FAULT DETECTION IN BALL BEARINGS : A MACHINE LEARNING APPROACH

A Project Report Submitted

by

BATCHU MADHU SRI KIRAN,

(1901ME15)

CHARANTHU MEENAKSHI,

(1901ME17)

In Partial Fulfilment

of the Requirements for the award of the degree

BACHELOR OF TECHNOLOGY

DEPARTMENT OF MECHANICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY PATNA

MAY 2023

THESIS CERTIFICATE

This is to certify that the thesis titled **Fault detection in ball bearings : A machine learning approach**, submitted by **Batchu Madhu Sri Kiran and Charanthu Meenakshi**, to the Indian Institute of Technology, Patna, for the award of the degree of **Bachelor of Technology**, is a bona fide record of the research work done by him under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. Chiranjit Sarkar
Supervisor
Assistant Professor
Dept. of Mechanical Engineering
IIT-Patna, 800 013

Dr. Mayank Tiwari
Supervisor
Professor
Dept. of Mechanical Engineering
IIT-Patna, 800 013

Place: Patna

Date: 6th May 2023

DECLARATION

Name of the students: **Batchu Madhu Sri Kiran, Charanthu Meenakshi**

Signature of the students:

B.Madhusrikiran C. Meenakshi

B.Tech Project Title: **FAULT DETECTION IN BALL BEARINGS : A MACHINE LEARNING APPROACH**

This is to certify that Mr. Batchu Madhu Sri Kiran and Ms. Charanthu Meenakshi

1. have sincerely worked on their project,
2. have contacted me regularly to update on the progress of the assigned project,
3. have received my comments on the preliminary version of the report and presentation and will address those prior to final presentation,
4. May be allowed to present the project before the department.

Name of Project Guide

Dr. Chiranjit Sarkar

Signature of Project Guide

Date : 06.05.2023

ACKNOWLEDGEMENTS

We would like to acknowledge the sincere efforts and valuable time given by our Professor in-charge Dr. Chiranjit Sarkar (Department of Mechanical Engineering, IIT Patna) and Dr. Mayank Tiwari (Department of Mechanical Engineering, IIT Patna) who have helped us in understanding the basic concepts required for the project and contributed immensely thereafter as well. Their valuable guidance and feedback had helped us immensely during the project. We would like to acknowledge the Mechanical Department of IIT Patna for providing us with an opportunity to broaden the horizon of our knowledge.

BATCHU MADHU SRI KIRAN
CHARANTHU MEENAKSHI

ABSTRACT

The detection and diagnosis of Ball bearing faults is very crucial for prevention of rotating machine breakdowns. Here in this Project we are using Back propagation to train and predict fault diagnosis. The Vibration response is taken from a case study from Case Western Reserve University. This data consisted of data samples collected at different frequencies at the driver end and fan ends of the motor. This consisted of both normal data and faulty data collected at different types of faults. We have prepared a custom data set from this data for our training. We have used the Backpropagation Algorithm with Stochastic Gradient Descent for training the model and K-fold Cross Validation technique for validating the model so that it can give better estimation of the results with lower bias.

TABLE OF CONTENTS

DECLARATION **ii**

ACKNOWLEDGEMENTS **iii**

ABSTRACT **iv**

1 INTRODUCTION **1**

1.1 Background and Motivation 1

1.2 Objective 1

1.3 Overview 2

2 LITERATURE REVIEW **3**

2.1 Anjali Thorat, Priyanka Walunj(Apr, 2022) 3

2.2 H.Saruhan, S.SarÕdemir, A.Çiçek and I.Uygur(Jun, 2014) 5

3 MODEL AND APPROACH **8**

3.1 Back propagation Algorithm 8

3.2 Backpropagation Algorithm With Stochastic Gradient Descent . . . 9

3.3 K-fold Cross Validation 12

4 TRAINING AND RESULTS **14**

4.1 Description of data records 14

4.2 Pre-processing 14

4.3 Training dataset 15

4.4 Results 20

5 CONCLUSION **25**

5.1 Conclusion 25

5.2 Future Scope 26

6 REFERENCES **27**

CHAPTER 1

INTRODUCTION

1.1 Background and Motivation

The most common type of movement in machines is rotation. The rotation allows it to reach a high-speed and high-power motion in limited space. A major part of rotating machineries around the world rely on Rolling element bearings (REBs). These provide rotating freedom along with transmitting load. As the contribution for the REBs in rotary machines increases, the demand for REBs also increases. These rotary machines have many complex parts, even failure of a single part can lead to failure of the machine. In some cases there may be catastrophic failures which may cause heavy damage to machinery as well as the humans who are working along with the machinery. So, to minimize the cost and danger of damage and rundown time,diagnosis of a bearing is a very important engineering task.

1.2 Objective

Given vibrational data-samples of rolling bearing elements calculated in different speeds, using machine learning we made a back propagation model and validated it through K-fold cross validation method. Deterministic methods are connected with vibration spectrum analysis such as spectral analysis, Fourier transform, wavelet analysis, etc. Most of the stochastic methods are connected with machine learning and artificial neural networks (ANNs). These deterministic methods are used to prepare a data set using filters and transforms and the machine learning methods are applied to solve a fault diagnosis problem. The failures in REB can be caused by many types of defects like:

1. Misalignment
2. Cracked shaft
3. Unbalanced shaft
4. Shaft rubbing

5. Bearing defects
6. High operating speed, large load

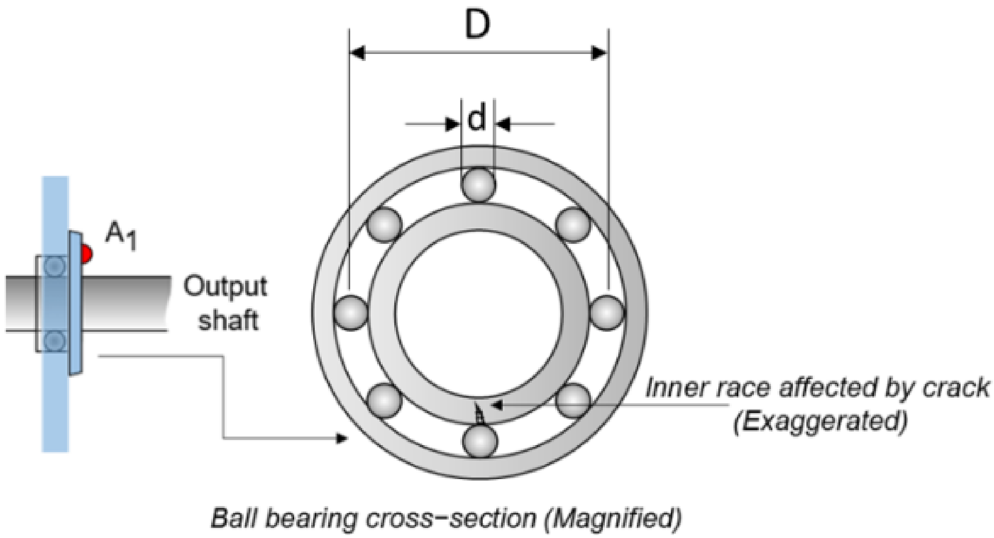


Figure 1.1: An image of Ball bearing cross section with inner race defect

1.3 Overview

We have taken data in two stages for our project. First the data sets required were taken from a case study from, Case Western Reserve University. The second is data directly from an industry. In the first case, the case study provided the ball bearing test data for normal and faulty bearings. experiments were performed utilizing a 2 hp Reliance Electric motor, and acceleration data was gathered from locations both near to end remote from the motor bearings. To simulate different degrees of wear and tear, the motor bearings were seeded with faults using electro-discharge machining (EDM). Faults from 0.007, 0.014, 0.021, 0.028 inches in diameter were introduced separately at the inner race, rolling element (i.e. ball) and outer race. Subsequently, the faulty bearings were reinstalled into the test motor, and vibration data was recorded for motor loads varying from 0-3 horsepower, corresponding to motor speeds ranging from 1797, 1772, 1750 and 1730 RPM. Data was collected for normal bearings, single-point drive end and fan end defects.

CHAPTER 2

LITERATURE REVIEW

Here we have a brief review of pre-existing works done on Rolling Element Bearing Defects.

2.1 Anjali Thorat, Priyanka Walunj(Apr, 2022)

In their work, they have taken the vibrational data from Rotor - stator and preprocessed it before performing time and frequency domain analysis of data. They developed a model to identify where peak fluctuations occur from the data.

The vibration data from different cases with each faulty part and remaining good parts is acquired. This data is preprocessed via cleansing so we can have data with lesser noise. Then this is normalized by mean centering and scaling the data. Then this data is viewed as smooth data which is visible with lesser noise.

