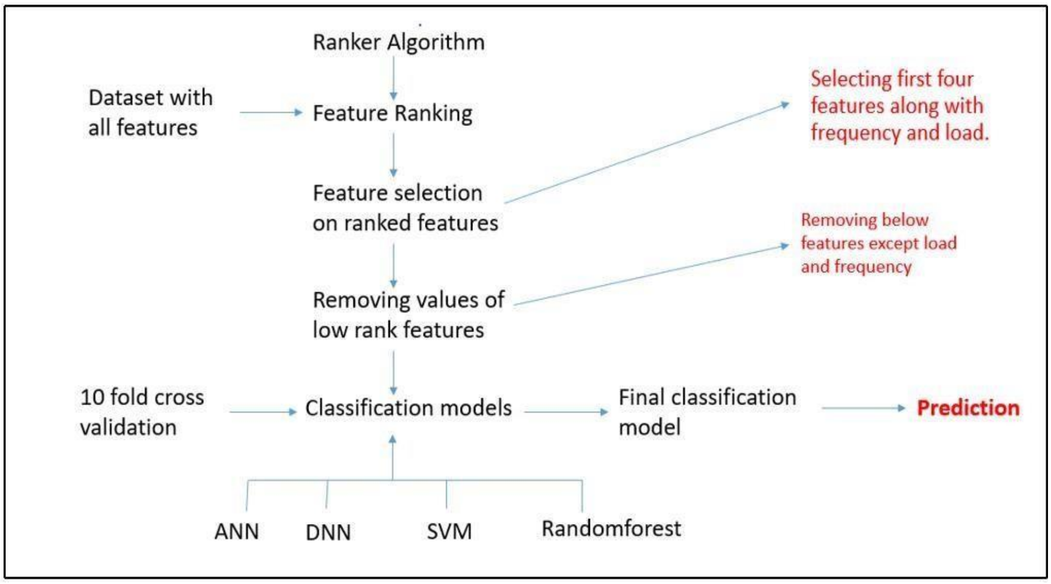



Figure 13: Representation of ranking working model

The best accuracy received in output is by using first four features along with load and frequency. The cross-validation method used is 10-fold cross validation has been taken for validation of the output in which the total samples are distributed into 10 parts and from those 9 parts are taken for training data and 1 part is taken for testing data. This is repeated for 10 times.

# Chapter 5: Results and Observation of stage 1 of the project

## 5.1 Ranking results given by all three evaluators

| Ranking by evaluator | Feature name and direction |
|---|---|
| 1 | Kurtosis |
| 2 | Crest factor |
| 3 | Standard deviation |
| 4 | Range |
| 5 | Variance |
| 6 | Skewness |
| 7 | Mean |

(a)

| Ranking by evaluator | Feature name and direction |
|---|---|
| 1 | Kurtosis |
| 2 | Range |
| 3 | Crest factor |
| 4 | Skewness |
| 5 | Variance |
| 6 | Standard deviation |
| 7 | Mean |

(b)

| Ranking by evaluator | Feature name and direction |
|---|---|
| 1 | Kurtosis |
| 2 | Range |
| 3 | Crest factor |
| 4 | Variance |
| 5 | Skewness |
| 6 | Standard deviation |
| 7 | Mean |

(c)

Table 2: Extracted feature ranking by (a) info gain attribute evaluator, (b) gain ratio attribute evaluator (c) one R attribute evaluator.

## 5.2 Resulting accuracy of the model before and after applying the ranking technique:

| Ranker -> Classifier | Before ranking Accuracy in %age | Info gain Accuracy %age | Gain ratio Accuracy %age | One R Accuracy %age |
|---|---|---|---|---|
| ANN | 85.5556 | 90 | 91.1111 | 90 |
| DNN | 86.1111 | 91.6667 | 91.6667 | 91.1111 |
| SVM | 93.3333 | 95 | 95.5556 | 94.4444 |
| Random Forest | 95.5555 | 96.1111 | 97.1111 | 96.1111 |

Table 3: Represents the classification accuracy before and after the application of attribute evaluator.

In the above table the column is representing the classification models and the row is denoting the type of attribute evaluator. The ANN classifier is giving 85.5556% accuracy before application of ranker algorithm and the accuracy is increased to 91.1111% as highest for this classifier with gain ratio attribute evaluator. Similarly, for DNN the accuracy before feature ranking is 86.1111% and the highest achieved is 91.6667% with gain ratio attribute evaluator. Again, for SVM and random forest classifier the accuracy before applying ranker algorithms are 93.3333% & 95.5555% respectively and the highest is 95.5555% and 97.1111% achieved by gain ratio attribute evaluator respectively.

## 5.3 Conclusion:

In this paper the model accuracy for different rankers with different models have been shown and we can see that by using the attribute evaluator the model accuracy is improved. For all the cases, the random forest classifier is giving the highest accuracy followed by SVM. It can also be seen in the table 3 that the highest accuracy for random forest and SVM has been achieved by gain ratio attribute evaluator.

For every classifier, the gain ratio is giving the highest accuracy. On the basis of above a conclusion can be made that for the better classification of the faults random forest along with gain ratio evaluator should be used.

# Chapter 6: Decomposition of signal using wavelet analyzer.

## 6.1Requirement of decomposition

In the previous work, it can be noted easily that we can improve accuracy to 100% by adding more techniques or by hybridization. For that we can use pre-processing techniques to analyze the signal because real world data have much more information. It has slowly changing trends and oscillations punctuated with transients. Therefore, to accurately analyze the exact behavior of the signals we require the new class of functions required that are well established in time domain and frequency domain.

Some authors used Discrete wavelet transform to analyze the signal and then applied adaptive neuro fuzzy technique to classify and compare the healthy gear, single local defect, multilocal defect, etc. [Jian-Da Wu, Chuang-Chin Hsu, 2009]. Authors also performed discrete wavelet transform to extract the feature was used and the various wavelets have different energy levels and the wavelet having highest potential energy is the parameter on the basis of which the gear faults are classified. The extracted features form the wavelet decomposition is given to neural network for decision making. The neural network performs better in both classification and decision making. It was concluded that neural network has a high potential and effective in gearbox fault diagnosis. [N. Saravanan, K.I. Ramachandran, 2014].

