MULTI-STAGE PREDICTION OF BEARING FAILURE

ADRIAN LIM JERN EE

A project report submitted in partial fulfilment of the requirements for the award of Bachelor of Engineering (Honours.) Mechanical Engineering

Lee Kong Chian Faculty of Engineering and Science
Universiti Tunku Abdul Rahman

April 2019

DECLARATION

I hereby declare that this project report is based on my original work except for citations and quotations which have been duly acknowledged. I also declare that it has not been previously and concurrently submitted for any other degree or award at UTAR or other institutions.

Signature   : _____

Name   : Adrian Lim Jern Ee

ID No.   : 1402577

Date   : _____

I certify that this project report entitled **"MULTI-STAGE PREDICTION OF BEARING FAILURE"** was prepared by **ADRIAN LIM JERN EE** has met the required standard for submission in partial fulfilment of the requirements for the award of Bachelor of Degree (Honours.) Mechanical Engineering at Universiti Tunku Abdul Rahman.

Approved by,

Signature   :

Supervisor   :

Date   :

Signature   :

Co-Supervisor :

Date   :

ACKNOWLEDGEMENTS

This research was supported by University Tunku Abdul Rahman. I would like to thank all the involved parties who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations of this thesis.

First of all, I would like to personally thank Prof. Dr Andy Chit Tan for all the guidance, assistance as well as supports provided. The biweekly meeting does provide me with lots of insights on this particular project. At the same time, this project wouldn't be completed without Prof. Dr Andy Chit Tan sparing his precious time guiding me through all the important aspects of this project with his expertise.

I would also like to show my gratitude to Ts Dr Yeo Wei Hong for sharing his pearls of wisdom with me during my presentation of FYP 1. The questions asked made me further my research and the flaws pointed out improve the project directly.

Furthermore, I would like to thank Mr Mark for guiding me through the programming section. Thank you for teaching me even though I had no programming background.

Lastly, I would also like to thank the lab staffs for being cooperative and understanding.

ABSTRACT

Bearing is a commonly used item in many fields to reduce frictional force in rotating machineries such as manufacturing field, aerospace field and so on. Therefore, bearing plays an important role in rotating machineries where its lifespan is crucial. A bearing, which is changed way too early before it fails results in wastage of materials whereas a bearing that fails during operation causes losses due to unplanned breakdown. However, a large percentage of industries are still relying on human experience prediction to predict the lifespan of bearings, which is inefficient and inconsistent. Therefore, this work provides a series of techniques which predicts the lifespan of bearing. This series of technique includes signal processing, diagnosis as well as prognosis. It was noted that the sensors used in this project includes acoustic emission (AE) sensor, thermocouple as well as accelerometer. The main emphasize of this work would be on the steps of feature selection as well as prognosis. For feature selection, two main elements were included which is neighbourhood component analysis (NCA) as well as recursive feature elimination (RFE). RFE is mainly to exclude the unimportant features and provide carefully analysed weightage for each unique feature. NCA then make use of the weightage computed through RFE to produce a health indicator. On the other hand, prognosis uses support vector regression (SVR) to further predict the remaining useful life (RUL) of bearing. Firstly, SVR uses the health indicator to predict the RUL of each individual training test. The result generated will be then compiled into combined training test. Finally, when the combined training test is matured, the training data will be used to predict online test. This work also emphasizes on the technique used for grid search as well as cross-validation to tune the parameters. By carrying out the series of technique mentioned above, the online test conducted achieve accuracy as high as 81.95 %.

TABLE OF CONTENTS

CHAPTER

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS / ABBREVIATIONS

A,B,C,D,E	Weightage Assigned for Each Feature
AAC	Artificial Ant Clustering
ALE	Adaptive Line Enhancer
ANN	Artificial Neural Network
AR	Autoregressive
b	Bias Term
BPF	Ball Passage Frequency
BPFI	Ball Pass Frequency of Inner Ring
BPFO	Ball Pass Frequency of Outer Ring
BPV	Ball Passage Vibration
BSF	Ball Spin Frequency
CNS	Conditionally Non-Stationary
du	Distribution or a Measure
DWPA	Discrete Wavelet Packet Analysis
DWT	Discrete Wavelet Transform
EXP	Math Constant *e* Raised to Power
$e^{\frac{i2\pi kn}{N}}$	Nth primitive root of unity
EEMD	Ensemble Empirical Mode Decomposition
EMD	Empirical Mode Decomposition
f(x, ϖ)	Denotes a Linear Model
FCO	Fault Characteristics Order
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FTF	Fundamental Train Frequency
F(u)	Real Value Function
$g_j(x), j=1,\dots,m$	A Set of Non Linear Transformation
HFRT	High Frequency Resonance Technique
HHT	Hilbert-Huang Transform
HMM	Hidden Markov Model
HSMM	Hidden Semi-Markov Model
H(x)	Hilbert Transform of Real Valued Function

IFCF	Instantaneous Fault Characteristics Frequency
IIR	Infinite Impulse Response
IMFs	Intrinsic Mode Function
K	Radial Basis Function (RBF) Kernel Function
K(-i)	Kernel Matrix by Elimination of ith Features in Input x
LDA	Linear Discriminant Analysis
m	Dimensional Feature Space
MED	Minimum Entropy Deconvolution
MoG-HMMs	Mixture of Gaussian Hidden Markov Model
N	Number of Terms
NB	Naïve Bayes
P	Average Power
RDR	residual frequency components-to-defect frequency components ratio
RS-ESg	reverse sequence enveloped spectrogram
RS-SES	reversed sequenced squared enveloped spectrum
RUL	Remaining Useful Life
SIDL	Shift-Invariant Dictionary Learning
SK	Spectral Kurtosis
STFT	Short Time Fourier Transform
SVM	Support Vector Machine
T	Time
VES	Vibration Analysis Expert System
W	Weightage
WPD	Wavelet Packet Decomposition
WPT	Wavelet Packet Transform
X	Training Data, Input, Vector
X'	Point of Reference Vector
$\dot{x}$	Mean Value of Signal
x_k	DFT Coefficient
x(i)	Amplitude of Signal at Specific Time
x(t)	Frequency Constant of Signal
x_n	Complex number
$X_{1...5}$	Amplitude of Each Feature

α_k	Lagrange Coefficient
ϵ	Insensitive Loss
$\| \quad \|$	Euclidean Distance
σ	Kernel Parameter

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1   General Introduction

Bearing life and its multistage bearing failure is of main essence in this study. First of all, bearing consists of several parts, which includes the outer race, inner race, cage, and ball. The main purpose of bearing is to reduce friction between two moving parts and at the same time limiting its motion to rotate in desired directions. When John Harrison first introduced bearing in the mid-1740s, industrial started to implement and install bearings into their machineries. By doing so, it simply increases the lifespan of the machines as bearing reduces not only friction, but at the same time reduces heat, noise as well as vibration. Bearing continues to improve machineries efficiency and life span even in our current generation.

Bearing usually fails in several ways such as wear in the inner or outer race, dislocation of ball bearing, debris contamination in the cage or outer race. The main reason behind each failure is due to friction, heat produced, uneven distributed force, shocks as well as vibration.

However, the bearing failure can be detected by using several sensors such as acoustic emission sensor, accelerometer, thermocouple, and so on. These sensors will be able to pick up trails of the bearing before, after as well as during the failure. For instance, thermocouple will be able to detect the temperature of the bearing, which is useful. A healthy bearing has a temperature temperate range, if the temperature exceeds the range; it is likely to fail sooner. Accelerometer plays the role of measuring proper acceleration of bearing. This data is useful in analysing the remaining useful life of bearing. The Acoustic Emission sensor picks up trails of mechanical loading from material and structures as well as structural changes that generates specific sources of elastic waves. These data are the main data used in analysing the remaining useful life of bearing. The data collected can then be used in bearing prognosis to develop an algorithm. The algorithm will be useful to predict the bearing life span.

Prediction of the bearing's lifespan will be a very promising improvement in the industrial field as it promises reduction in downtime; reduction in spare inventory, reduction in maintenance cost as well as a safe environment.

After many years of research and development, researches came out with a series of techniques to understand more about bearing failure and lifespan. First of all, the first technique used is the Conditioned Based Maintenance, which locates the bearing's defect as well as identify the type of failure. Prognosis then plays its role to analyse the remaining useful life of bearing based on the graph plotted. There are three techniques in bearing prognosis, which includes the physics-based prognosis technique, data-based prognosis technique as well as hybrid-based prognosis technique. Physics-based technique generally uses a series of fundamental physics formula to predict the lifespan of the bearing. Data based technique uses historical data of past bearings to predict the lifespan of the bearing. However, hybrid based technique combines both physics based as well as data based technique to analyse and estimate the lifespan of bearing. Each technique was unique and had its pros and cons. Therefore, the user should carefully choose the technique based on the environment, technology available, budget as well as many other factors.