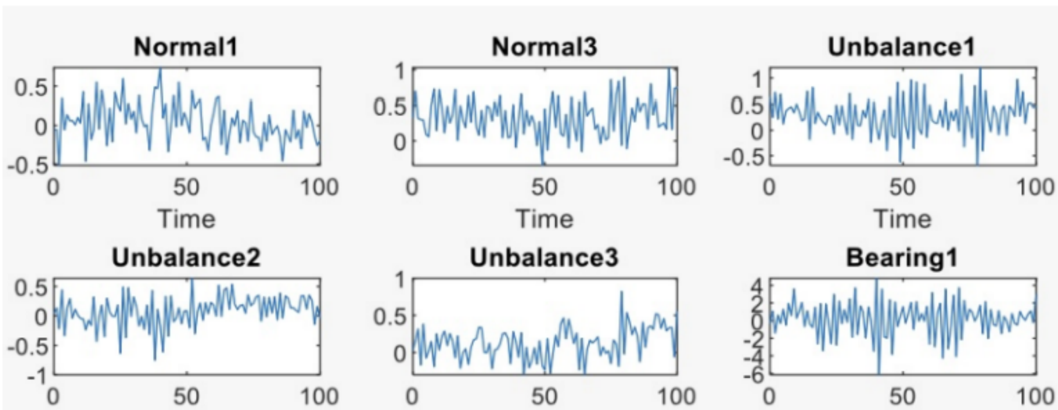


Figure 2.1: An image showing vibration data with noise

Then this smooth data is converted using fourier transform and peaks are observed. From the below figures even after using smooth data we can see so much noise in the figure 3. This represents higher noise or higher frequency of fluctuations which indicates system defects.

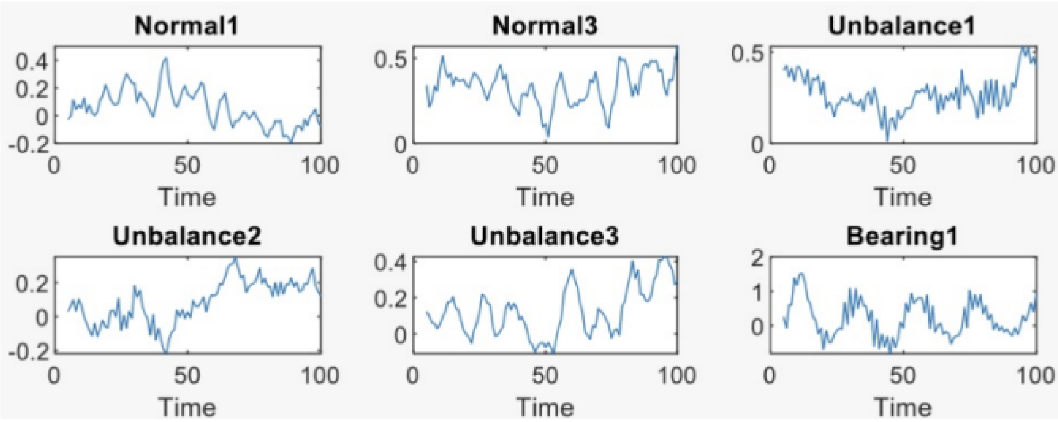


Figure 2.2: An image showing vibration data with less noise

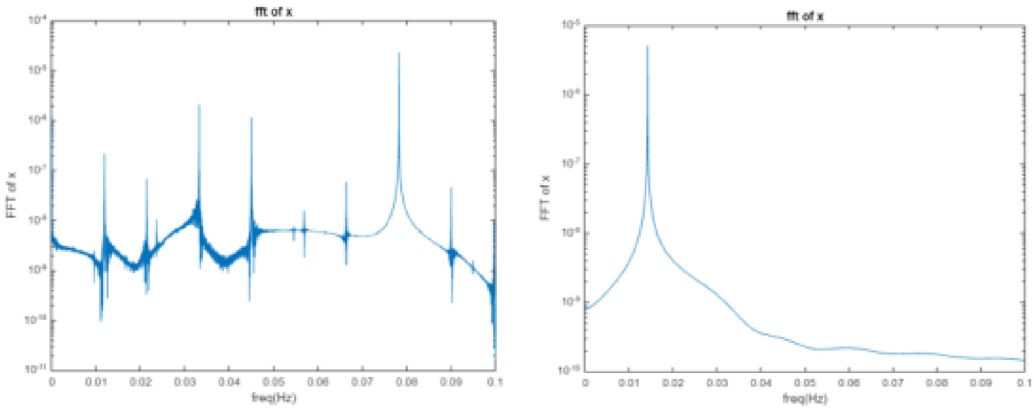


Figure 2.3: An image showing vibration data with less noise

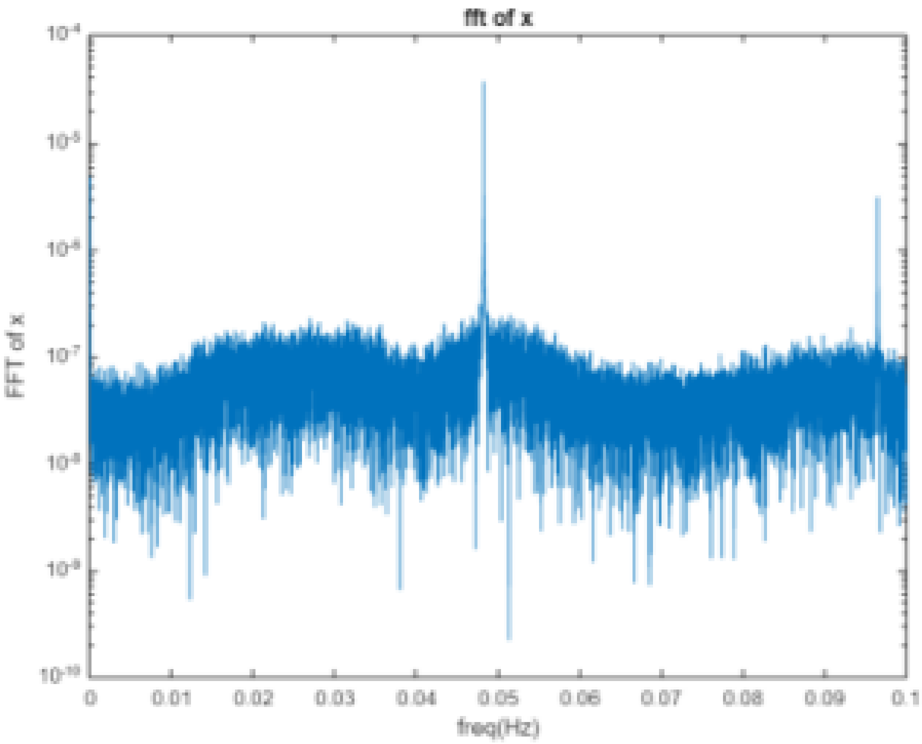


Figure 2.4: An image showing vibration data with less noise

Feature extraction is performed by using time-domain analysis on data like variance, standard deviation, mean, median, kurtosis.

2.2 H.Saruhan, S.SarÕdemir, A.Çiçek and I.Uygur(Jun, 2014)

In this study, the vibration signals of bearings with different types, including normal bearings (GOOD) and inner raceway (INNER) , outer raceway (OUTER), ball (BALL) and combination of these defects (COMB), and to determine the unique characteristics of these signals that can be used to detect the type and severity of the defect. The study uses experimental data collected from a test rig equipped with vibration sensors to measure the vibration signals of the bearings under different operating conditions. normal state condition bearing and deliberately defected bearings were tested under different shaft running speeds (17 Hz, 25 Hz, 33 Hz, and 41 Hz) with two load levels(0 kg, 5.04 kg).

Each bearing element has a characteristic defect frequency that depends on mechanical dimensions of the bearing.From the specified design of the system, We find Ball pass frequency for outer race (BPFO) , Ball pass frequency for inner race (BPFI), and Ball spin frequency (BSF), which are as follows.

$$BPFO = \frac{Nb}{2}\left(1 - \left(\frac{Nb}{dp}\right)\cos\alpha\right) \tag{2.1}$$

$$BPFI = \frac{Nb}{2}\left(1 + \left(\frac{Nb}{dp}\right)\cos\alpha\right) \tag{2.2}$$

$$BSF = \frac{dp}{2db}\left(\left[\left(\frac{db}{dp}\right)\cos\alpha\right]^2\right) \tag{2.3}$$

This BPFO, BPFI, BSF are taken for every shaft speed along with its harmonics. This is collected for every case of defect (defected inner race / outer race / ball / combi-

Table 2.1: Table showing the specifications of ball bearing system

Bearing specifications	**MB ER - 12K**
Outer diameter , D(mm)	47.00016
Inner diameter , d(mm)	19.05
Pitch diameter dp(mm)	33.50006
Ball diameter , db (mm)	7.9375
Outer ring width, B(mm)	15.8496
Number of balls, Nb	8
Contact angle (alpha) (degrees)	0

nation of defects) with and without loads. The data collected during running tests were observed, analyzed and presented the defect frequencies. From the data collected, the maximum amplitude and its respective frequency are compared in various cases. This study analyzed 40 different test cases of rolling element bearings, including 5 bearings with different health conditions, 4 running speeds, and 2 load levels. In each bearing case spectrum values and harmonics of three characteristic frequencies: BPFO, BPFI, and BSF are collected. The results were presented with the bearing defect harmonics presented for frequencies ranging from 1 to 13 times the shaft running speed rate.