Researchers have also used wavelet transforms like discrete wavelet transforms (DWT) is applied on the signals acquired by the accelerometer for the rotating machinery. DWT gives the sequence of detailed signals with various scales. For all those various detailed multiscale signals, the variances are calculated. At last, the multiscale wavelet-based estimation of the features from the logarithmic slope of variances. It was concluded that the multiscale wavelet-based slope features have the advantage of greater accuracy and sustainability in classifying different fault conditions of gearbox, bearing or other rotatory components. [Gang Cheng, Xihui Chen, Hongyu Li, Peng Li, Fanrang Kong, Qingbo He, Yongbin Liu, 2018]. Transmission errors (TE) are also put into consideration. [Sungho Park, Seokgoo Kim, Joo-Ho Choi, 2018]. The transmission errors for the gears having different faults like with spall, crack etc. are evaluated respectively. Ensemble empirical mode decomposition is germane for feature extraction under the vibration from the evaluated TE. The variation of the crack and spall is accurately judged by the intrinsic mode function’s selected features. After that the K Nearest Neighbor technique is used for the classification of the faults in the gears using the extracted features.

## 6.2 Problem formulation and objective

Here we are seeking for somehow improving the accuracy of the classification. In the stage 2 the by having the literature review a point is considered that the signal acquired by the machine will specify the actual condition world data have much more information. It can be done using DWTs, CWTs, fuzzy logic etc. here we are using a modified DWT i.e. Maximal Overlap Discrete Wavelet Transform (MODWT) and empirical mode decomposition. the wavelets are applied to decompose the signal such that the signal could be analysed properly without any loss of information. Some of the wavelets are Morlet, Daubechics, Coiflets, Biorthogonal, Maxican Hat, Symlets, etc.

## Objective:

In the literature review it has been seen that the authors are using wavelet transforms like DWT, CWT and EMD along with one of the classifiers and gave the results as in terms of classification but the consideration on the comparison of wavelet transform technique is much lower. So, the objective here is to compare the wavelet transform technique like with the usage of SVM classifier and Ranker algorithm. The wavelet transform used in this experiment is MODWT and EMD and the ranker algorithm used id ReliefF.

## 6.3 Maximal overlap discrete wavelet transform

As compared to discreate wavelet transform, maximal overlap discreate wavelet transformcan be used for any size of the sample. In case of DWT the sample size should be the power of two whereas there is no such condition of in case of MODWT.Other features of DWT like (MRA) multiresolution analysis Similar can be performed using MODWT but there is some advantage using MODWT over

DWT that it gives more smoothness and detailed decomposition that shifts along original signal. In MODWT discreate wavelet power spectrum is similar as that of original signal and circular shift. The equation below gives the expression of a continuous wavelet transform. [N. Saravanan, K.I. Ramachandran, 2014]

$$T(a, b) = \int_{-\infty}^{\infty} f(t)\varphi *_{(a,b)}(t)dt \quad [8]$$

Here * denotes the complex conjugate.

$$\varphi_{(a,b)}(t) = \frac{1}{\sqrt{a}}\varphi\left(\frac{t-b}{a}\right) \quad (a, b \epsilon R; a \neq 0) \quad [9]$$

The equation for DWT is driven out by discretization of CWT wavelet $\varphi_{(a,b)}(t)$

$$\varphi_{(j,k)}(t) = \frac{1}{\sqrt{2^j}}\varphi\left(\frac{t-2^j k}{2^j}\right) \quad [10]$$

Approach called pre-processing of data focuses on the processing of the dataset by decomposition the originally acquired time domain signal into more stable or static and constant subseries which are in general more advantageous to analyze. The decomposition can help in the elimination of the non-useful and redundant feature from the original data set. By this we get more contingent informative decomposed series and the informative training data is further responsible for increment in the overall accuracy.

Fourier transforms are the base by which the DWTs are developed. The equation of DWT mathematically used to decompose time domain frequency signal in various sub components. The common advantage of the DWT over Fourier transforms is that while using DWT we have the perfect in-depth analysis of the resulting decomposed signal good scale of resolution. Since DWT collects the proper and relevant information at every decomposed level, the DWT increases the capability of the model. Additionally, the DWT is suitably preferred for analysis of the data because the transients and the abrupt changes in the real world data will specify the exact condition of machine and these changes and non-periodic are analyzed using DWT.

## 6.4 Empirical mode decomposition

(Empirical Mode Decomposition) is a method of decomposition of signal on the basis of intrinsic mode function. It is a suitable to process the signals which are non-stationary and having abrupt changes and transients. EMD partitions the original time domain signal into series which are called modes on the basis of IMFs.

Decomposition by EMD works as follows:

- Let we have a data set x(t)
- Next step includes the interpolation and identification of local maxima and local minima in order to have upper and lower envelop U(x) and L(x).
- Evaluate the mean of the envelop identified in previous step

$$m(t) = \frac{U(x) + L(x)}{2} \quad [11]$$

- Extraction of the mean signal from the original signal. Resulting signal is h(t).
  h(t) = x(t) – m(t)
- Check weather resulting signal, the IMF condition are satisfied or not. If yes stop shifting, If no then h(t) = x(t) and keep shifting.
- The criteria for the sifting process to stop: standard deviation computed for two consecutive sifting should be in nearly equal value.

$$SD = \sum_{t=0}^{T} \frac{[|(h_{1(k-1)}(t) - h_{1k}(t)|^2]}{[h_{1(k-1)}(t)]^2} \quad [12]$$

$h_{1k}: residue\ after\ kth\ iteration\ of\ the\ 1st\ IMF$

## 6.5 Relief F ranker algorithm

Relief F is a ranker algorithm developed by Rendell and Kira inspired by instance-based learning. it is a feature selection process using which the ranking of the features are done. in relief F proxy statistics is calculated for every feature by which the feature quality and relevance is been estimated. it is a univariate method. the feature statistic is generally the feature weight [w(A) = weight of feature 'A'] or it also referred to as the score of the feature and ranges from worst (-1) to best (+1). [Yu Yang, Yigang He, Junsheng Cheng, Dejie Yu, 2011].