If research wasn't done on bearing failure, it will leads to more losses. As fourth industrial revolution has already begun, the rates of growth of technology, no matter hardware or software were higher than before. This simply indicates the increasing amount of machineries in the current era and indirectly increases the use of bearings as well. If bearing wasn't optimized, it might either lead to longer downtime of machine, which reduces efficiency, waste the remaining useful life of bearing, or even damage the machineries. Neither of the above scenarios is desirable. According to Roberts (2017), it was claimed that the failure of bearings, which isn't predicted, causes a loss over, 400,000 USD in a middle-class firm in a year. This clearly shows the importance in predicting bearing failure.

In this project however, it was planned predict the bearing's lifespan through several steps which simply includes signal processing of raw data, diagnosis of processed data to verify the fault mode of bearing, and lastly prognosis to predict the remaining useful life of bearing.

1.2   Importance of the Study

Bearing prognosis is important mainly to the industrial field as it simply allow user to estimate or predict the lifespan of the bearing. By doing so, it reduces the downtime of the machine, as unscheduled breakdown will cost the industry to incur a greater amount of loss. It also results in reduction on the maintenance cost. This is because it

allows the user to change the bearing before failure occurs and at the same time not wasting the bearing remaining useful life. This will also indirectly reduce the chances of damaging the machine due to defective bearing. Besides, it also increases the safety factor of the industry, which provide workers with a safer working environment.

1.3   Problem Statement

In recent study and research, it was highlighted that bearing failure has become a major problem to the current industry. This is mainly because it is difficult to estimate the exact bearing life. If the bearing fails before estimated time, this would simply cause an unplanned breakdown, which the industry will incur, losses. If the bearing were changed way before the bearing should fail, this would result in unutilized use of bearing as well as low efficiency. Therefore, it is essential to estimate remaining useful life of bearing. Below are the main problems of bearing failure:

1. Low safety factor without estimation of RUL
2. High potential of losses due to unplanned breakdown
3. Low efficiency if bearing were not utilized
4. High maintenance cost incurred
5. High potential in damaged machineries
6. Unsafe environment for industrial workers without accurate estimation

1.4   Aims and Objectives

The aim of this study is to come up with a prognosis algorithm to achieve multi stage prediction of bearing failure by using the data collected. The objectives of this study include:

1. Conducting experiment by running bearing till it fails and collecting data at the same time.
2. Conducting background noise elimination by applying signal processing technique.
3. Develop diagnosis technique to capture failure of bearing.
4. Develop prognosis technique and algorithm to predict bearing failure.

1.5   Scope and Limitation of the Study

In the experiment, the test bearing operates under a 3 kN radial force. The rotating speed of the test bearing is 1200 RPM. These are the fixed variable to ease the study

of bearing prognosis. Sensors used in the experiment to accumulate data include acoustic emission sensor, thermocouple as well as accelerometer. Once the data were collected, the files will then be converted from WFS files to CSV files in order for the Personal Computer to process it. The data will then undergo noise elimination by signal processing and further prognosis analysis. The limitation of the study includes:

- Daily bearing running test only last for eight hours as it requires human observation.
- The shaft only consists of one size, which limits the test bearing to one size.
- Only two seconds of AE data were collected in between 5 minutes interval due to software limitations.

1.6   Contribution of The Study

Although the current industry has technology advances at a fast pace, majority of the industries are still relying on human experience to maintain bearings within its rotary machines. These estimations used in industry are not reliable, accurate as well as consistent as human error plays a very important role. Instead of relying on the judgement based on what human sees, what human hears or even smell, it would be a wise approach to apply a much more systematic and consistent way of predicting bearing's lifespan.

Therefore, this study is mainly to contribute in developing a series of algorithm to accurately predict the bearing's lifespan. These series of techniques include signal processing, diagnosis as well as prognosis. Signal processing serves the purpose of denoising the data collected, diagnosis serves the purpose of identifying the fault mode of the bearing. Prognosis however serves the purpose of being trained with historical data and finally execute decision making in prediction when applied online. By doing so, bearing's RUL will not be wasted and at the same time prevents unplanned breakdown of machineries which is due to bearing failure.

1.7   Outline of The Report

This thesis consists of several parts and it begins with a general introduction which introduce the problems faced in the industries as well as the importance of the study. The thesis continued with literature review section which includes some discussion of other researches which are relevant to this study. The following part will be methodology which includes how the test were conducted and the series of proposed

technique to analyse the data collected. Each technique will be further introduced and explain in details. The results and discussion section however present the raw as well as analysed data of each technique in a chronological order. This involves the results and explanation on signal processing, diagnosis as well as prognosis. Finally, the conclusion section concludes the results obtain from this study as well as recommendation to further improve this study.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This section of the report consists of 2 main relevant parts in the field of machine condition monitoring and prognosis. Those parts are:

2.1.1 Signal Processing and Diagnostics

Signal processing is a procedure where signals data collected through sensors were processed in order to eliminate the background noise. It serves several other purposes such as improving signal transmission accuracy, storage effectiveness as well as subjective quality. In order to extract features as well as move on to diagnostics and prognosis, signal processing is vital. Signal processing performs noise elimination which allows individuals or software to determine which are the crucial signal indication of bearing health. Without performing this step, it would confuse the model on the bearings exact health conditions. Signal processing has several different techniques such as Fast Fourier Transform (FFT), infinite impulse response (IIR) filter, finite impulse response (FIR) filter as well as the others. These techniques are catered for different condition as well as parameters.

Diagnostics were usually performed together with signal processing. It simply serves the purpose of detecting the types of failure in the failed bearing, acting as the health indicator and also extracting feature from separate graphs to understand bearings latest condition. Although there are many different diagnostics methods but high resonance frequency technique will be applied in this study.

Wang et al., (2016) proposed a method to diagnose bearing fault which consists of Hidden Markov Model (HMM) and this method emphasizes on learning method rather than a fixed formula or fixed methodology. The Shift-Invariant Dictionary Learning (SIDL) plays the role of adaptive feature extraction technique which analyse both simulated as well as experimented signal. The HMM then identify the fault types. The accuracy of this technique is as high as 96.25%.

Kang et al., (2015) proposed an envelope analysis by utilizing the (genetic algorithm) GA-based adaptive filter bank. The method proposed enhances 134% accuracy, 46% accuracy, and 9% accuracy of (residual frequency components-to-

defect frequency components ratio) RDR_{BCO}, RDR_{BCI}, and RDR_{BCR} values compared to the conventional methods. It was explained that conventional methods do not reveal fault information details due to sub band signals influenced by harmonics of working frequency.

Nguyen and Kim (2015) proposes a method which combines (wavelet packet transform) WPT-based kurtogram together with an advance vector median-based feature analysis method. It first extracts main features from bearing health condition of time domain, frequency domain as well as envelope power spectrum using WPT-based kurtogram. Then it uses (linear discriminant analysis) LDA technique to identify important fault modes of bearing. Once the selected fault features were chosen, (naïve Bayes) NB classifier will then determine bearing fault conditions. It was claimed that this method produces average accuracies of 91.11% for rotational speed 300 rpm, 96.67% for rotational speed 350 rpm, 98.89% for rotational speed 400 rpm, 99.44% for rotational speed 450 rpm as well as 98.61% for rotational speed 500 rpm.

Wang et al. (2014) proposes a method to extract signals related to rotational speed. This method consists of four steps which is signal filtering through spectral kurtosis (SK), short time Fourier transform (STFT), instantaneous fault characteristics frequency (IFCF), and lastly fault characteristics order (FCO). This method was claimed to solve problem of bearing diagnosis with varying speed (speed changes smoothly without dramatic variations) without a tachometer.

Alonso-Betanzos et al. (2013) proposes an automatic method for bearing fault detection. A one-class v-SVM was used to differentiate differences between healthy and faulty bearing. It then uses Hilbert transform and bandpass filters to produce envelope spectrum to analyse and determine location of the fault. It was claimed that with the rotation speed of 1740 rpm, the method achieves 100% accuracy in detection.

Ricci, et al. (2013) proposed reversed sequenced squared enveloped spectrum (RS-SES) which is a unique way of obtaining squared envelope signal. It is based on the order tracking of envelope signal and it was also proven to be accurate in numerical simulation as well as experimental data. This technique consists of several steps such as angular-resampled signal of computed data, band pass filtering signal, squaring enveloped signal, applying discrete Fourier transform to produce envelope spectrum and finally squaring absolute value to obtained squared enveloped signal. Reverse sequence enveloped spectrogram (RS-ESg) was also proposed to cope with highly

variable load conditions as well as pre-whitening step to enhance quality of the technique.