Running Speed 41 Hz			Corresponding Frequency (xHz) x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13
Calculated	With Defect	BPFO	125.1	250.2	375.4	500.5	625.7	750.8	875.9	1001.1	1126.2	1251.4	1376.5	1501.6	1626.8
		BPFI	202.8	405.7	608.5	811.4	1014.2	1217.1	1419.9	1622.8	1825.7	2028.5	2231.4	2434.2	2637.1
		BSF	81.6	163.3	244.9	326.6	408.2	489.9	571.6	653.2	734.9	816.5	898.2	979.9	1061.5
	GOOD	BPFO	2.08	4.16	6.24	8.33	10.41	12.49	14.57	16.66	18.74	20.82	22.91	24.99	27.07
		BPFI	3.38	6.76	10.14	13.53	16.91	20.29	23.68	27.06	30.44	33.83	37.21	40.60	43.98
		BSF	1.36	2.72	4.08	5.44	6.80	8.16	9.52	10.88	12.25	13.61	14.97	16.33	17.69
Experiment (two disks without loader)	OUTER	BPFO	121.8	243.7	365.6	487.5			853.1	975	1096.8		1340.6	1462.5	1584.3
		BPFI	203.1	406.2		812.5	1015.6		1421.8	1625				2437.5	2640.6
		BSF	81.2	162.5	243.7	325.0	406.2	487.5	568.7	650.0		812.5		975.0	
	INNER	BPFO		243.7		487.5		731.2							
		BPFI				812.5	1015.6		1421.8	1625.0		2031.2	2234.3		2640.6
		BSF	81.25	162.5	243.7	325.0		487.5			731.2	812.5		975.0	1056.2
	BALL	BPFO	121.8	243.7	365.6		609.3	731.2	853.1			1218.7	1340.6	1462.5	1584.3
		BPFI	203.1	406.2	609.3			1218.7	1421.8	1625.0	1828.1	2031.2	2234.5	2437.5	2640.6
		BSF	81.2	162.5	243.7		406.2		568.7	650.0	731.2		893.7		
	COMB	BPFO	121.8	243.7	365.6	487.5	609.3	731.2	853.1	975.0	1096.8	1218.7		1462	1584
		BPFI	203.1	406.2	609.3		1015.6	1218.7	1421.8	1625.0	1828.1	2031.2	2234.3	2437.5	
		BSF	81.2	162.5	243.7	325.0	406.5	487.5	568.7		731.25	812.5	893.7	975.0	
Experiment (two disks with loader)	OUTER	BPFO	121.8	243.7	365.6	487.5	609.3	731.2	853.1	975.0	1096.8	1218.7	1340.6		
		BPFI	203.1	406.2	609.3			1218.7							2640.6
		BSF	81.2	162.5	243.7	325.0	406.2	487.5	568.7	650.0	731.2		893.7	975.0	
	INNER	BPFO	121.8	243.7		487.0	609.3	731.2	853.1	975.0		1218.7	1340.6		
		BPFI	203.1	406.2	609.3			1218.7			1828.1	2031.2			
		BSF	81.2		243.7	325.0	406.2	487.5	568.7	650.0			893.7	975.0	1056.2
	BALL	BPFO	121.8	243.7	365.6	487.5	609.3					1218.7	1340.6	1462.5	1584.3
		BPFI	203.1	406.2	609.3		1015.6	1218.7		1625.0	1828.1		2237.5	2437.5	2640.6
		BSF	81.2	162.5	243.7	325.0	406.2	487.5	568.7						1056.2
	COMB	BPFO	121.8	243.7	365.6	487.5	609.3	731.2		975.0		1218.7		1462.5	
		BPFI	203.1	406.2	609.3			1218.7				2031.2	2234.3		2640.6
		BSF	81.25	162.5	243.7	325.0	406.2	487.5	568.7	650.0	731.2			975.0	

Figure 2.5: The harmonics of defected bearing elements frequencies for running speed 41 Hz

From the data, we found that as the shaft running speed increases it will raise the bearing defect frequencies and decrease the bearing life. Also the defect frequencies with increased amplitude of harmonics appear in the vibration spectrum data. The frequency and amplitude spectrum graphs give the frequencies with higher fluctuations.

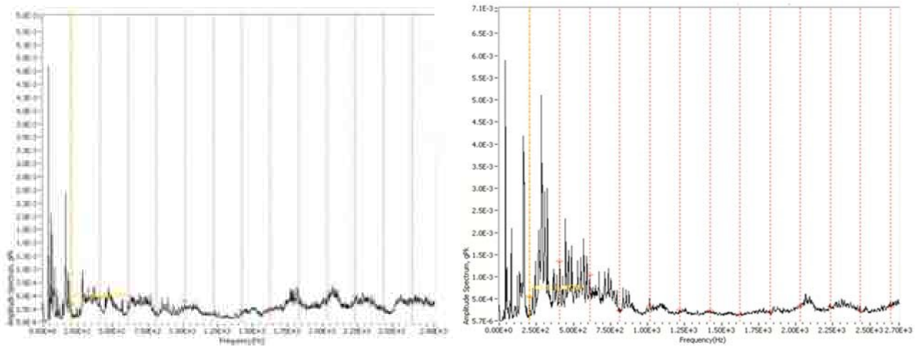


Figure 2.6: Amplitude spectrum of the COMB for the BPFI at 41Hz without loader and with loader

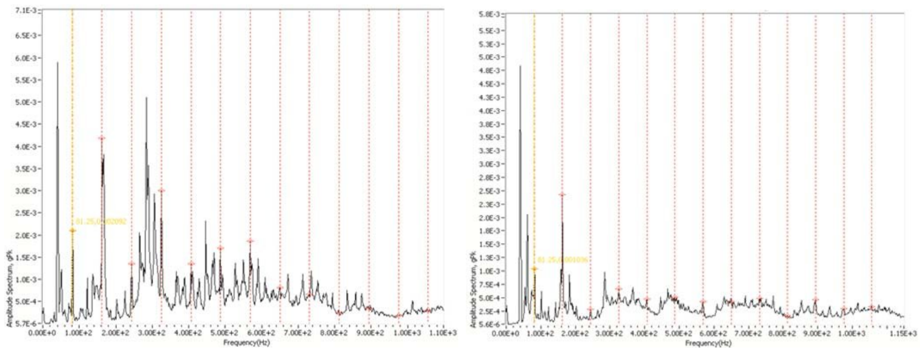


Figure 2.7: Amplitude spectrum of the COMB for the BSF at 41Hz with loader and without loader

CHAPTER 3

MODEL AND APPROACH

We have used Back propagation

3.1 Back propagation Algorithm

Backpropagation is a widely used algorithm in the field of artificial neural networks, and it is used to train multi-layered feedforward neural networks. The algorithm works by iteratively adjusting the weights of the connections between neurons, in order to minimize the difference between the actual output of the network and the desired output. Backpropagation is a type of supervised learning, and it requires a dataset of input-output pairs to train the network.

Backpropagation calculates the gradient of a loss function with respect to all weights in the network. This method is commonly referred to as "backward propagation of errors." The algorithm efficiently computes one layer at a time, unlike direct computation, and applies the chain rule to calculate the gradient of the loss function for a single weight. This algorithm generalizes the computation in the delta rule and is a standard method for training artificial neural networks.