In every iteration, the feature value (x) which belong to one of the particular instances and the feature value of the instance nearest to x from each class. the nearest similar instance class is called 'near-hit' and the nearest variational instance class is called 'near-miss'. finally updating the weight value as:

$$W_{i+1} = W_i - (x_i - nearHit_i)^2 + (x_i - nearMiss_i)^2 \quad [13]$$

# Chapter 7: Decomposition level for healthy and faulty gear

## 7.1 MODWT decomposition results

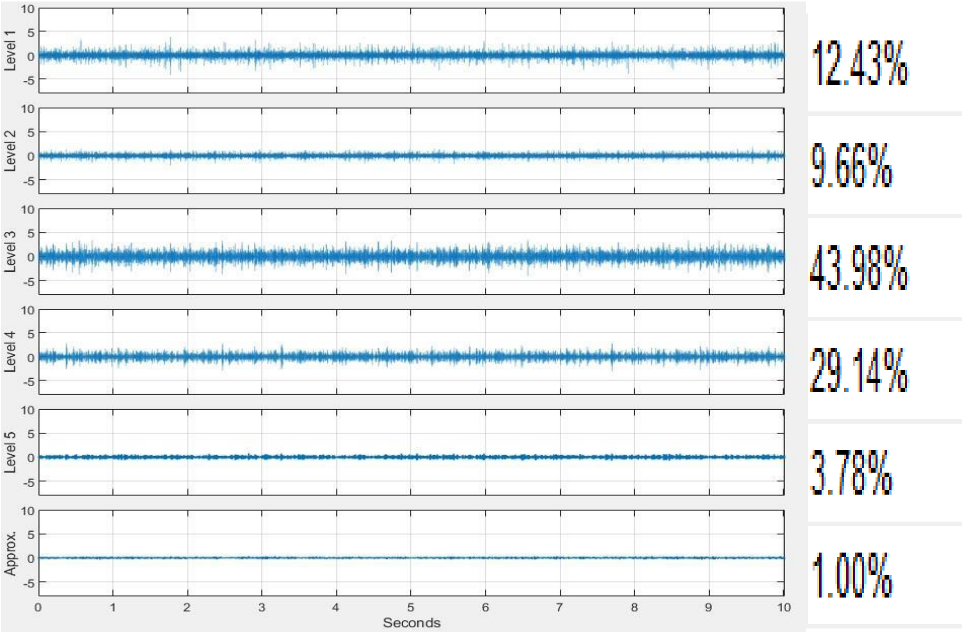



Figure14: Results of MODWT decomposed signal up to level 5 for healthy gear under no load condition.

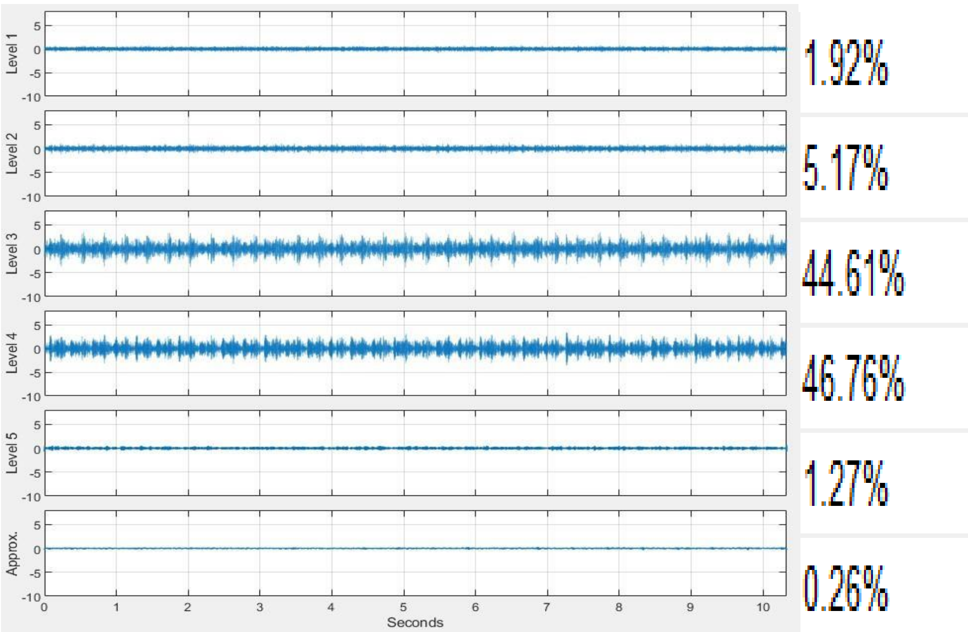



Figure15: Results of MODWT decomposed signal up to level 5 for chipped tooth gear under no load condition.

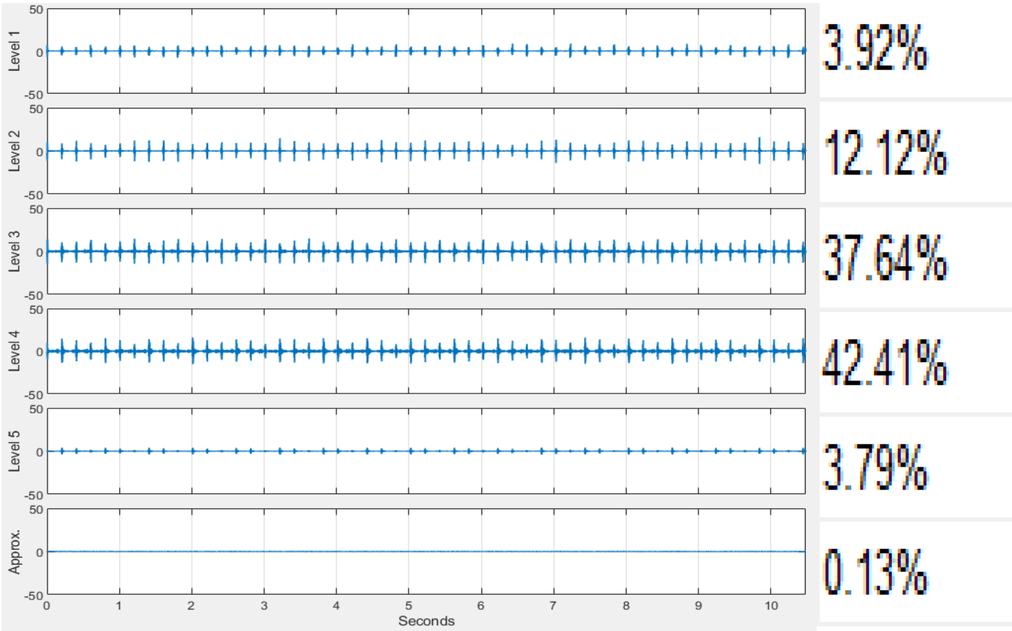



Figure16:Results of MODWT decomposed signal up to level 5 for missing tooth gear under no load condition.