Ming et al., (2011) proposed a method which was establish from the cyclic Wiener filter as well as envelope spectrum analysis. This method makes use of cyclic Wiener filter to obtain spectral coherence theory and the original signal was replicated and shifted in frequency domain. The method then uses a filter-bank to minimize the noise to go on with envelope spectrum. Finally, the essence of the fault is extracted for further diagnosis.

Liang and Bozchalooi (2010) proposed a new method of detecting bearings fault. It is based on a simple transformation followed by spectral analysis step. This technique first transforms signal of fault frequency and amplitude modulation, it then increases strength of amplitude-demodulation impulses with energy operator. This technique was claimed to be advantageous because no parameters as well as no location of bearing was needed. It was also claimed that the technique is effective on revealing both inner and outer race faults characteristics.

Randall and Antoni (2010) explains the method of envelope analysis. It consists of several steps which is the separation of bearing signals from discrete frequency noise. By filtering the noise, it will be more advantageous to do bearing diagnosis. Methods of filtering noise includes linear prediction which obtains deterministic part of signal based on immediate past samples, adaptive noise cancellation which separates 2 uncorrelated components by using a reference signal, self-adaptive noise cancellation which uses transfer function between deterministic part and the delayed signal do to separation. The next step will be the enhancement of the bearing signals which uses minimum entropy deconvolution (MED) technique to eliminate effect of transmission path and then STFT-based spectral kurtosis to detect trails of transients in a signal. The envelope analysis is then applied. Envelope analysis is where signal being bandpass filtered and fault impulses were amplified by resonance structure. Demodulating the amplitude then formed envelope signals.

Immovilli et al. (2009) explains generalized roughness damage on bearings produces characteristics fault frequency spreading. This makes spectral and envelope analysis hard to detect faults. It was claimed that spectral-kurtosis energy of vibration technique could solve the problem. The technique works in a way that kurtogram of SK was meant to locate the stronger effect of fault within the bandwidth. The

amplitude of the signal in the mentioned bandwidth will then be used as diagnostics index.

Rafsanjani et al., (2009) proposed an analytical model to understand nonlinear dynamic behaviour of bearing which includes surface defects. It was first modifying Newmark time integration technique in order to resolve equations of motion numerically which produces results in frequency response, time series as well as phase trajectories. It was then further compared with experimental results and applying classical Floquet theory to discover linear stability of defective bearing. Lastly, peak-to-peak frequency response, basic routes to periodic as well as disorderly motions for different internal radial clearances are obtained.

Onel and Benbouzid, (2008) conducted experiments of bearing failure detection and diagnosis in induction motor. They compared between Park and Concordia transforms approaches in bearing failure. They claimed that Park transform shows better diagnosis. The reason is because Park transform is speed sensor based (bearing diagnosis) and 50% of bearing failure accounts to the induction motor.

Sreejith et al. (2008) proposed feedforward neutral network technique for diagnosis of bearing faults. They claimed that by inserting 12 sets of 10 normalized features, differences between normal and defective cases could be effectively identified. However, when the experiment was repeated with eight input features lesser, the network performs similar diagnostics accuracy as before. The inputs of the time domain parameters include normal negative log-likelihood value and kurtosis value.

Liu et al. (2008) explains extended wavelet spectrum analysis technique. It consists of two different approaches which is two statistical indexes as well as extended Shannon function. Two statistical indexes quantify resulting wavelet function. It was claimed that wavelet function is more effective when used over appointed frequency bands. Shannon function on the other hand synthesize wavelet coefficient over designated bandwidths. The average autocorrelation power spectrum is then applied to highlight bearings characteristics.

Nataraj and Harsha (2008) proposed a model for non-linear behaviour of unbalanced rotor bearing system due to cage run out. The model focuses on Hertzian contact forces, cage run out as well as the outcome of transitions which are non-linear. It was claimed that when the cage run out condition happens, ball passage vibration

(BPV) is parametrically excited and ball passage frequency (BPF) will lead to peak amplitudes of vibration.

Ebersbach and Peng (2008) claimed that the expert system developed which is vibration analysis expert system (VES) was developed specifically to analyse vibration, oil and wear debris. This system is also able to produce a single correlated condition report. This system allows limited operator input but yet having high fault detection capability. An example was also provided when a horizontal 4000Hz spectra shows evident of defects in loose and eccentric gears.

Rai and Mohanty (2006) claimed that many techniques including high frequency resonance technique (HFRT), fast Fourier transform (FFT), wavelet transform (WT) and so on are able to predict health of bearings but only few can localize the defects. The Hilbert-Huang transform (HHT) however can detect localized bearing defects as well as predict health of bearings. HHT uses empirical mode decomposition (EMD) process, which applies on non-stationary data to obtain intrinsic mode function (IMFs). IMFs can then be further process by ARMA. For instance, when FFT is applied alone, defect frequencies conspicuous cannot be detected. WT on the other hand uses fixed decomposition scales for analysis, which causes leakage problems.

Antoni (2006) stated SK is able to show the existence of series of transient as well as their location presented in frequency domain. This technique is not popular due to the lack of information. However, it is able to detect many problems, which are useful not only in features extraction but also diagnosis. It was also claimed that power spectral density is a measure of position, which is time-average, on the other hand SK is a measure of dispersion which is also time-variance. STFT-based estimator of SK was also proposed and when it was applied to non-gaussian driven conditionally non-stationary (CNS) process, minimum of 75% overlap is recommended to be used to get shift invariant results.

Yu et al. (2005) claimed that EMD energy entropy can analyse and obtain different results of vibration signal that fluctuate in distinctive frequency bands when bearing fails. It was also claimed that energy features extracted from IMFs has most important fault information which could be used as input vectors of artificial neural network. First of all, EMD will be used to pre-process several vibration signals. ANN was then applied to pre-processed data of roller bearing which is under working

condition. EMD energy entropy will vary with work condition of roller bearing which indicates different faults.

Stack et al. (2005) came up with fault signature modelling and detection of inner race bearing faults. It was claimed that inner race defects comprised of peaks separated by F_{IRF} as well as peak within groups separated by F_S. When combining it with principle of phase coupling, fault detector can be developed. The detector was then applied to data and peak that were spaced by F_{IRF} were used as fault index. It was also claimed that inner race defects and bearings without single points defects which has low fault index count (1 to 2) are easily detected.

Purushotham et al. (2005) proposed a new method on detecting localized bearing defects by using wavelet analysis and hidden Markov model. Discrete wavelet transforms (DWT) is first used to detect bearing race faults; the wavelet transform approach then provides a variable resolution time-frequency distribution which is mainly due to repetitive force impulses. When results were compared with feature extraction data as well as spectrum analysis, it was claimed that DWT is more efficient in detecting faults. When HMM was applied, vibration data were mainly to extract complex cepstral coefficients for wavelet transform time windows at Mel-frequency scales. The accuracy was as high as 99% when HMM was combine with Mel-frequency complex cepstrum analysis method.

Duque et al. (2005) explains spectral analysis of line current in the case of sinusoidal supply. It was claimed that the proposed formulation can be used with motor driven by frequency converters and it has high accuracy which further spectral analysis is not necessary. The experimental results show the motor current spectral analysis is reliable and remarkable at the presence in the spectra of components corresponding to k =5.

Silva and Cardoso (2005) introduces a spectral analysis of motor current Park's Vector modulus for identifying bearing faults in three-phase induction motor. It was claimed that this approach shows positive results even for early stage damage in the case of ⌀2 mm hole drilled to outer race. This technique also claims to be strong analysis as it doesn't require access to the motor. It also claims that extend park's vector approach (EPVA) is more sensitive compared to conventional vibration analysis itself.

Samanta and Al-Balushi (2003) proposes a fault diagnosis technique for bearing using artificial neural network (ANN). It is learned that wavelet transform did

not enhance the results. However, multiple signals (original signals and high frequency components of signals) is indeed helpful in diagnostics approach. ANN-based approach has limitations such as unable to differentiate normal and defective with several fault types as well as levels of severity. Apart from that, some specific ANN structure is not applicable for all machine test and therefore requires customization.

Altmann and Mathew (2001) claimed that discrete wavelet packet analysis (DWPA) is an accurate technique to detect and diagnose low-speed rolling-element bearing failure. It was claimed that the technique overcome limitations of determining frequency bands of interest. Notable gains in signal-to-noise ratio are shown even under conditions where resonant frequency was not well known. When work together with autoregressive (AR) spectral analysis, it shows enhanced diagnostics results for low-speed-rolling-element bearing vibrations.