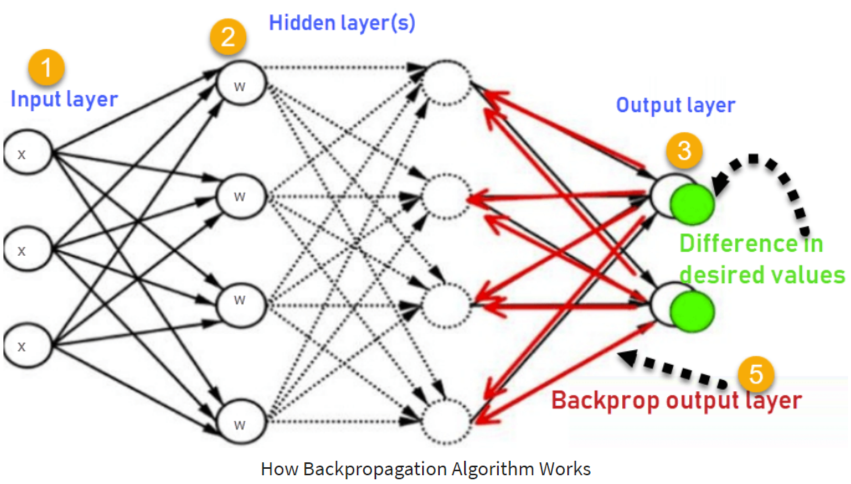


Figure 3.1: Back propagation in neural network

This is how the Backpropagation works:

1. Inputs X, arrive through the preconnected path
2. Input is modeled using real weights W. The weights are usually randomly selected.
3. Calculate the output for every neuron from the input layer, to the hidden layers, to the output layer.
4. Calculate the error in the outputs

ErrorB= Actual Output – Desired Output

5. Travel back from the output layer to the hidden layer to adjust the weights such that the error is decreased.
6. Keep repeating the process until the desired output is achieved

3.2 Backpropagation Algorithm With Stochastic Gradient Descent

Backpropagation Algorithm with Stochastic Gradient Descent (SGD) is a popular supervised learning algorithm used in artificial neural networks. It is an extension of the standard Backpropagation Algorithm, which updates the weights of the neural network based on the error calculated from the entire training set.

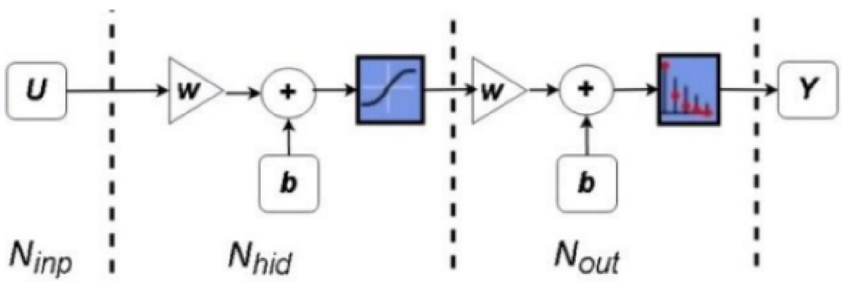


Figure 3.2: backpropagation of a input signal

STOCHASTIC GRADIENT DESCENT:

A learning rule is an algorithm for updating the weights of a network in order to achieve a particular goal. One particular common goal is to minimize an error function associated with the network, also known as cost function. SGD is a variant of gradient descent that updates the weights of the network incrementally after processing each training example, rather than updating the weights after processing the entire training set. In addition, the random shuffling of the training examples during each epoch can help SGD escape from local minima.

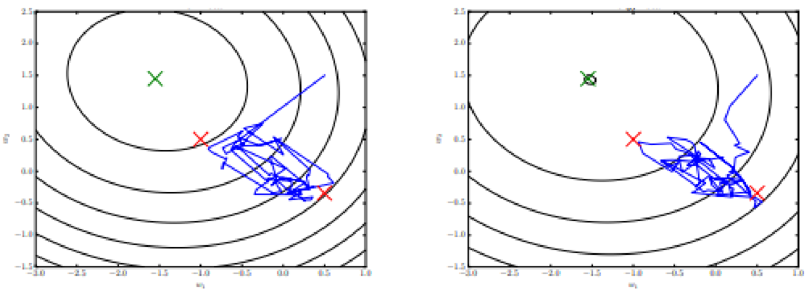


Figure 3.3: Total error surfaces for a set observations x, t that consists of two subsets, each with a single solution in weight space with zero MSE

SGD is important in using the backpropagation algorithm because it: Updates the weights incrementally, making it computationally efficient Can escape local minima due to random shuffling of training examples Can generalize better than batch gradient descent, especially with non-convex loss functions.

Consider a linear neuron with inputs xi and weights w. Assume a linear activation function g(hi) = hi, so the actual output of the neuron is $y_i = g(h_i) = h_i = w^T x_i$.

$$\mathbf{w} := \mathbf{w} - \eta \sum_{i=1}^{n} \nabla Q_i(\mathbf{w}) \tag{3.1}$$

where, $Q_i(w)$ - error corresponding to the i-th observation,

n - constant learning rate.

MEAN SQUARE ERROR (MSE): A popular loss function used to measure the difference between the predicted and actual output values of a neural network.

$$E(\mathbf{w}) = \sum_{i=1}^{n} E_i(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^{n} \left(t_i - \mathbf{w}^\top \mathbf{x}_i\right)^2 \tag{3.2}$$

where, t_i - expected outcome of the network for the i-th observation.

FEED-FORWARD: The input x is fed into the input layer, and forward propagates through the network. Store the output of each neuron i as oi.

ERROR AT THE OUTPUT LAYER: The error delta(j) of the j-th neuron in the output layer is computed as

$$o_i\delta_j = \frac{\partial Q}{\partial w_{ij}} \tag{3.3}$$

where W_{ij} - weight from the i-th neuron to the j-th neuron
o_i - output of the i-th neuron, or conversely, the input into the j-th

BACKPROPAGATE TO THE HIDDEN LAYERS: Since the expected output of neurons in the hidden layers is not known, it is instead estimated by errors in the layer closer to the output . If W_{hq} is the connection from the h-th to the q-th neuron in the hidden layer, then the error is estimated as

$$\delta_h = \sum_q w_{hq}\delta_q \tag{3.4}$$

WEIGHT UPDATES: Update the weights of each connection as

$$w_{ij} := w_{ij} - \eta o_i \delta_j \tag{3.5}$$

3.3 K-fold Cross Validation

Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample.

The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into. As such, the procedure is often called k-fold cross-validation. When a specific value for k is chosen, it may be used in place of k in the reference to the model, such as k=10 becoming 10-fold cross-validation.

Cross-validation is primarily used in applied machine learning to estimate the skill of a machine learning model on unseen data. That is, to use a limited sample in order to estimate how the model is expected to perform in general when used to make predictions on data not used during the training of the model.

It is a popular method because it is simple to understand and because it generally results in a less biased or less optimistic estimate of the model skill than other methods, such as a simple train/test split.

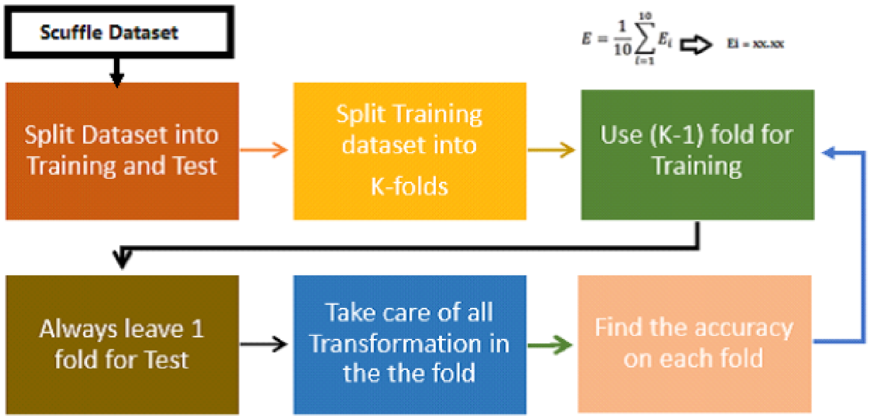


Figure 3.4: Cycle of K-fold cross validation

HOW K-FOLD CROSS-VALIDATION WORKS?

1. Shuffle the data randomly to ensure that the data is unbiased and does not follow any pattern
2. Split the data into k-folds or partitions, with one fold kept aside for validation
3. Train the model on the remaining k-1 folds
4. Validate the model's performance on the fold kept aside for validation and calculate

the accuracy or error

5. Repeat steps 2 to 4 for k times with each fold being used as validation data once

6. Calculate the average accuracy or error of the k models

Figure 3.5: five folds sectioning of training dataset

In our project, we have put together both the Stochastic Backpropagation algorithm and K-fold cross validation to train our model.