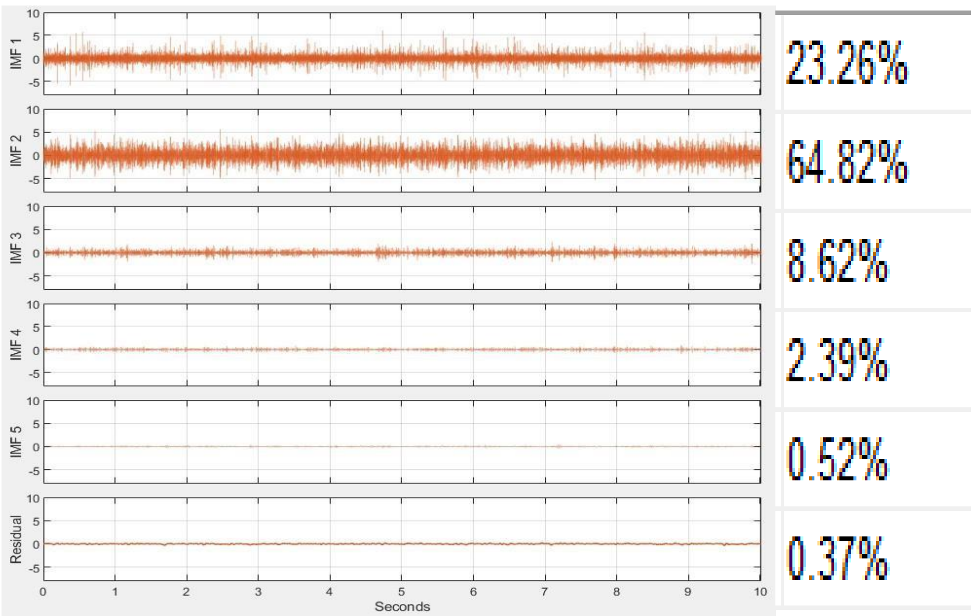



Figure17: Results of EMD decomposed signal up to level 5 for healthy gear under no load condition.

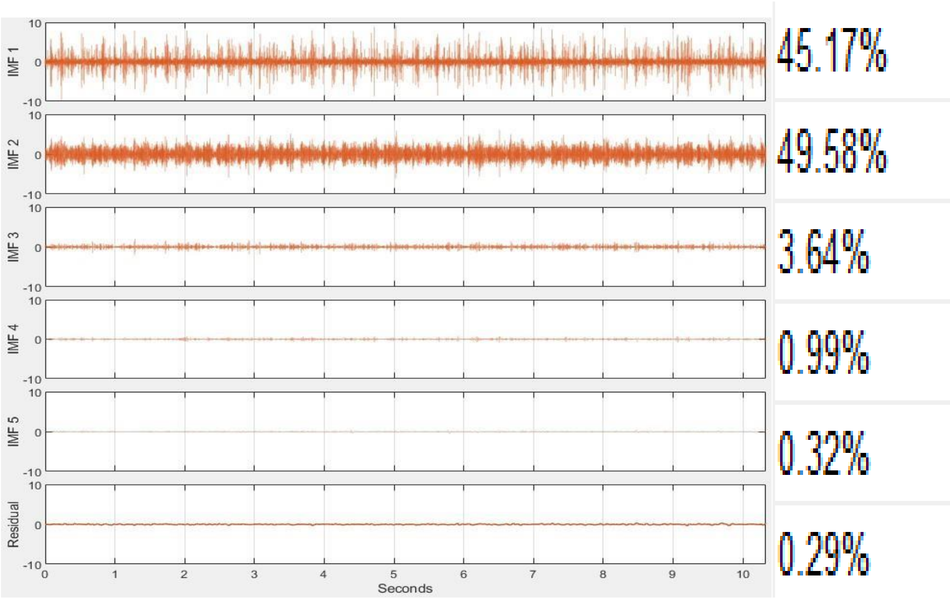



Figure18: Results of EMD decomposed signal up to level 5 for chipped tooth gear under no load condition

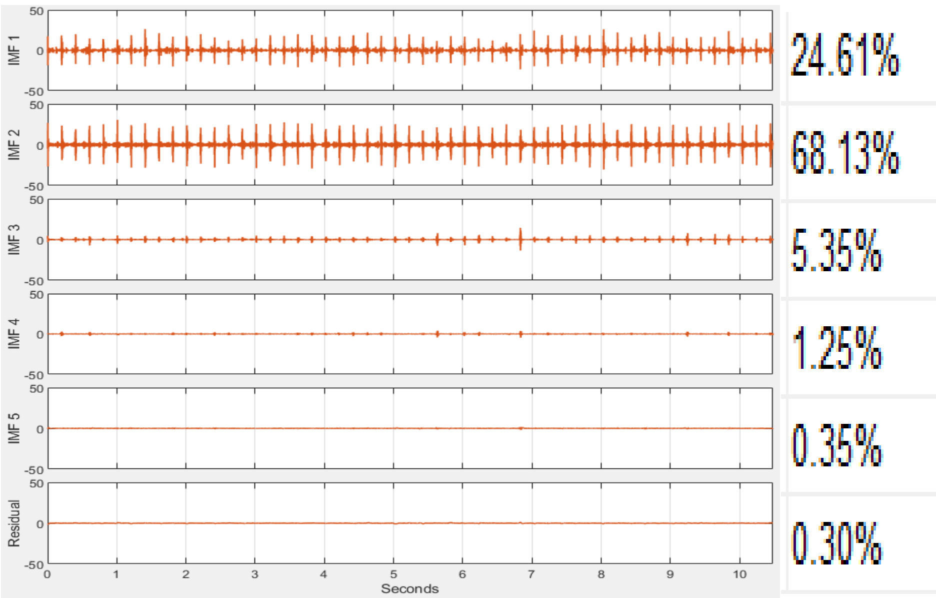



Figure19: Results of EMD decomposed signal up to level 5 for missing tooth gear under no load condition