Williams et al. (2001) conducted run-to-failure lifetime testing on bearing and discovered that the AE sensors are able to pick up distinct peaks when inner race fails even though elastic waves that are generated traveled from inner race to the housing. However, the failure of outer race did not show large and distinct peak. The technique used to improve frequency spectrum includes adaptive line enhancer (ALE) and HFRT. These techniques have proven to improve diagnosis of bearing failure.

2.1.2   Prognosis

Prognosis is basically classified into three main categories as what was stated by Dawn An et al. (2015). These three categories of prognosis include data based prognosis, physics-based prognosis as well as hybrid-based prognosis. Each category has their strengths and weaknesses; they should be selected based on the suitability of the current research conditions.

The first techniques introduced by physicians were the physics based prognosis technique. This technique focuses on equations as well as formula, which are theoretically solid. The general formula simply uses the main force applied to the bearing as well as the basic dynamic force to generate a prediction. However, this technique does not provide accurate results and prediction when applied in real life condition, as there are many more variables, which needs to be taken into consideration. These considerations include the impurities trapped in the bearing, precision of the bearing when manufactured, external environment of the bearing and so on. It is also

not possible to consider all the variables into the equation as it will not only be complex but at the same time very time consuming.

The second technique introduced is the data-based prognosis technique. This technique simply requires a lot of data to study the trend and bearing lifespan to predict the remaining useful life of the bearing. By collecting lots of data, it allows software to plot graph and combine them together in order to predict the average lifespan of a bearing. At the same time, this technique also studies the machine's condition while predicting the lifespan of bearing. However, this technique might not be accurate as there are many other considerations which must be taken into account. This technique also requires a whole lot of data in order to predict a much accurate lifespan of bearing life. Therefore, this technique requires a long time to develop.

The third technique however combines the above techniques which is known as the hybrid-based prognosis technique. This technique is known to be more accurate than the other two techniques mentioned above as it combines both ideas and optimized it. Although it is much more accurate and reliable, but it simply requires a lot of time, effort and technology to develop the model.

Huang et al. (2015) proposed a fault diagnosis methodology based on ensemble empirical mode decomposition (EEMD) and Bayesian network. When compared to conventional EEMD, this method uses all useful information besides sensor signals. When all the information was combined, a generalized three-layer Bayesian network can be developed and this can improve diagnostic accuracy and capacity. Comparing fault features extraction of EEMD with ANN and support vector machine (SVM), it improves the average diagnosis accuracy by 4.5% and 3.5%.

Soualhi et al. (2014) proposed a methodology that extracts features by time domain from vibration signals which later classified as health indicator. They then sus artificial ant clustering to detect degradations of the bearing. Hidden Markov model will then be used to analyse the next degradation rate while multistep time series prediction and adaptive neuro-fuzzy inference system will analyse the time of the next degradation. The proposed technique was claimed to be effective when used in artificial ant clustering (AAC) to detect degradation states of bearings.

Sloukia et al. (2013) proposed a data-based prognosis technique which is based on transformation of data delivered by sensors. A hybrid method between mixture of gaussian hidden Markov model (MoG-HMMs) method and support vector machine (SVM) method was introduced. It was claimed that SVM is a better classifier and

generalized well but it can't represent temporal evolution which is vital to distinguish between vibratory signals. However, MoG-HMMs requires more learning models in order to produce accurate results. When both techniques were compared, the relative error between both techniques is 21.85%. When both techniques were combined, the accuracy is higher than both models.

Tobon-Mejia et al. (2012) proposed a technique which compose of MoG-HMMs and wavelet packet decomposition (WPD). This method started off with the first phase (off-line phase) where features extraction was performed in the form of WPD coefficients. This step is to estimate parameters of the technique as well as establish learning algorithm. The second step (online phase) make use of generated model to estimate remaining useful life (RUL) of bearing life. The results were claimed to be accurate but it was also advice to implement precise detection for better results.

Peng and Dong (2011) presented an age dependent hidden semi-Markov model (HSMM) based prognosis method. It was claimed that conventional HSMM method's assumptions are stating bearing with a less healthy state which does not increase with age. The conventional transition matrix was improved with new aging factors and hazard rate was also introduced to predict RUL of bearing. The technique introduced was claimed to be accurate but adding workloads into the parameters would further enhance the result.

Tobon-Mejia et al. (2011) present utilization of Mixture of Gaussians Hidden Markov Models (MoG-HMMs). The technique works in such a way that it started with feature extraction and followed by prognosis. The first phase of prognosis (learning) is to extract useful features which are then used as inputs of learning algorithm. The second phase (evaluation) then input features extracted to the generated model. It was claimed that the RUL estimated has the accuracy of 95 to 99.5%.

Kankar et al. (2011) compare several techniques used in bearing prognosis till date. It was claimed that model-based prognosis methods produce accurate results but it is difficult to produce analytical models and limited in practice. The experience-based prognosis methods simply based on simple reliability functions rather than complex models. The setback is the accuracy especially in new systems. The data-driven model uses online data to predict RUL of bearing. The procedure can even automatically adapt to changes in environmental factors. It has advantage over both models mentioned above.

Dong and He (2007) proposed a methodology which is based on segmental hidden semi-Markov models (HSMMs). This approach has more powerful modelling and analysis as well as prognosis capability unlike HMM that follow unrealistic Markov chain assumptions. It was claimed that the model has a recognition rate higher than 96% when applying on real world hydraulic plant health monitoring. When compared to HMM, it has 29.3% higher recognition rate.

2.1.2.1 Support Vector Regression Prognosis LR

Abdenour et al. (2014) proposed a technique to monitor degradation of bearing through SVM (support vector machine) which is also categorized as supervised classification. It was claimed that a series of techniques such as HHT, SVM as well as SVR could produce high accuracy predicted RUL results. However, this series of technique can only be implemented when the ample of training data are available to be trained.

Theodoros et al. (2013) proposed a series of way in optimizing E-SVR hyper-parameters to implement E-SVR which claimed to achieve high accuracy. This approach also includes a Bayesian Network analysis on error bounds. In order to optimize the hyper-parameter, a quantitative grid search should be done which claimed to achieve accuracy as high as 97 percent.

Benkedjouh et al. (2013) proposed a technique to accurately predict the RUL of bearing which emphasize on SVR. This technique proposed consists of two steps which is online as well as offline step. The offline step should first to carried out to train SVR model on the degradation trend on bearing's RUL. The online step however applied what was trained during the offline step to predict bearing's RUL. It was claimed that the accuracy obtained by the technique mentioned was as high as 96.95 %.

2.1.2.2 Recursive Feature Elimination

David et al. (2015) claimed that SVM-RFE is a very promising technique but may be biased when containing high correlated features. Therefore, a new series of technique was introduced which is SVM-RFE-CBR (correlation bias reduction) which claimed to overcome the cons of SVM-RFE technique. It was further claimed that SVM-RFE-

CBR has higher accuracy when compared to SVM-RFE. However, the proposed technique could still be further enhanced by implementing ensembled method.

Marc et al. (2010) discussed the technique of Recursive feature elimination (RFE). It was claimed that this technique could remove variables which are not desired. By doing so, this could improve the general result of accuracy greatly. It was also claimed that removing unwanted features provides better learning condition for model and at the same time improves the computational time.

Pablo et al. (2006) compared two different techniques which is random forest-recursive feature elimination (RF-RFE) and SVM-RFE. It was claimed that RF-RFE has higher accuracy when compared to SVM-RFE. It was further explained that RF-RFE performs better when it comes to locating minor features that has high discrimination levels. At the same time, it was also claimed that RFE has high potential in overfitting the feature selection process that could further leads to error.

2.2   Summary

In order to carry out prognosis on bearing test, signal processing and diagnostics must first be conducted. In the literature review mentioned above, several ways of signal processing such as GA-based adaptive filter, combination of SK, STFT, IFCF and FCO, as well as the other methods, which shows enhancement in accuracy over conventional method. Besides, comparison between different diagnostics technique were also made to compare differences in accuracy under certain conditions. New diagnostics technique was also proposed which simply combines two different techniques into one. Furthermore, hybrid technique on prognosis was also mentioned such as MoG-HMMs method and SVM, which produces higher accuracy than the original method. Comparison between prognosis techniques was also made such as differences between EEMD, SVM and ANN method. All these reviews aids in selecting the essential way of carrying out signal processing, diagnostics and prognostics in this study. However, HRFT technique shows reasonably high accuracy due to a systematic series of technique applied. NCA also shows much superior characteristics and more accurate results when dealing with machine learning compared to PCA. In order to analyse weightage accurately, RFE is easy to use and it shows relatively high accuracy. Finally, SVM is also a popular technique recently as it is a very promising technique in terms of dealing with a moderate amount of data. Therefore, this study applies HRFT for signal processing and diagnosis, RFE for

weightage analysing, NCA for health indicator distribution and finally SVM for prognosis.