CHAPTER 4

TRAINING AND RESULTS

4.1 Description of data records

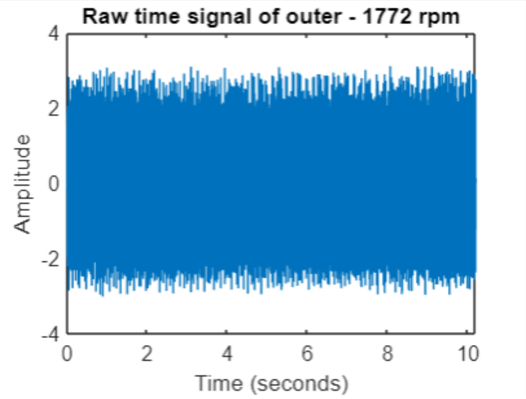
The datasets are classified into four categories based on the sample rate (12 or 48 kHz) and faulty bearing location, namely 48k baseline, 12k drive end fault, 48k drive end fault, and 12k fan end fault. Each category includes data sets for normal baseline, rolling element (ball) faults and inner and outer race faults. The outer race faults are grouped into three categories based on their fault position relative to the load zone: 'centred' (fault in the 6.00 o'clock position), 'orthogonal' (3.00 o'clock), and 'opposite' (12.00 o'clock). Furthermore, the data sets are classified by fault size (0.007 to 0.028 in.) and motor load (0-3 hp) corresponding to motor speeds (1730-1797 rpm).

4.2 Pre-processing

In the context of vibration analysis, pre-processing is necessary to extract and analyze the relevant information contained in the vibration signals obtained from the machine system. In vibration analysis of bearings, the envelope spectrum is an important tool for detecting and diagnosing faults.

ENVELOPE ANALYSIS : The envelope spectrum is obtained by first applying a bandpass filter to the vibration signal to isolate the frequency range of interest, typically around the bearing fault frequencies. The filtered signal is then rectified to obtain the absolute value of the signal, which effectively removes any negative values. The resulting signal is then passed through a low-pass filter to remove the high-frequency components and obtain the envelope of the signal. The envelope spectrum is then obtained by taking the Fourier transform of the envelope signal. It involves two steps:

1. Data Resampling
2. Data Filtering (Band-pass and Low-pass filtering)



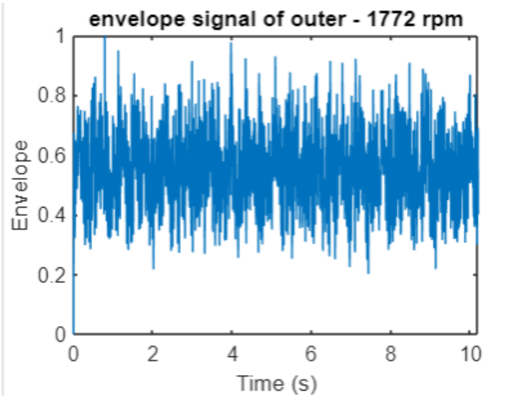


Figure 4.1: (a) raw time signal of outer raceway defected data at 1772 rpm, (b) envelope signal of outer raceway defected data at 1772 rpm

Data Resampling: This method is used to change the sampling rate of a vibration signal to match the requirements of the analysis method for more accurate envelope analysis. It also helps to reduce the computational complexity of the analysis.

Data Filtering : The raw vibration data were filtered using a bandpass filter to remove any unwanted frequencies. The filter cutoff frequencies were set to 100 Hz and 1 kHz to remove any low-frequency noise and high-frequency noise. The band-pass filter is used to remove the noise and extract the frequency content of interest that relates to the bearing faults. The low-pass filter is used to remove the high-frequency content in the signal that may not be relevant to the analysis.

4.3 Training dataset

For each RPM, each defected data set (outer race, inner race, ball bearing) consisted of around 1.2 lakhs points and normal data had 4.8 lakhs points each. We have performed Fast-Fourier transform on these data sets which is assumed to be sampled at a frequency of 12000 Hz. It then finds the top peaks in the magnitude spectrum, i.e., the frequencies with the highest amplitudes, which consisted of 1000 first peak data points. Upon doing the same for all 16 data sets-4 RPMs each having outer race defected, inner race defected, ball bearing defected and normal(4x4 = 16) i.e., 3 defect types and 1 normal at 4 different rpms which gives us 12 sets of faulty data and 4 sets of normal data in total. We have mixed the values in all the 16 datasets(each containing 1000 points) into one bigger dataset consisting total 16000 points. We used this data set for training our model and validating the model using this data.

FFT of those 16 data sets are:

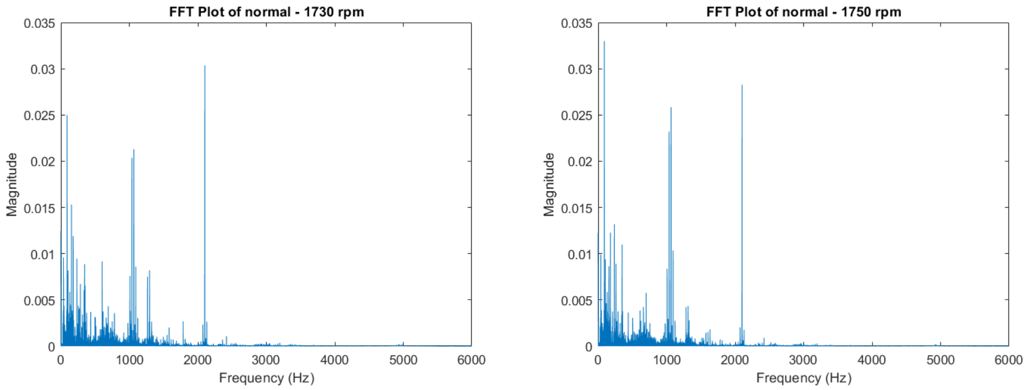


Figure 4.2: fft plot of normal bearings at 1730, 1750 rpm

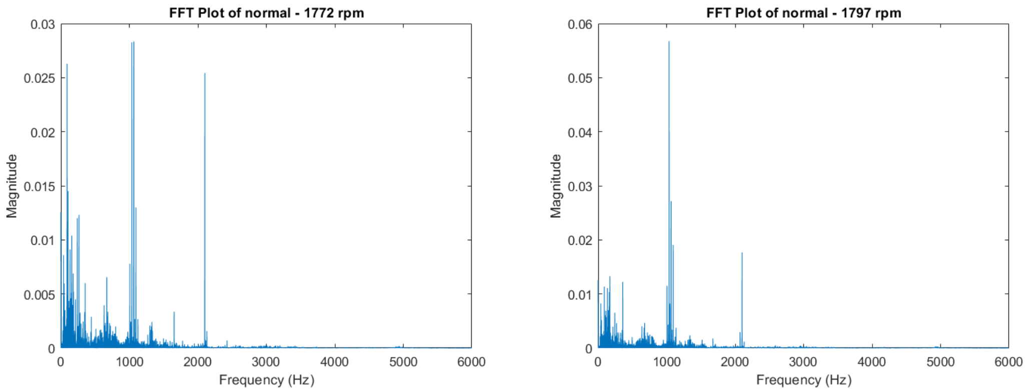


Figure 4.3: fft plot of normal bearings at 1772, 1797 rpm

We had Inner race, ball bearing, outer race defected data and normal data for 1797, 1772,1750,1730 rpm each. As each set is taken for 1000 peaks, the size of the training data set is 16,000 units of data.

So, Data consists of :-

12 x 1000 = 12000 data points of defected data(1,2,3 types each 4000 data).

4 x 1000 = 4000 data points of normal data (0) which is taken from different frequency spectrums at different RPMs(1730,1750,1772,1797).