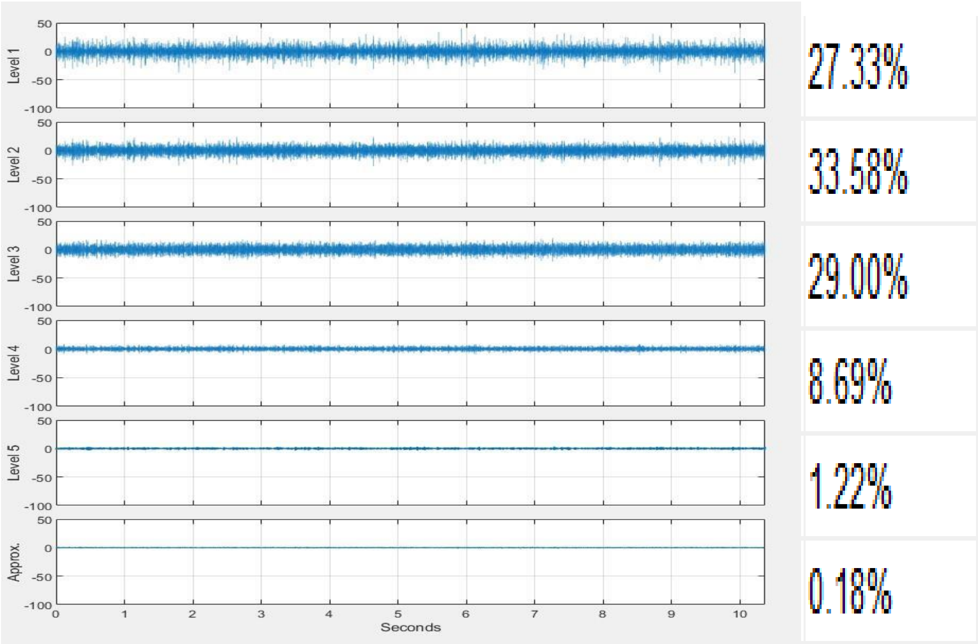



Figure 20: Results of MODWT decomposed signal up to level 5 for healthy gear under full load condition

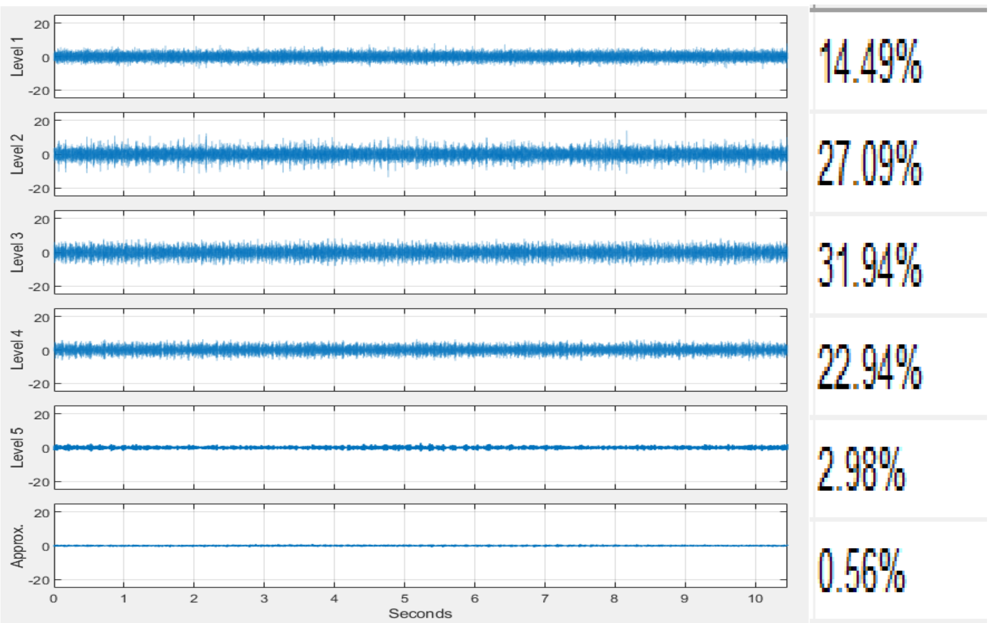



Figure 21: Results of MODWT decomposed signal up to level 5 for chipped tooth gear under full load condition

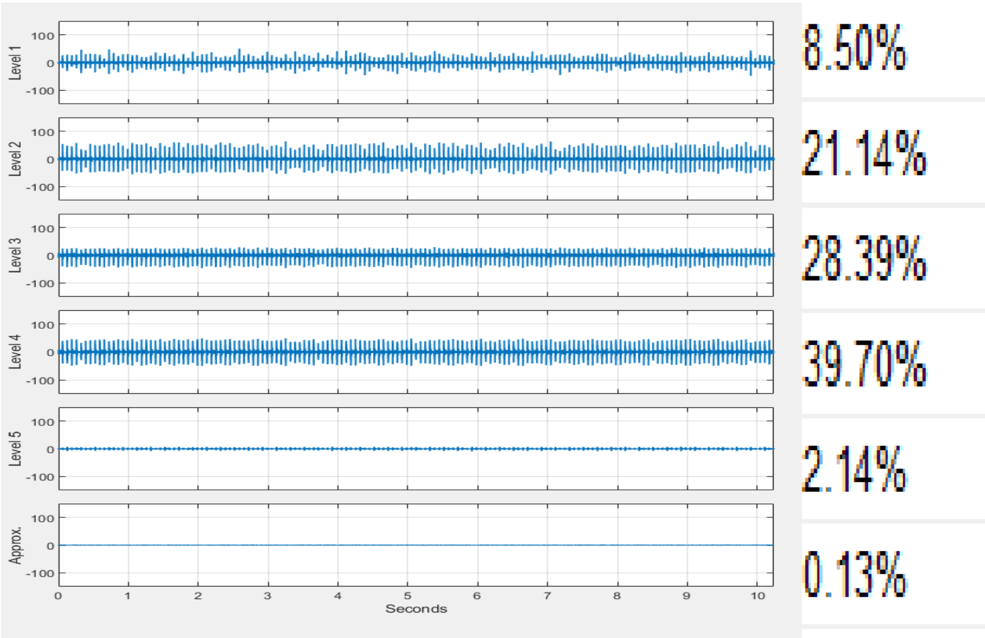



Figure 22: Results of MODWT decomposed signal up to level 5 for missing tooth under full load condition