CHAPTER 3

METHODOLOGY AND WORK PLAN

3.1   Work Plan

The work plan in final year project part one consists of several steps such as problem formulation and project planning, literature review, research methodology, report writing and presentation, as well as data gathering, shown in Table 3.1. First of all, problem formulation was done together with professor Andy to assure the direction of the study was right. Then, literature review on different authors was conducted in order to understand the topic as well as to learn different techniques better. In the section of research methodology, it is to create a standard operating procedure for the test and report writing is to present the final work done. Presentation will however be conducted under the supervision of Prof. Dr Andy Chit Tan and data gathering will be a continuous process which was also done consistently since week two.

Table 3.1: Work Plan (Part 1)

No.	Task	Week													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
M1	Problem formulation and project planning (Schedule)	█	█												
	(Actual)	▒	▒												
M2	Literature review (Schedule)		█	█	█	█	█	█	█						
	(Actual)			▒	▒	▒	▒	▒	▒	▒	▒	▒			
M3	Data gathering (Schedule)						█	█	█						
	(Actual)		▒	▒	▒	▒	▒	▒	▒	▒	▒	▒			
M4	Data investigation (Schedule)									█	█	█	█	█	
	(Actual)					▒	▒	▒	▒	▒	▒	▒			
M5	Report writing and Presentation (Schedule)												█	█	█
	(Actual)						▒	▒	▒	▒	▒	▒	▒	▒	▒

█	Schedule
▒	Actual

The work plan in final year project part two consists of several steps such as developing prognosis algorithm, refine prognosis parameters, analyse and evaluate results, conduct test and analyse data, as well as report writing and presentation, shown in Table 3.2. First of all, developing prognosis algorithm was done under the guidance of Prof. Dr Andy Chit Tan as well as Mr. Mark (who is a software engineer). Then, refining prognosis parameters was done by reviewing more literature on support vector regression. Analysing and evaluating results was done once the algorithm and parameters were defined. Presentation and report writing will however be conducted under the supervision of Prof. Dr Andy Chit Tan. Test was only conducted from week seven to week twelve as developing algorithm for prognosis was extremely time consuming (lack of coding background).

Table 3.2: Work Plan (Part 2)

No.	Task	Week													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
M1	Development of Prognosis Algorithm (Schedule)	■	■	■	■										
	(Actual)	▨	▨	▨	▨	▨									
M2	Refinement of Prognosis Parameters (Schedule)					■	■								
	(Actual)					▨	▨	▨							
M3	Analyse and evaluate result. Conduct tests and analyse data. (Schedule)							■	■	■	■	■			
	(Actual)							▨	▨	▨	▨	▨	▨		
M5	Report writing and Presentation (Schedule)												■	■	■
	(Actual)											▨	▨	▨	▨

■	Schedule
▨	Actual

3.2   Requirement/ Specification/ Standards

In this study, it consists mainly of several parts such as signal processing, bearing diagnostics as well as bearing prognostics. In order to achieve the objective of this project, signal processing must first be conducted to eliminate noise of signals. Bearing diagnosis then plays its role to determine the fault type as well as the fault modes. Finally, bearing prognosis serves the purpose of determining the remaining useful life of the bearing. The test bearing used has specifications shown in Figure 3.1.

Configured Specifications

Raceway Ring Shape	Ball	Inner/Outer Ring Material	SUJ2 Equiv.
Bearing Type	Open	Outer Ring Type	Flat
Precision (JIS)	Class 0	Inner Dia. d(Ø)	15
Outer Dia. D(Ø)	32	Width B (or T) (mm)	9
Basic Load Rating Cr (Dynamic)(N)	5600	Allowable Rotational Speed(rpm)	24000
Specifications, Environment	Standard	Load Direction	Radial
Number of Raceway Ring Rows	Single Row	Size Standards	Metric System
Rolling Element Material	[Steel] SUJ2	Basic Load Rating Cor (Static)(N)	2830
Seal Part Structure	Open	RoHS	6

Figure 3.1: Configured Specification

3.2.1   Test rig

In this study, a conventional milling machine, which is equipped with a 3-phased induction motor. The milling machine was modified into a vertical bearing test rig to conduct bearing run-till-failure test. Figure 3.2 shows the configuration of the test rig including labels of it.
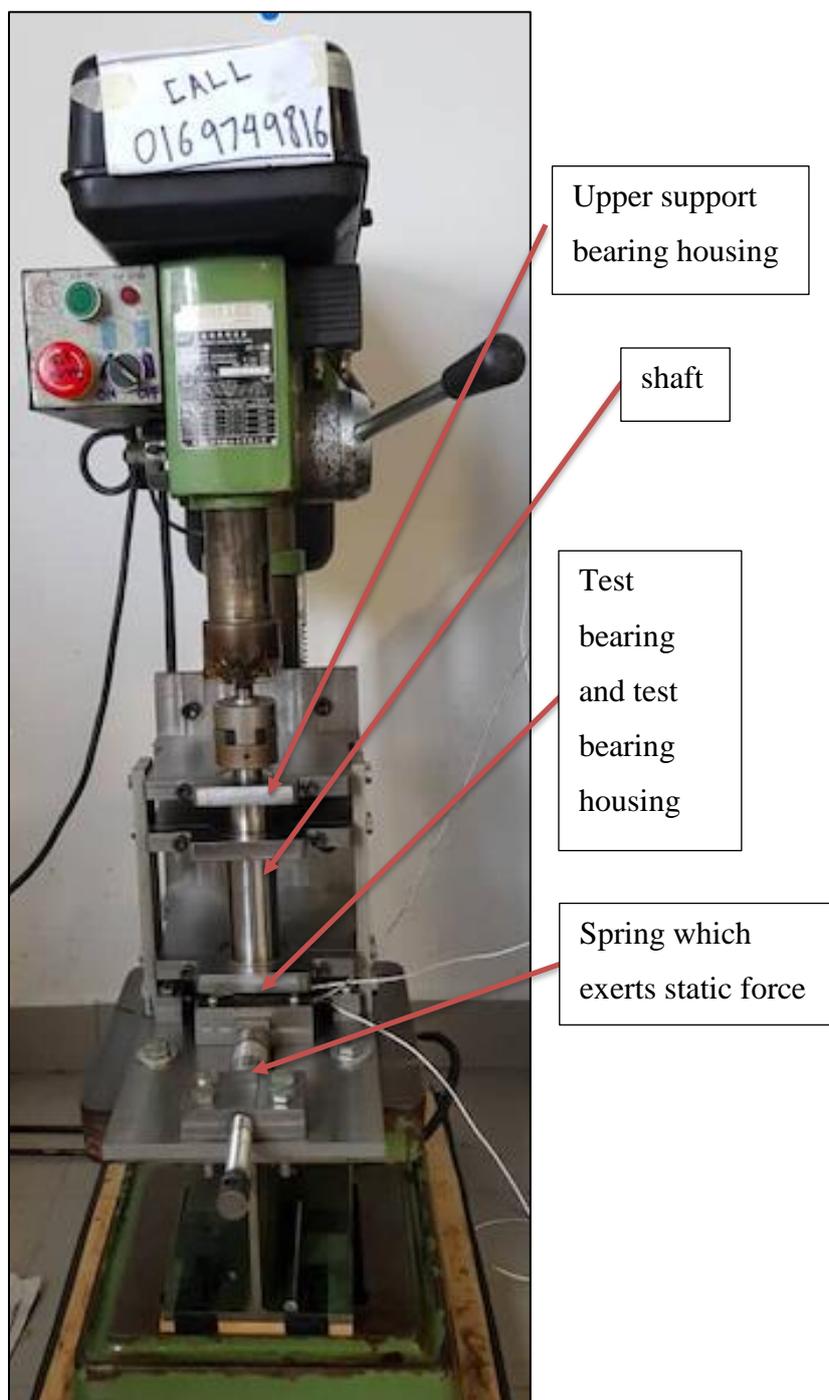