Here in fig 4.10 the defect column,

0 - normal / no defect

1 - defect in ball bearing

2 - defect in inner race

3 - defect in outer race

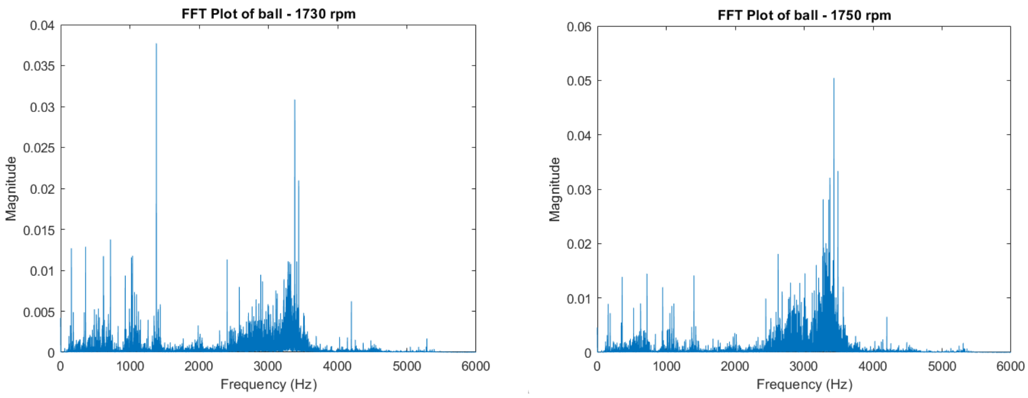


Figure 4.4: fft plot of ball defected bearings at 1730, 1750 rpm

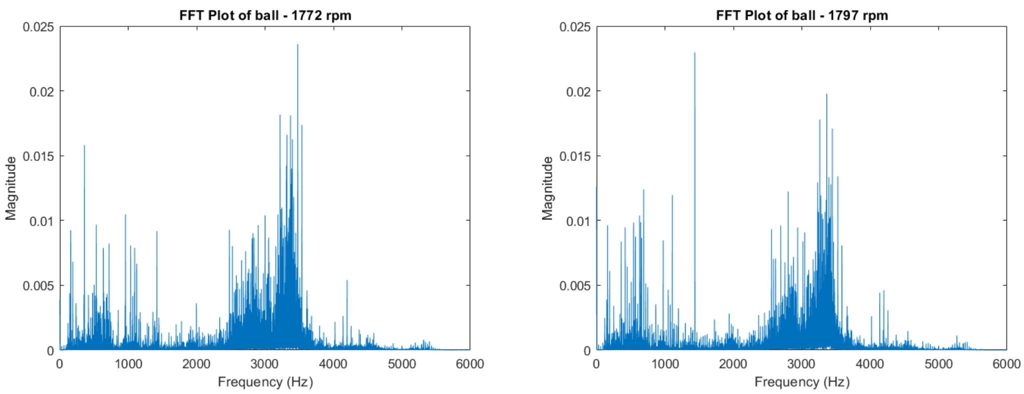


Figure 4.5: fft plot of ball defected bearings at 1772, 1797 rpm

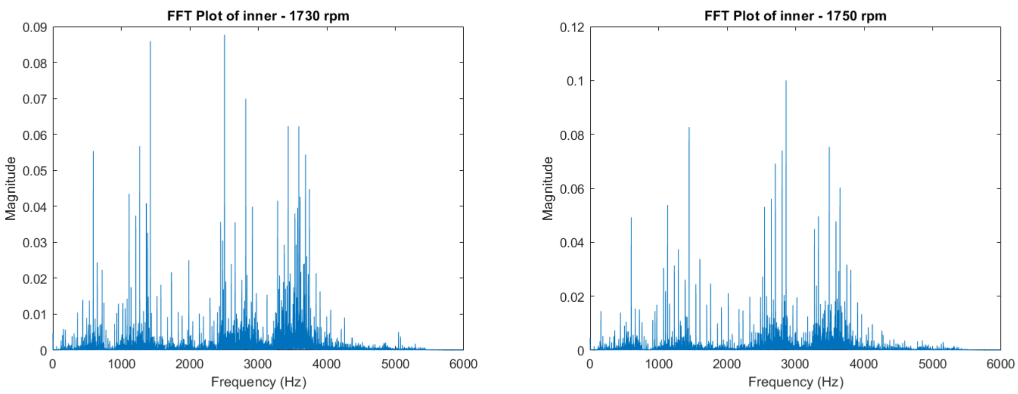


Figure 4.6: fft plot of inner raceway defected bearings at 1730, 1750 rpm

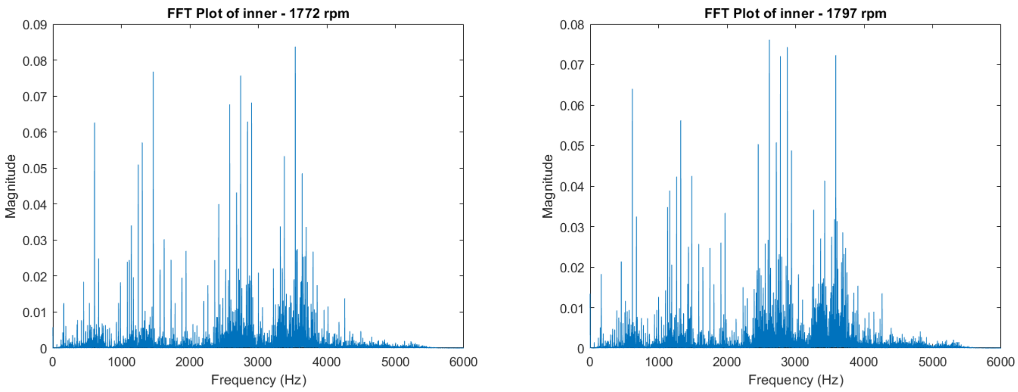


Figure 4.7: fft plot of inner raceway defected bearings at 1772, 1797 rpm

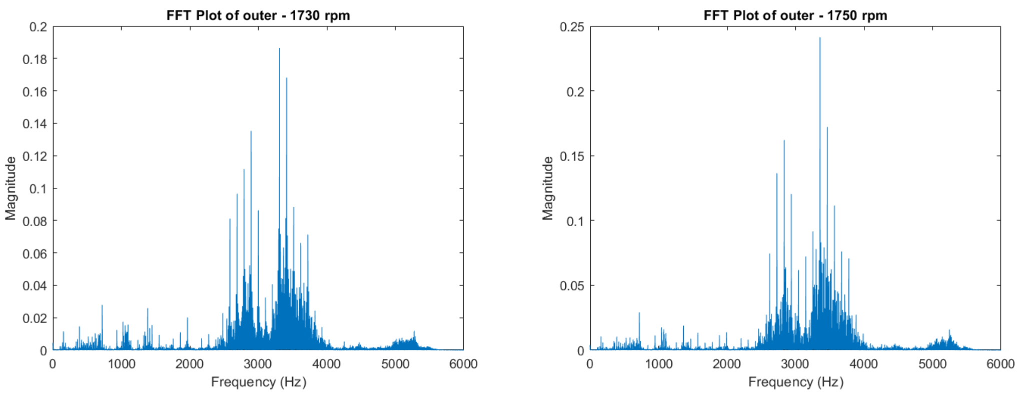


Figure 4.8: fft plot of outer raceway defected bearings at 1730, 1750 rpm

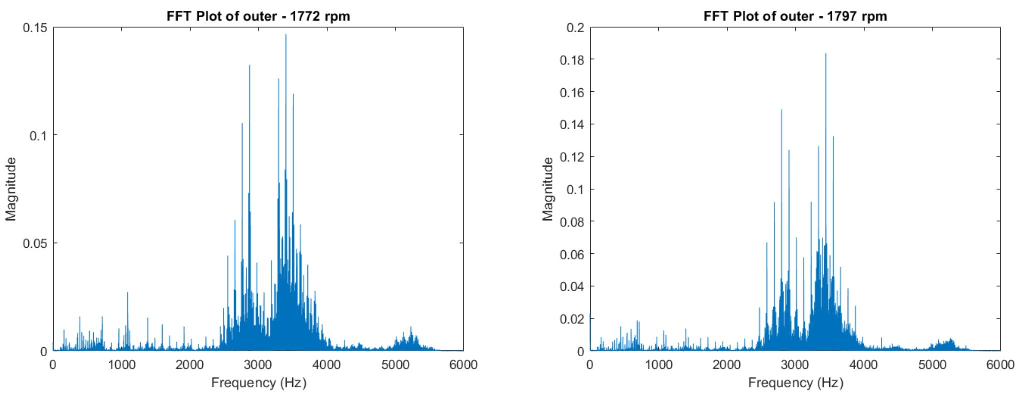


Figure 4.9: fft plot of inner raceway defected bearings at 1772, 1797 rpm

1	Frequency	Amplitude	rpm	defect
2	3338.671	0.002303	1772	1
3	3641.189	0.003229	1730	2
4	3347.187	0.002513	1797	1
5	2911.538	0.015664	1772	3
6	3574.985	0.005045	1772	2
7	3621.316	0.002906	1772	2
8	215.317	0.000699	1797	0
9	2935.987	0.011194	1797	3
10	2655.71	0.004568	1772	3
11	2924.677	0.001534	1750	1

Figure 4.10: first 10 points of training dataset

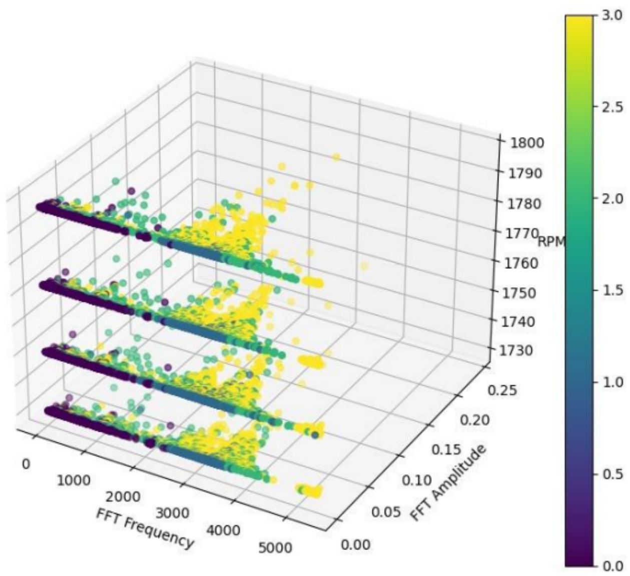


Figure 4.11: 3D visualization on training data

In the fig 4.11 we can see the training data represented in 3d showing the defect against frequency, magnitude(amplitude in image). This gives us better understanding of the nature of the data in training data.