## 7.2 EMD decomposition results

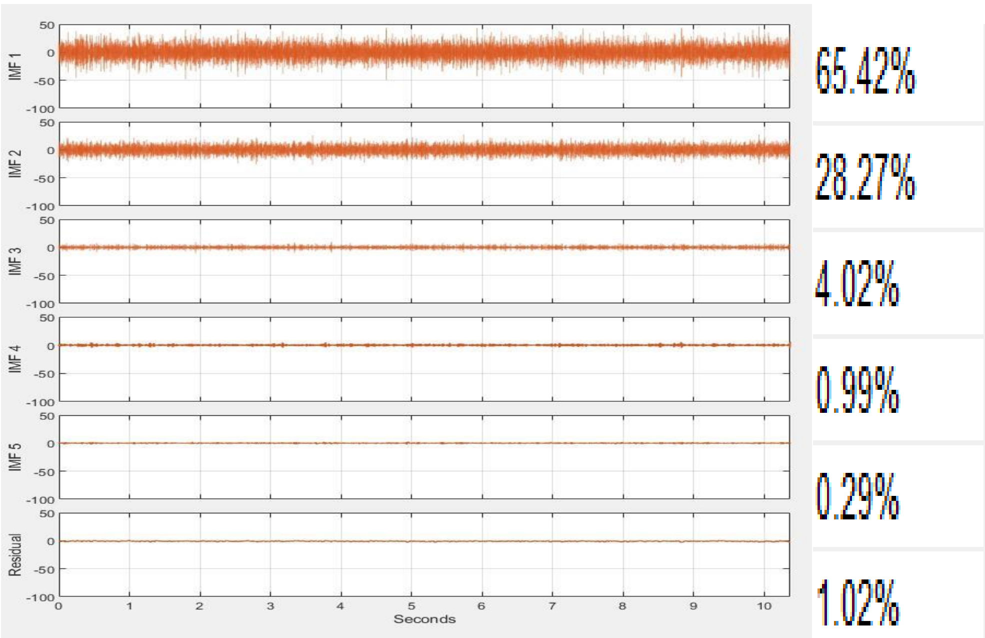



Figure 23: Results of EMD decomposed signal up to level 5 for healthy gear under full load condition.

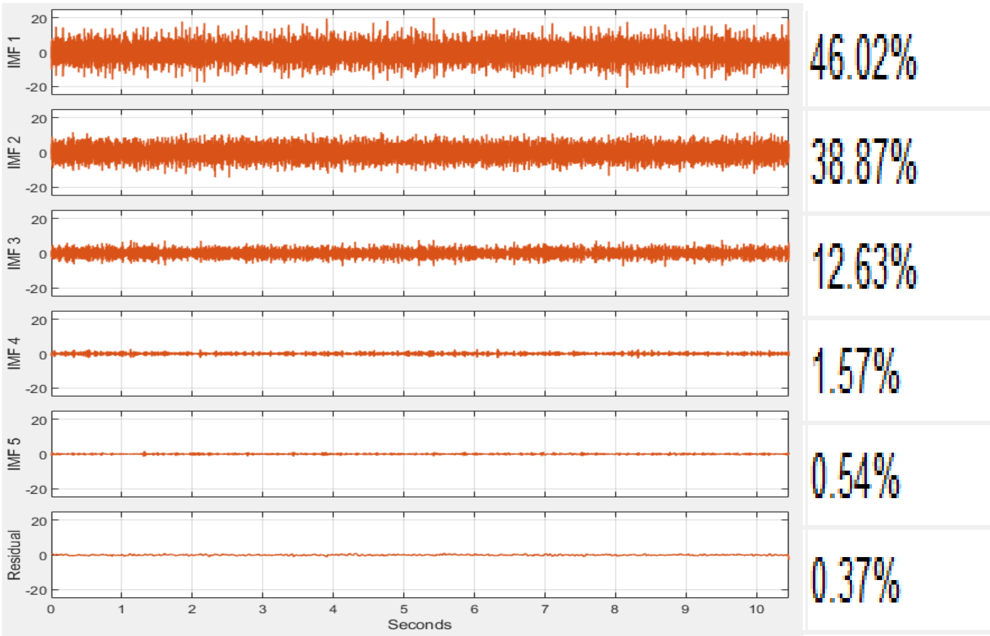



Figure 24: Results of EMD decomposed signal up to level 5 for chipped tooth gear under full load condition

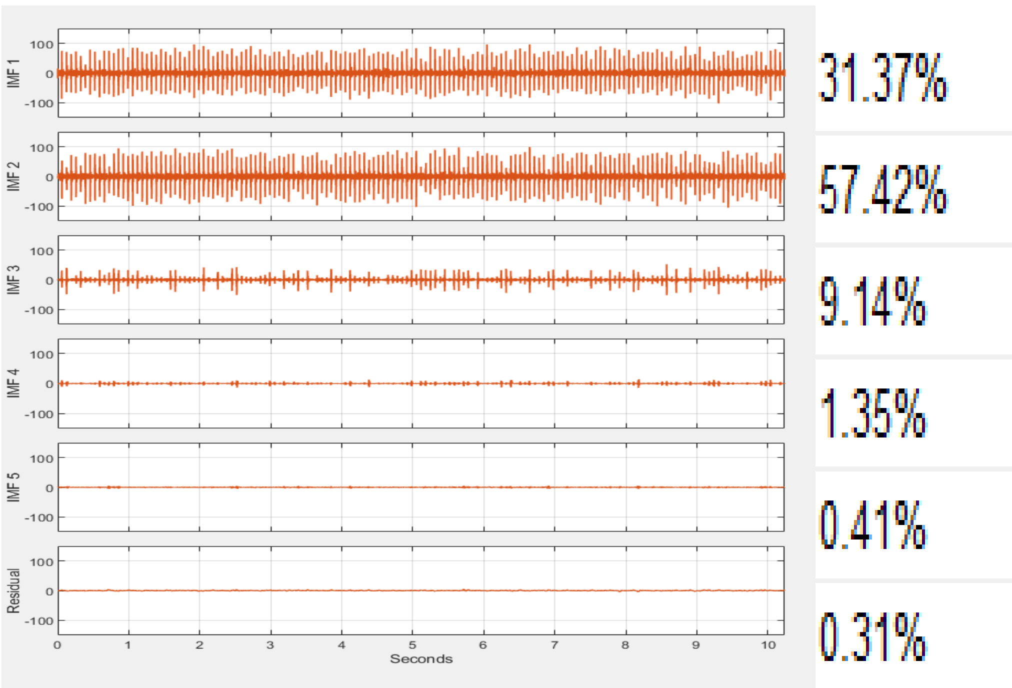



Figure 25: Results of EMD decomposed signal up to level 5 for missing tooth gear under full load condition.