Figure 3.2: Test Rig Machine

1. Test bearing is force fitted onto the shaft by using a hydraulic jack.
2. Spring, which exerts static force, will be placed into position first.
3. The shaft will then be positioned on the test rig and upper support bearing housing will be installed, shown in Figure 3.3. (To make sure shaft was help firmly in position)
4. AE sensor, Thermocouple as well as accelerometer will be installed on test bearing housing. Based on the view of the diagram, the AE sensor will be placed on the left and the other sensors will be placed on the right.
5. The lower support bearing housing will then be installed once the above are completed.
6. Speed of the motor will be set to 1200 RPM by placing the belt onto the right position.
7. All the bearings will be lubricated daily.
8. Electricity will then be supplied to the test rig to turn it on.
9. A static load of 3 kN will be applied onto the test bearing.
10. Once the bearing fail, the static load will first be unloaded followed by cutting off the current supply.
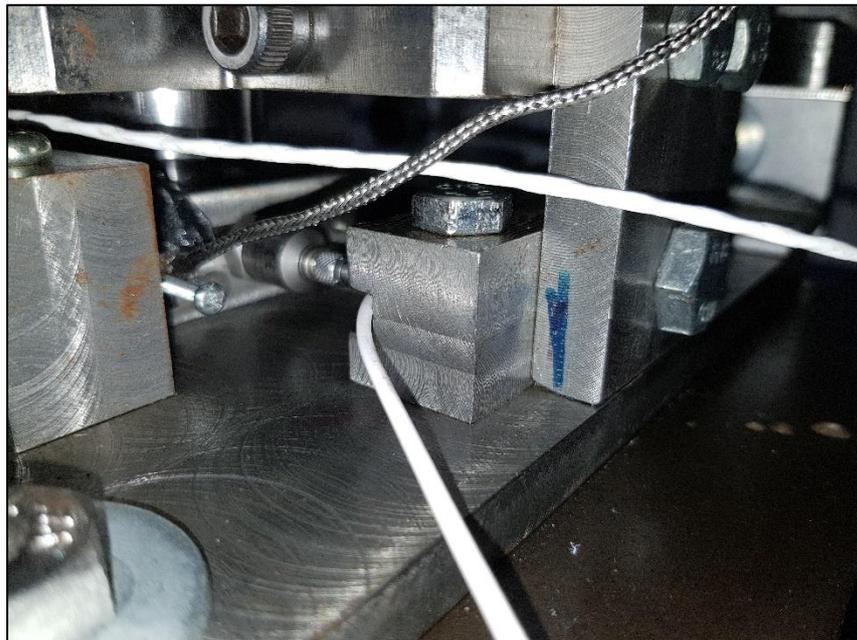

Figure 3.3: Test Bearing Housing and Sensors Connected To It

3.2.2   Data collection and data type

Data was collected through several sensors such as AE sensor, thermocouple and accelerometer. These data were collected in several types of formats and processed by different software. The data collected will go through three main stages, which includes signal processing, diagnostics and prognostics. Each type of data is equally important and plays its own role to increase the accuracy of predicting RUL of bearings. This technique works in such a way that it makes the state visible to users where the only parameter is state transition probabilities. As each state has a probability distribution over possible output tokens which allows this technique to produce results on the sequence of states.

3.2.2.1   Accelerometer

It is a sensor, which measures the proper acceleration. In other words, it is to measure acceleration of a body in its own instantaneous rest frame, which is different from coordinating acceleration. The accelerometer used in this study is KSTLER type 8704B50T, shown in Figure 3.4, which is suitable in the field between 0 to 10 kHz. This accelerometer was chosen because the bearings working and defective frequency are within the range. The software used for this study is LabView, which also stores data every 5 minutes interval. This software also serves the purpose of illustrating an FFT graph, which can pick up trails of defective frequency.

Figure 3.4: AE Sensor

3.2.2.2   Acoustics Emission

AE sensor serves the purpose of receiving and translating radiation of acoustic waves in the bearing when the material undergoes permanent changes in the internal structure.

The data are important because it can locate the source of the defects, evaluate material mechanical performance as well as monitor health of bearing. The model of AE sensor used in this study is MISTRAS AE sensor, shown in Figure 3.5. As the signals generated are not significant enough, it requires a preamplifier, which is at 40 dB differential mode. The software used to store data and display important information is AE Win. This software is also responsible for transforming the file format from DTA to CSV, which is compatible for MATLAB.

Figure 3.5: Accelerometer

3.2.2.3 Thermocouple

The function of a thermocouple is to measure the temperature of the test bearing housing from time to time. Although placing the thermocouple directly on the bearing will be desirable, but it is not possible. Therefore, thermocouple was placed somewhere near to the test bearing where the differences of temperature between the test bearing and the test bearing housing is negligible. After collecting the data from the thermocouple, software of Agilent 34970A acquisition unit, shown in Figure 3.6 will then be used to store the data for every five minutes interval. This software also serves the purpose of illustrating temperature against time graph. The temperature of bearing will indicate the bearing's condition as well as health status. When the bearing goes beyond a certain temperature, it is likely to fail soon.

Figure 3.6: Data Acquisition Unit

3.3 Proposed method

In order to choose the optimum method to predict remaining useful life of bearing, the considerations simply include the availability of data, the tools used, the codes available, the capability of hardware as well as condition of testing machine.

The proposed method will involve carrying out experiment continuously to collect data, develop signal processing, diagnosis, and prognosis algorithm. The collected data can then be filtered through signal processing technique, and diagnosis can be carried out to identify bearing fault type. Lastly, the data can be trained by prognosis algorithm and be applied onto online test.

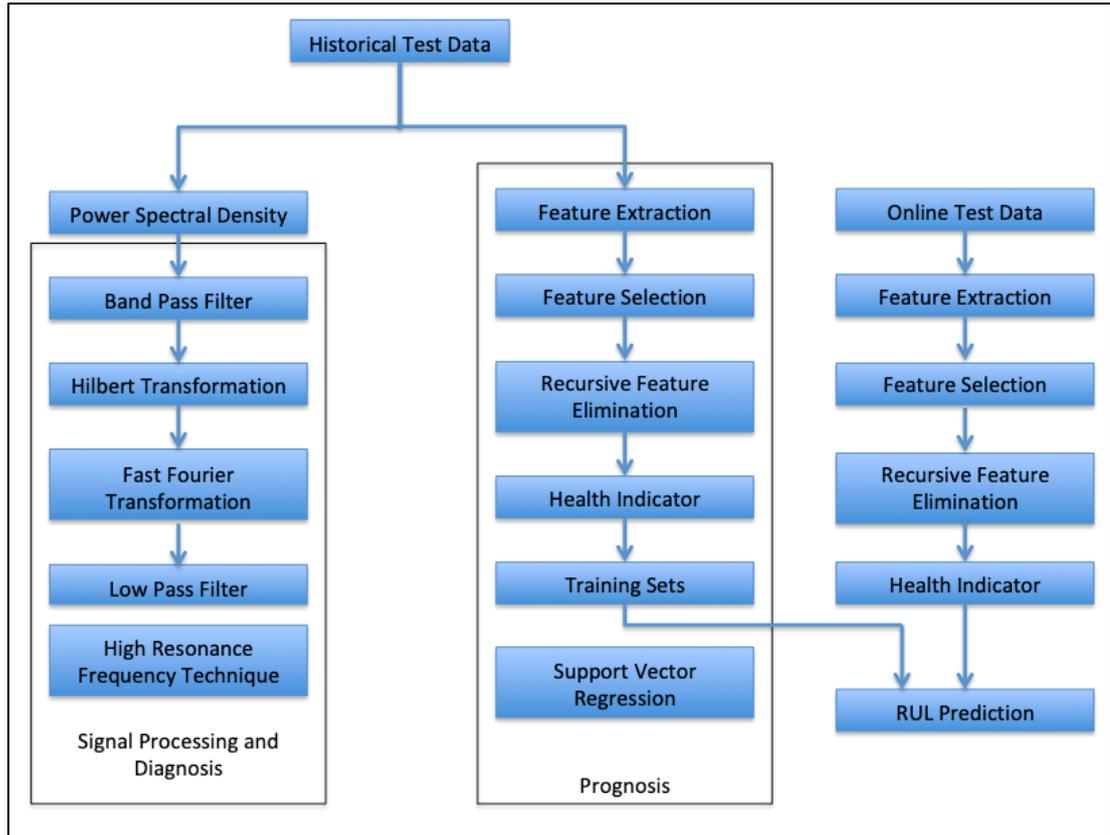The project flow chart is shown in Figure 3.7.

Figure 3.7: Project Flow Chart

3.3.1   Signal processing and diagnostics

As signal processing is a vital step to unmask the noise in the data, and diagnostics serves the purpose of identifying fault mode, high resonance frequency technique (HRFT) was embraced. This technique involves several steps in order to complete the analysis and these steps are

1. Power spectral density: it is the frequency domain analysis (transform time domain to frequency domain, to filter the noise)
2. Bandpass filter: used to filter mainly unwanted frequencies
3. Hilbert transformation and FFT: in order to obtain envelop of the signals
4. Lowpass filter: mainly passes low frequencies.