We can clearly see at which rpm and frequency we were getting which type of defect. From this we can see that towards the 0 - 1600 frequency we have normal data without a defect. We find the ball bearing defect mostly in around 2000- 3500 frequency region, while we find the inner race defect can be spread across 0- 4000 region. But the Outer race defect seems a bit different but it also is observed across 2000- 4000 and around 5000 region.

4.4 Results

Upon using this data for training our model by taking different values of input parameters like number of k -folds, no of hidden layers in backpropagation neural network, Learning rate and no of iterations we were able to get different accuracies for our model. Manually changing the values, we were able to get accuracy of more than 70Given, the input parameters:
(n) k-fold = 10
no. of hidden layers = 10
learning rate = 0.02
no. of epochs = 500
we were able to find average accuracy of 71.6%.

To find a pattern between accuracy and dataset, we have taken a different approach. In this we were to train and test the model with different input datasets.
In first case, we have taken 1000 data points from each of 16 data sets,(The original data set which consisted of the top 1000 datapoints when put in descending order.) and trained and tested the data. In the second case, we have decided to take first 750 datapoints when all fft magnitudes are placed in descending order. Similarly, in the next case we have taken first 500 datapoints, in the next case the first 250 and in the next case we have taken first 100 datapoints.

Here are the accuracy results of K-fold cross validation after Stochastic backpropagation algorithm Given the input parameters are:

(n) k-fold = 5

no. of hidden layers = 10

learning rate = 0.02

no. of epochs = 200

The results are as follows:

Here Confusion matrix shows no of predicted against the actual values which helps in better understanding of which data is predicted better.

1. For 1000 data points from each data set = 16000 points in total
2. For 750 data points from each data set = 12000 points in total
3. For 500 data points from each data set = 8000 points in total
4. For 250 data points from each data set = 4000 points in total
5. For 100 data points from each data set = 1600 points in total

```
Scores: [72.34375, 72.15625, 72.21875, 73.25, 74.25]
Mean Accuracy: 72.844%
```

Figure 4.12: five folded accuracies and mean accuracy for 1000-peaks

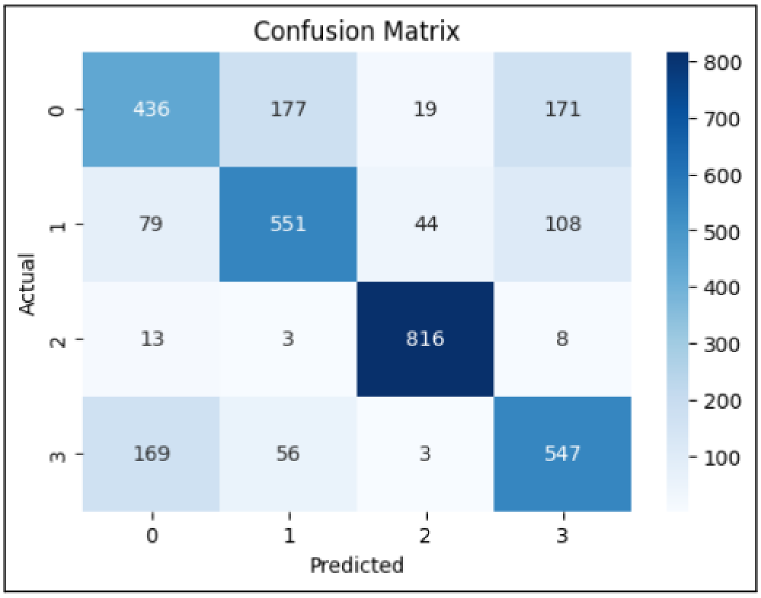


Figure 4.13: confusion matrix for 1000-peaks

```
Scores: [73.45833333333334, 72.20833333333333, 73.0, 73.875, 74.41666666666666]
Mean Accuracy: 73.392%
```

Figure 4.14: five folded accuracies and mean accuracy for 750-peaks

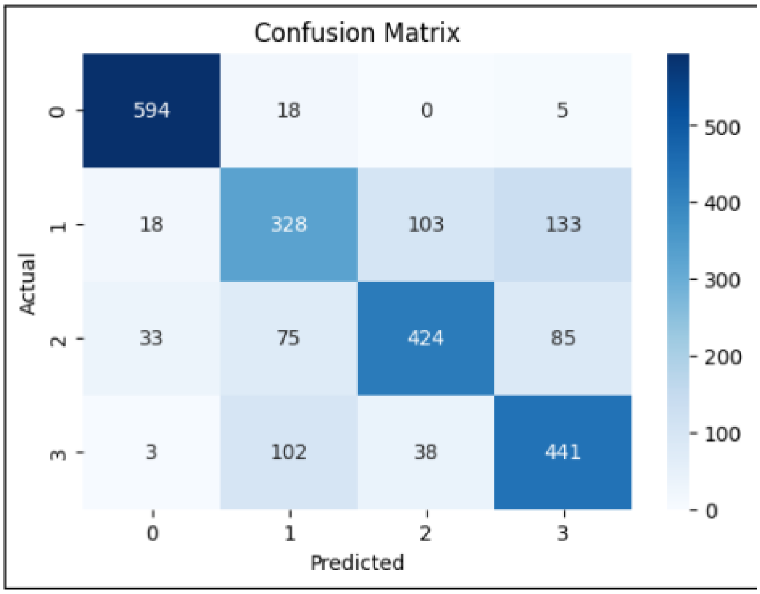


Figure 4.15: confusion matrix for 750-peaks

```
Scores: [73.5625, 73.5625, 73.1875, 73.875, 73.9375]
Mean Accuracy: 73.625%
```

Figure 4.16: five folded accuracies and mean accuracy for 500-peaks

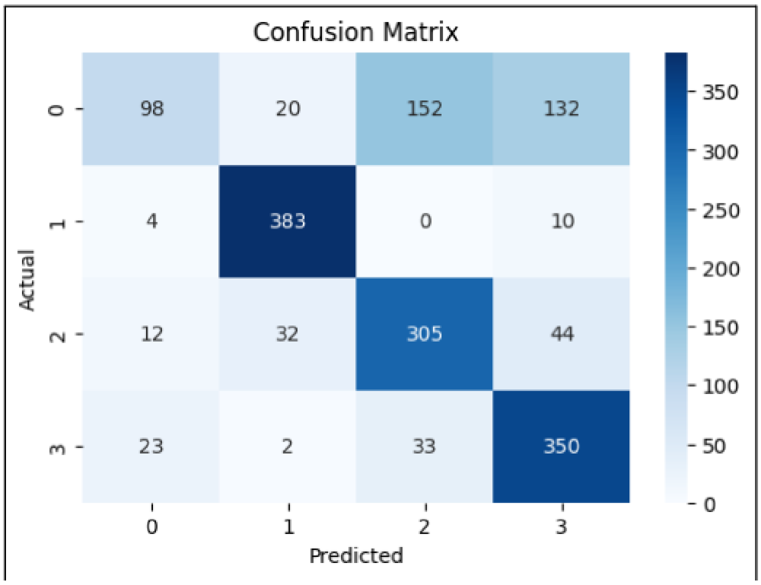


Figure 4.17: confusion matrix for 500-peaks

```
Scores: [76.25, 72.0, 76.0, 72.875, 74.75]
Mean Accuracy: 74.375%
```

Figure 4.18: five folded accuracies and mean accuracy for 250-peaks

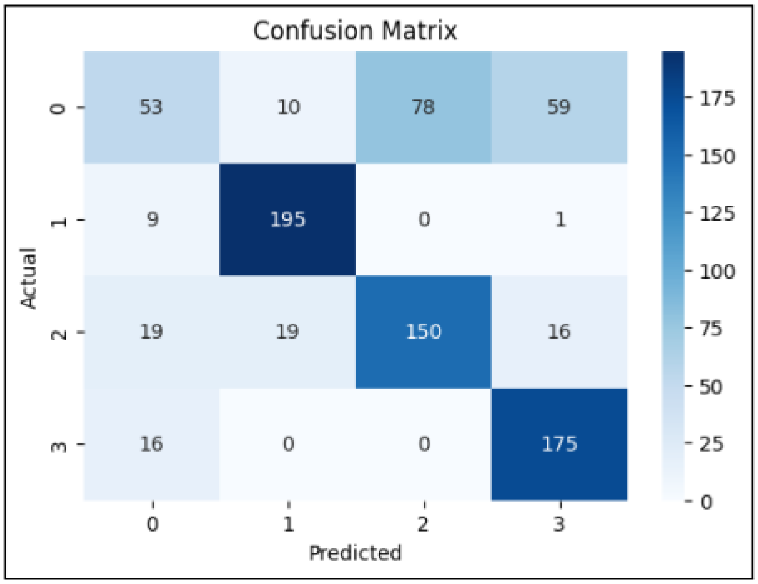


Figure 4.19: confusion matix for 250-peaks

```
Scores: [76.875, 76.5625, 74.0625, 74.6875, 80.9375]
Mean Accuracy: 76.625%
```