## 7.3 Results of relief F ranker algorithm

| Rank | Features along z axis |
|---|---|
| 1st | Kurtosis |
| 2nd | Crest |
| 3rd | Range |
| 4th | Standard deviation |
| 5th | Skewness |
| 6th | Variance |
| 7th | Mean |

(a)

| Rank | Features along z axis |
|---|---|
| 1st | Range |
| 2nd | Crest |
| 3rd | kurtosis |
| 4th | Standard deviation |
| 5th | variance |
| 6th | Mean |
| 7th | Skewness |

(b)

Table 4: (a)Ranking of features after MODWT is applied to vibration signal (b) Ranking of features after EMD is applied to vibration signal.

# Chapter 8: Overall result and conclusion

## 8.1 Classification results using support vector machine technique

| | |
|---|---|
| **Total Number of Instances** | **180** |
| Correctly Classified Instances | 145 |
| Incorrectly Classified Instances | 35 |
| Classification accuracy | 80.5556 % |
| Classification error percentage | 19.4444 % |

(a)

| | |
|---|---|
| **Total Number of Instances** | **180** |
| Correctly Classified Instances | 142 |
| Incorrectly Classified Instances | 38 |
| Classification accuracy | 78.8889 % |
| Classification error percentage | 21.1111 % |

(b)

| | |
|---|---|
| **Total Number of Instances** | **108** |
| Correctly Classified Instances | 103 |
| Incorrectly Classified Instances | 5 |
| Classification accuracy | 95.3703 % |
| Classification error percentage | 4.6297 % |

(b)

| | |
|---|---|
| **Total Number of Instances** | **108** |
| Correctly Classified Instances | 96 |
| Incorrectly Classified Instances | 12 |
| Classification accuracy | 88.8889 % |
| Classification error percentage | 11.1111 % |

(d)

Table 5: (a) Classification accuracy for MODWT by taking 5 levels (b) Classification accuracy for EMD by taking 5 levels (c) Classification accuracy for MODWT by taking 3 levels (d) Classification accuracy for EMD by taking 3levels

## 8.2 Classification results using Randomforest technique

| | |
|---|---|
| **Total Number of Instances** | **180** |
| Correctly Classified Instances | 157 |
| Incorrectly Classified Instances | 23 |
| Classification accuracy | 87.2223 % |
| Classification error percentage | 12.7777 % |

(a)

| | |
|---|---|
| **Total Number of Instances** | **180** |
| Correctly Classified Instances | 152 |
| Incorrectly Classified Instances | 28 |
| Classification accuracy | 84.4445 % |
| Classification error percentage | 15.5555 % |

(b)

| | |
|---|---|
| **Total Number of Instances** | **108** |
| Correctly Classified Instances | 106 |
| Incorrectly Classified Instances | 2 |
| Classification accuracy | 98.2222 % |
| Classification error percentage | 1.7778 % |

(b)

| | |
|---|---|
| **Total Number of Instances** | **108** |
| Correctly Classified Instances | 101 |
| Incorrectly Classified Instances | 7 |
| Classification accuracy | 93.5185 % |
| Classification error percentage | 6.4815 % |

(d)

Table 6: (a) Classification accuracy for MODWT by taking 5 levels (b) Classification accuracy for EMD by taking 5 levels (c) Classification accuracy for MODWT by taking 3 levels (d) Classification accuracy for EMD by taking 3levels.

## 8.3 Accuracy for each classifier with different energy signals taken

| Techniques / Classifier | Taking highest energy 3 signals MODWT | Taking highest energy 3 signals EMD | Taking highest energy 5 signals MODWT | Taking highest energy 5 signals EMD |
|---|---|---|---|---|
| SVM | 95.3703 % | 88.8889 % | 80.5556 % | 78.8889 % |
| Random forest | 98.2222% | 93.5185 % | 87.2223 % | 84.4445 % |

Table 7: Representing the classification accuracy for corresponding to each technique and classifier.

## 8.4 Conclusion:

It has been noted that the MODWT is giving higher accuracy for each case as compare to EMD. It has been also noted that random forest is giving higher accuracy as compare to support vector machine.

The highest accuracy has been achieved by applying MODWT along with relief F and random forest which is 98.2222 %.

From the above result it can be concluded that maximal overlap discrete wavelet transform technique gives higher accuracy as compare to empirical mode decomposition. So, on the basis of classification accuracy it is better to use MODWT along with relief F algorithm to achieve maximum accuracy.

## 8.5 Overall conclusion of the thesis:

In this experiment we have taken the vibration signal for healthy condition, chipped tooth condition and missing tooth condition of the gear for our study.

In the first stage of the project the features have been extracted from the signals using mathematical formulae shown in table 1 and then the features are ranked on the basis of the useful information or we can say less entropy information delivered by them that helps in classifying faults more accurately. Three ranking algorithms have been used to rank the features along with four classifiers which results a combination of ranker and a classifier giving highest accuracy. We have the ranking os feature for all three ranker algorithms shown in table 2. It can be concluded by table 3 that the accuracy of classification will always be increase by using ranking algorithm as the accuracies in all classifiers are greater after ranking. For all the classifiers the gain ratio attribute evaluator is giving higher frequency i.e. 91.11% for ANN, 91.67% for DNN, 95.55% for SVM and 97.1111**%** for random forest. It is concluded that amongst all above specified accuracy the gain ratio attribute evaluator with random forest is the best combination for attaining maximum accuracy. It can also be concluded that for all attribute evaluators, the random forest and SVM are giving higher accuracies than ANN and DNN. So, only random forest and SVM classifiers are used in second stage.

In the second stage of the project the work is focused to improve the accuracy more than what we got in first stage. To achieve the same a class of functions that are used to study the abrupt changes and transients are used which is known as wavelet transform. An advanced version of discrete wavelet transforms i.e. MODWT and EMD is used to decompose the signals and then the features are calculated for all. those features are ranked using relief F ranker algorithm and finally the classification is done using SVM and random forest classifier. The ranking results are shown in table 4 for MODWT and EMD. The signals are decomposed to level 5 and the forward technique (first taking only 1$^{st}$ level decomposed signal than two levels and so on) is used to calculate the highest accuracy. It can be concluded that random forest is giving the higher accuracy as compare to SVM as shown in table 4. The higher accuracy is given by taking 3 level decomposition for both wavelet transform. MODWT is giving higher accuracy than EMD. Also the highest accuracy achieved is by using 3 level MODWT decomposed signals with relief F ranker algorithm using random forest classifier. So, it is suggested to use this combination to achieve the best accuracy for fault classification of bevel gear box vibration.