This technique works in a way that it captures the impulse of vibration generated by the bearing. As this impulse generated has several characteristics such as it happens in a very short duration of time and the energy of that impulse generated was widely distributed over a range of frequency, it simply makes it tougher for the other

techniques such as conventional spectrum analysis to capture it. This technique captures the impulse of vibration by detecting the high frequency generated, which is also the by-product of resonance. With the data extracted, it allows users to determine the fault mode by examining the irregularity of the graph.

Each technique applied in HRFT will be further discussed in the following subsection.

3.3.1.1 Power Spectral Density

Power spectral density (PSD) is a powerful technique when applied in signal processing to de-noise signals. Besides, it can also be applied onto several other fields such as transaction on speech and audio processing, geophysical fields, gravity and magnetic fields, which the data's collected will be extremely complex. The main reason why this technique was chosen rather than the other techniques is due to the complexity that this particular technique can handle. This technique's approach is also unique which produces high accuracy when compared to other techniques in terms of de-noising.

This particular technique is also known as a measure of signals power content versus frequency. Therefore, it first has to convert the raw data from time domain to frequency domain. After the conversion to frequency domain, the PSD can then outline the broadband random signals. Normalization will then be carried out by spectral resolution, which digitized the signals.

PSD is suitable to process continuous signal over time. One such example would be stationary process. A stationary process is a random process that can be shifted in time but yet its unconditional joint probability distribution remains the same. It was also noted that PSD describes power as either actual physical power or squared value of signal.

In this study, users are allowed to determine the amount of power a particular type of noise has in a particular frequency band. In other words, this technique plots a graph of power versus frequency and from the graph generated, users are able to determine the strength of the power at specific frequencies.

PSD plays an important role by providing users with the image of power at particular frequency. By having the results generated through PSD, users are able to

get the desired power generated by the bearing and at the same time eliminate the noise generated by impurities as well as vibration of test equipment.

After applying PSD, the raw data, which is in time domain could be transformed into frequency domain which can further filter the noise through band pass filter.

$$P_{(f_k)} = \frac{1}{N} \left| \sum_{n=0}^{N-1} x|N| \exp(-j2\pi f_k n) \right|^2 \tag{3.1}$$

Where

N= number of terms

3.3.1.2 Band Pass Filter (BPF)

Signal processing was advancing over the years, which can produce more precise results as more steps or processes were added in order to further de-noise the signals. There is no "one best technique" to filter all the noise, however a series of carefully planned techniques could do a decent job. After applying the technique PSD, there is still a reasonable amount of noise, which can be further reduced in order to obtain a much accurate result. By doing so, a band pass filter (BPF) will be added in this series of process to smoothen the data and at the same time further reduce the noise.

BPF is a simple but yet efficient technique to trim off data, which is proposed to be noise. Instead of trimming off the signal point one at a time which is also time consuming, this technique set a boundary for all data set. Any signal, which exceed or fall short of the boundary defined by the user, the signal will be labelled as noise and further eliminated. In the case of BPF, it does not require any complex mathematics equations but only a set of boundaries.

The boundaries set in this project would be the frequency between 80 kHz as well as 180 kHz. Any signal that overshoots 180 kHz or fall short of 80 kHz will be treated as noise and therefore eliminated. These parameters were chosen based on historical data test. After running this particular project for several months, it was learned that most of the data were within 80 kHz to 180 kHz. Therefore, it safe to assume that signal which lies beyond the boundary was indeed noise.

Even if the signal trimmed off is not entirely noise, it would not actually affect the final results much as the essence of the signal lies at the peak. Therefore, it would

be a great trade off when we compromise some information, which are not of essence and eliminate noise.

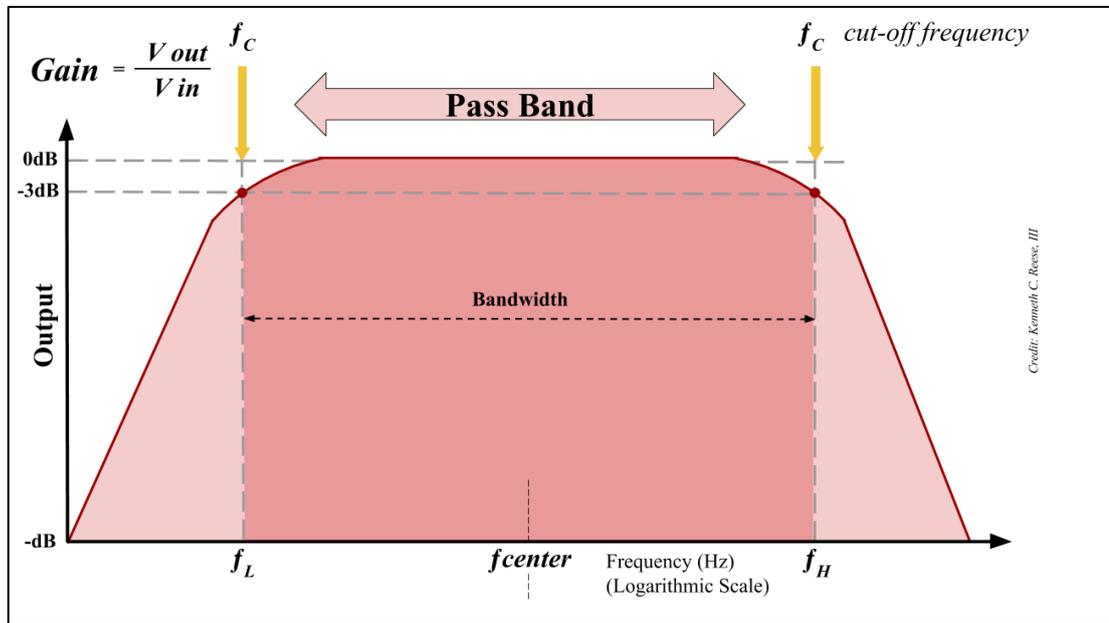A sample band pass filter is shown in Figure 3.8.

Figure 3.8: Sample Bandpass Filter

3.3.1.3 Envelope Detection (Hilbert Transformation)

The following technique of diagnosis after the band pass filter is generally called envelope detection or more specifically know as Hilbert Transformation (HT). HT provides a much simple understanding to its user when applied as it produces an easy understanding graph in terms of frequency domain. By doing so, this technique phase shifts every single Fourier function by 90 degree. In layman term, HT produces "a true amplitude" when applied onto a situation with stationary phenomena.

When real data were applied, the amplitudes generally fluctuate up and down the origin, which consists of negative amplitudes as well as positive amplitudes. These readings were difficult to understand or even visualize. By applying HT, it literally transforms the data into a consistent manner where initial information remains but yet the amplitude was converted into positive value. The outcome mentioned above is also known as analytic signal.

However, this filter to phase shift the data does not change the data from time domain to frequency domain. Instead, time domain was phase shift to time domain

where envelop detection could be done. Envelop detection is a process to envelop scattered form of signals ad amplify their output to provide a better illustration and understanding for its users. Figure 3.9 shows and envelope detection illustration.
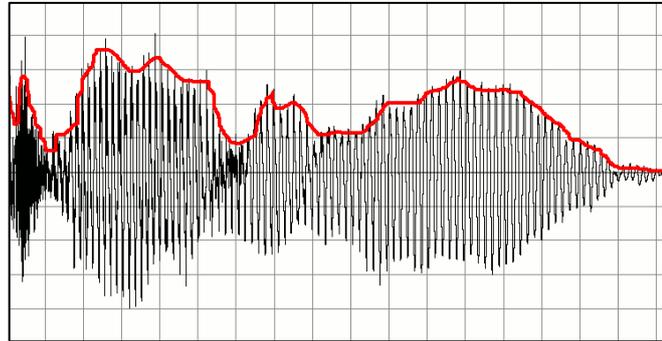

Figure 3.9: Envelope Detection

Above shows a figure which black represents the original signal and red represents the signal after envelop detection through HT. As it can be seen that the original data might be useful but users are not sure about the exact peak or fluctuation at a specific time. Therefore, envelop detection will envelop these scattered signals and amplify it to provide a better illustration which was presented in red. The Hilbert transform of fu can be thought of as the convolution of f(t) with the function h(x)=1/(πx), known as the Cauchy kernel.