Figure 4.20: five folded accuracies and mean accuracy for 100-peaks

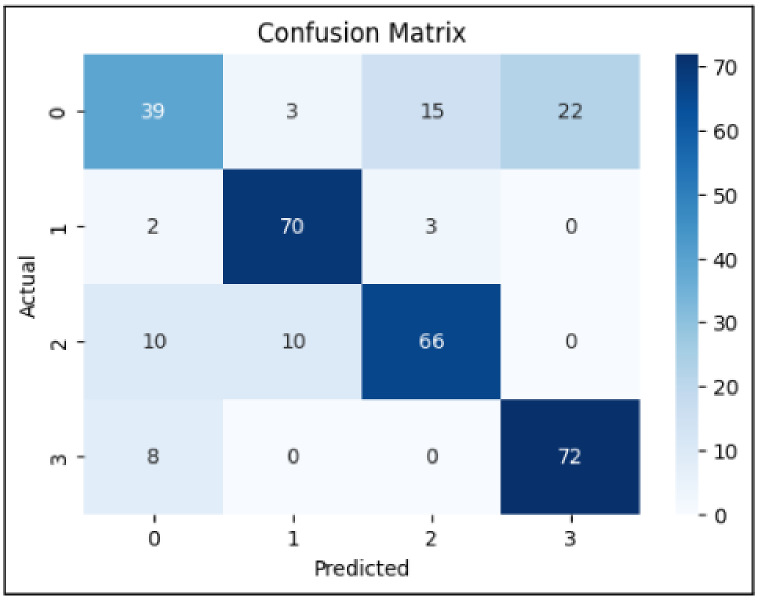


Figure 4.21: confusion matix for 100-peaks

Variable parameters of ANN	n_peaks	mean-accuracy %
n_epochs = 200	1000	72.84%
n_hidden = 10	750	73.392%
l_rate = 0.02	500	73.625%
n_folds = 5	250	74.375%
	100	76.625%

Figure 4.22: Comparing the results

Here we can see a trend of increase in mean accuracy when there is a decrease in no of peaks taken from each fft data for training data decreased. Upon visualising this on a graph we can clearly see the decrease in accuracy as the no of top peaks taken form each fft to form a training data increased.

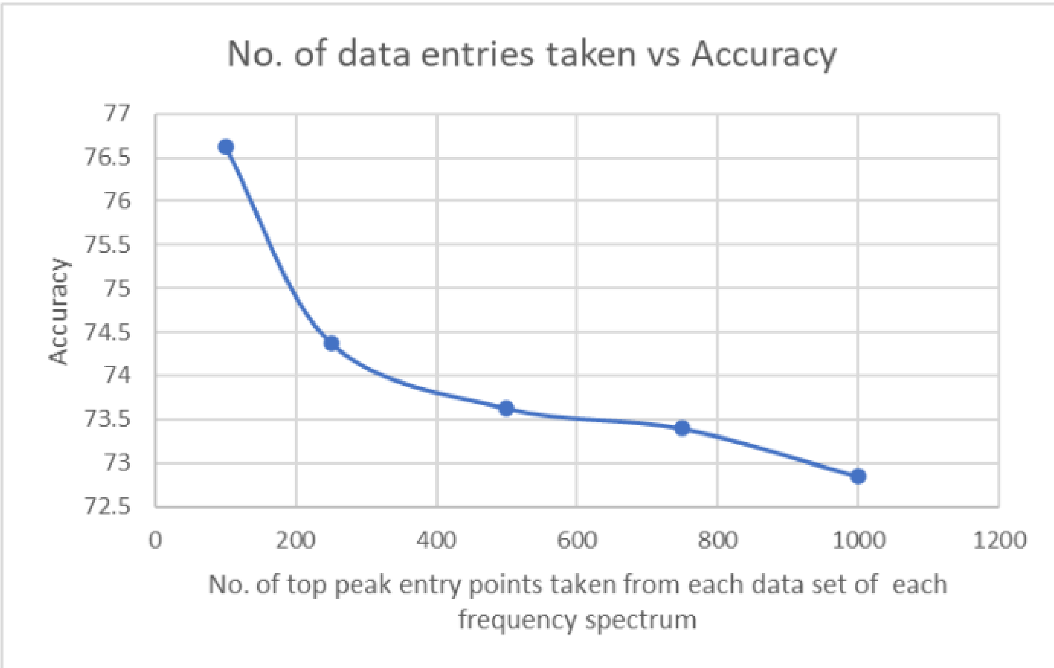


Figure 4.23: Graphical representation of the results of data points from each data set

CHAPTER 5

CONCLUSION

5.1 Conclusion

In this project our goal was to observe any recurring or similar pattern in the training data provided and predict future failures from testing data if any.

We were able extract the required data for our model from the datasets used in the case study and use them. We have used the extracted data and were able to successfully train our model to give upto 70% accuracy.

We have further tried to increase the accuracy of the model by varying different values but the increase in accuracy is not much significant. But we were able to observe a trend of increase in accuracy as the number of data entries of peak values taken from each dataset of each frequency spectrum decreases, i.e., Which means as the no of magnitude values taken from ffts of the main datasets to form a training data set decreased we were able to see an increase in the accuracy.

This is because the vibrational data produced by any machine has some noise/errors in it, the fft of this data also consists of noise. We can observe the noise in the fft plots which were shown previously. Thus taking more number of magnitude values(data points) from fft means increasing the number of data points with noise in the training data. Therefore, as a result, when the no of magnitude values taken from ffts of the main datasets to form a training data set decreased we were able to see an increase in the accuracy because now the training data has lesser in it.

We have also tried to test our model with some real data from industry, but due to lack of sufficient data, we were not able to get proper results. The data provided happened to be of different form and smaller quantity of data thus we were unable to go further with that data for testing our model.

5.2 Future Scope

There is huge scope for future works in this field. In our project we have diagnosis for the bearing data to predict the defect from the data. But this is underlying defect which is already there but unnoticed. But if it was possible to predict even before the defect formation then the goal of minimizing the cost and danger of damage and rundown time becomes even more close.

So Further the same methods can be used to make more accurate model by using the realtime industrial data. Thus after developing the diagnosis model, Regression algorithms are used for bearing life detection. Thus we can predict how long the bearing last before failure. This model can be made even more accurate by increasing the number and types of data given as training input, like along with vibrational signals, we can also add temperature data for detecting bearing life. Unlike fault detection which is classification problem, life detection is a regression based problem, Other methods which can be used for Bearing life detection are Support vector machines, Random forests which can be used for both regression and classification models and regression models like Gaussian process can be used.

CHAPTER 6

REFERENCES

[1] N Kornaev, E Kornaeva and L Savin(2020), "Application of artificial neural networks to fault diagnostics of rotor-bearing systems". IOP Conference Series: Materials Science and Engineering.

[2] Case Western Reserve University. (n.d.). Bearing Data Center. Retrieved March 1, 2023.

[3] H. Saruhan, S. SarÕdemir, A. Çiçek3 and ø. Uygur (2014), "Vibration Analysis of Rolling Element Bearings Defects".Journal of Physics: Conference Series, 551(1), 012006. doi:10.1088/1742-6596/551/1/012006

[4] Oliver K. Ernst. (2014), "Stochastic Gradient Descent Learning and the Backpropagation Algorithm". Brain Computer Interface (BCI)

[5] Analytics Vidhya. (2022, February 1). "K-fold Cross-Validation Technique and Its Essentials".

[6] Geron, A. (2019, February 11). "K-Fold Cross-Validation. Machine Learning Mastery".

[7] Smith, W. A., Randall, R. B. (2006). "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study". Mechanical Systems and Signal Processing.

[8] Guru99. (n.d.). "Backpropagation Neural Network". Retrieved May 4, 2023.

[9] MathWorks. (n.d.). "Rolling Element Bearing Fault Diagnosis - MATLAB Simulink - MathWorks", Retrieved May 4, 2023.

[10] Matt Mazur. (2015, March 17). A Step by Step Backpropagation Example.