So, it would be better to use wavelet transform along with ranker algorithm to get the best classification for the fault diagnosis.

## REFERANCE


[1] Anand Parey, Amandeep Singh, 2019, Gearbox fault diagnosis using acoustic signals, continuous wavelet transforms and adaptive neuro-fuzzy inference system: Applied Acoustics, volume 147, pages 133–140

[2] N. Saravanan, V. N. S. Kumar Siddabattuni, K. I. Ramachandran, 2010,Fault diagnosis of spur bevel gearbox using artificial neural network and proximal support vector machine:Applied soft computing Volume 10, Issue 1, Pages 344-360.

[3] S. Khan, T. Yairi, 2018, A review on the application of deep learning in system health management, Mech. Syst. Signal Process. Volume 107 pages 241–265.

[4] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R.X. Gao, 2019, Deep learning and its applications to machine health monitoring, Mech. Syst. Signal Process. Volume 115.pages 213–237.

[5] Ruonan Liu, Boyuan Yang, Enrico Zio, Xuefeng Chen, 2018, Artificial intelligence for fault diagnosis of rotating machinery: A review: Mechanical Systems and Signal Processing 108, 33–47

[6] L. Duan, M. Xie, J. Wang, T. Bai, 2018, Deep learning enabled intelligent fault diagnosis, Overview and applications, J. Intell. Fuzzy Syst. 35, 5771–5784.

[7] Jaehyun Kima, Heesu Kima, Subin Huha, Jinho Lee, Kiyoung Choi, 2018, Deep neural networks with weighted spikes: Neurocomputing 311, 373–386.

[8] Ch. Ravi Sekhar1, Minal2, and E. Madhu: Mode Choice Analysis Using Random Forrest Decision Trees: 11th Transportation Planning and Implementation Methodologies for Developing Countries, TPMDC 2014, pages, 10-12, Mumbai, India

[10] P. K. Kankar, Satish C Sharma, S.P. Harsha, 2011, Fault diagnosis of ball bearing using machine learning methods, Expert systems with applications, Volume 38, Issue 3, Pages 1876-1886

[11] Kane P.V., Andhare A.B., 2019, End of assembly line gearbox fault diagnosis using artificial neural network and support vector machine: International journal of acoustics and vibrations, vole 24, issue 1, pages 68-84

[12] T. Kathirvalava kumar, S. Jeyaseeli Subavathi, 2019, Neurocomputing, Neighbourhood based modifiedbackpropagation algorithm using adaptive learning parameters. Volume 72, Issues 16–18, Pages 3915-3921.

[13] Md. Zahangir Alam, M. Saifur Rahman, M. Sohel Rahman, 2019, A Random Forest based predictor for medical data classification using feature ranking, Informatics in Medicine Unlocked, Volume 15, Article 100180.

[14] Ethem Alpaydın, 2004. Introduction to Machine Learning, Third Edition. The MIT Press Cambridge, Massachusetts London, England.

[15] Asha Gowda Karegowda, A. S. Manjunath & M.A. Jayaram, 2010, Comparative study of attribute selection using gain ratio and correlation-based feature selection: International Journal of Information Technology and Knowledge Management, Volume 2, No. 2, pp. 271-277

[16] Mariela Cerrada, Grover Zurita, Diego Cabrera, René-Vinicio Sánchez, Mariano Artés,ChanLi, 2016,Fault diagnosis in spur gears based on genetic algorithm and randomforest: Mechanical Systems and Signal Processing pages 70- 71, volume 87–103

[17] Jian-Da Wu, Chuang-Chin Hsu, 2009, Fault gear identification and classification using discrete wavelet transform and adaptive neuro-fuzzy inference: expert system and applications: Volume 36, Issue 2, Part 2, Pages 3785-3794.

[18] N. Saravanan, K.I. Ramachandran, 2014, Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN) Expert Systems with Applications, Volume 37, pages 4168-4181.

[19] Gang Cheng, Xihui Chen ⇑, Hongyu Li, Peng Li, Houguang Liu, 2018,Study on planetary gear fault diagnosis based on entropy feature fusion of ensemble empirical mode decomposition: Measurement, volume 9, pages 140–154.

[20] V. Vakhariaa,V.K. Guptab, 2014,Bearing Fault Diagnosis Using Feature Ranking Methods andFault Identification Algorithms: 12th International Conference on Vibration Problems, ICOVP 2015, Procedia Engineering volume 144, pages 343 – 350.

[21] Yu Yang, Yigang He, Junsheng Cheng, Dejie Yu, 2011, A gear fault diagnosis using Hilbert spectrum based on MODWPT and a comparison with EMD approach: Measurement, volume 42, pages 542–551.

[22] Thoranin Sujjaviriyasup, 2017, A new class of MODWT-SVM-DE hybrid model emphasizing on simplification structure in data pre-processing: A case study of annual electricity consumptions, Applied Soft Computing, volume 54, pages 150– 163.

[23] Ryan J. Urbanowicz, Melissa Meeker, William La Cava, Randal S. Olson, Jason H. Moore, 2018, Relief-based feature selection: Introduction and review, Journal of Biomedical Informatics, volume 85, pages 189–203.

[24] Amel Abdoullah Ahmed Dghais, Mohd Tahir Ismail, 2013,A Comparative Study between Discrete Wavelet Transform and Maximal Overlap Discrete Wavelet Transform for Testing Stationarity: International Journal of Mathematical Science and Engineering Vol:7 No:12.

[25] R. Liu, B. Yang, E. Zio, X. Chen, 2018, Artificial intelligence for fault diagnosis of rotating machinery: A review, Mech. Syst. Signal Process. Volume 108, pages33–47.

[26] Sungho Park, Seokgoo Kim, Joo-Ho Choi, 2018, Gear fault diagnosis using transmission error and ensemble empirical mode decomposition: Mechanical Systems and Signal Processing, volume 108, pages 262–275.

[27] Yaguo Lei, Bin Yang, Xinwei Jiang, Feng Jia, Naipeng Li, Asoke K. Nandi, 2020, Applications of machine learning to machine fault diagnosis:A review and roadmap:Mechanical Systems and Signal Processing volume 138.

[28] Zhiqiang Tong, Gouhei Tanaka, 2015, A pruning method based on weight variation information for feedforward neural network: IFAC-papers online volume 48, issue 18, pages 221-226.