The formula used in HT is shown as follow:

$$H(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} f(u) \frac{1}{x - u} du \qquad (3.2)$$

Where:

F(u) = real value function

H (x) = Hilbert transform of a real-valued function

du = distribution or a measure (not really a function)

3.3.1.4 Fast Fourier Transform (FFT)

FFT is a very useful technique for time dependent phenomena analysis and the main objective of this technique is to transform the test data from time domain to frequency domain. In this project, FFT serves as an algorithm that generates the discrete Fourier transform (DFT). It was known that computing DFT could be a very complex and time-consuming process. FFT is therefore unique in a way that it simplifies the process of generating DFT. When comparing FFT to the conventional DFT, the time taken to generate results is a lot lesser in FFT. Moreover, FFT offers round-off error which in turn produces much more accurate result than DFT.

However, FFT also plays the role in breaking down vibrational signal into individual frequency. By applying this technique, AE signal emitted from the bearing could be dissected into individual frequency to further determine fault mode of the bearing.

The formula of FFT is derivable from DFT. Below shows the formula of DFT.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{i2\pi kn}{N}} = \sum_{n=0}^{N-1} x_n w^{-kn} \quad k = 0, \dots, N-1 \tag{3.3}$$

Where:

X_n = complex number in general

$e^{-\frac{i2\pi kn}{N}}$ = N-th primitive root of unity

N= number of terms

X_k = DFT coefficient

FFT works in a way it breaks down N-point of DFT in terms of smaller length, which any factorization could be applicable. In this project, the algorithm of FFT used is known as "eight-point DFT". Eight-point DFT was chosen mainly due to the practicality to apply it in real data. The formula for this algorithm will be as follows:

$$X(k) = G(k) + e^{-j\left(\frac{2\pi}{8}\right)k \times 1} H(k) \tag{3.4}$$

Where

$$H(k) = x(1)e^{-j\left(\frac{2\pi}{4}\right)k\times0} + x(3)e^{-j\left(\frac{2\pi}{4}\right)k\times1} + x(5)e^{-j\left(\frac{2\pi}{4}\right)k\times2} \tag{3.5}$$
$$+ x(7)e^{-j\left(\frac{2\pi}{4}\right)k\times3}$$

$$G(k) = x(0)e^{-j\left(\frac{2\pi}{4}\right)k\times0} + x(2)e^{-j\left(\frac{2\pi}{4}\right)k\times1} + x(4)e^{-j\left(\frac{2\pi}{4}\right)k\times2} + x(6)e^{-j\left(\frac{2\pi}{4}\right)k\times3} \tag{3.6}$$

3.3.1.5 Low Pass Filter (LPF)

The low pass filter (LPF) has similar theory with the BPF. They both do not require complex mathematical equations but just a set of boundaries. The question here lies in "why LPF was included into the process when there already is BPF?" As what was mentioned above, a good or suitable signal processing technique requires a combination of different techniques to further smoothen and de-noising the signals. The BPF serves as the very first barrier to remove the general noise, which is easily identifiable. It can be noticed that the barrier set in BPF is within 80 kHz to 180 kHz, which is quite a large range. However, after applying BPF there are still ways to eliminate noise in the signals, which can further improve the results. HT converts all the signals into the positive region, which allow LPF's barrier to set just within positive range.

The process was made easier when it was learned that the results could be obtain just within a certain frequency range. Therefore, an LPF was put into use to trim off the data, which the results don't require. In other words, any signal beyond 500 Hz was eliminated.

The parameters set in the LPF is that the frequency lies within 0 to 500 Hz. This boundary was based on historical data and results as well. The main objective of applying low pass filter is to generate frequency spectrum and diagnose bearing's fault mode. The fault mode frequency could also be calculated which will be further discussed in chapter four. These fault modes are the BPFO, BPFI, BSF as well as FTF and they are the essence in diagnosis. Based on historical results, these fault mode frequencies do not exceed 500 Hz. Therefore, it also can be safely assumed that the future results would also lies within 500 Hz. Samples of Low Pass filter and its effect is shown in Figure 3.10 and Figure 3.11.
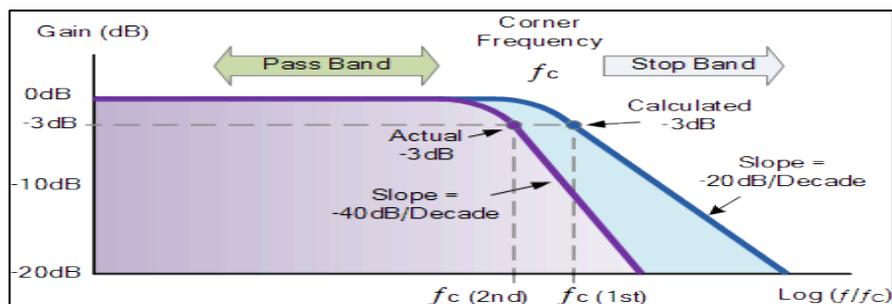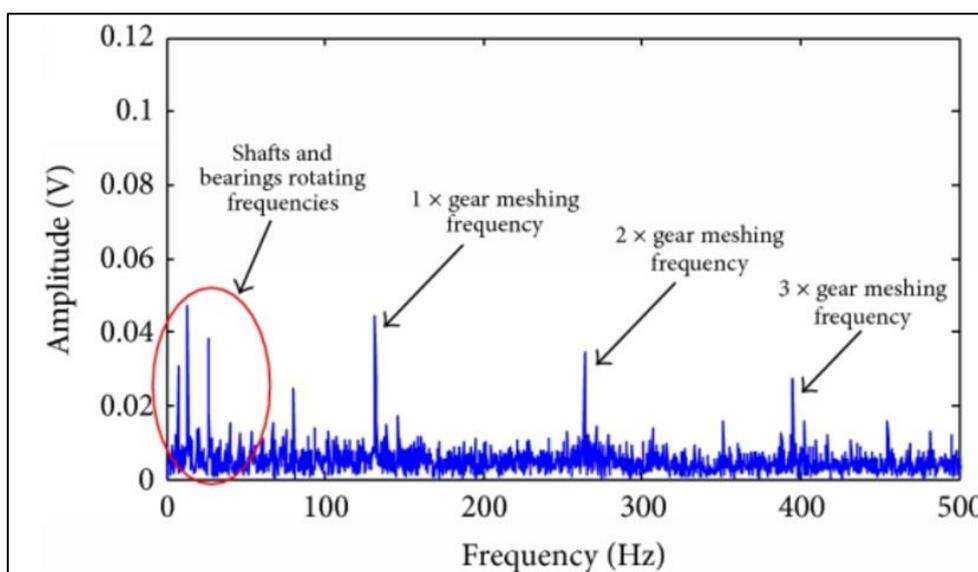
Figure 3.10: Low Pass Filter

Figure 3.11: Sample Low Pass Filter of Rotary Machine Diagnosis

Why is the bandwidth 0Hz to 500Hz? This is because we apply the standards of rotary machines diagnosis.

3.3.1.6 Sample HRFT Process of Test 17

In order to understand this technique better, the result of test 17 generated after applying each technique in HRFT will be illustrated below:

It was noted that these results generated were based on the results of test 17 after the bearing failed, where the graphs generated are in Figure 3.12 to Figure 3.15.
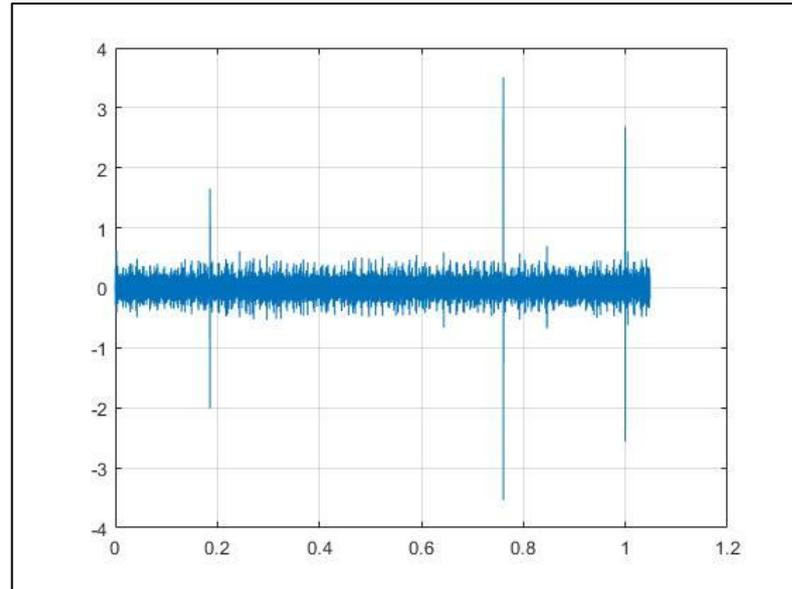
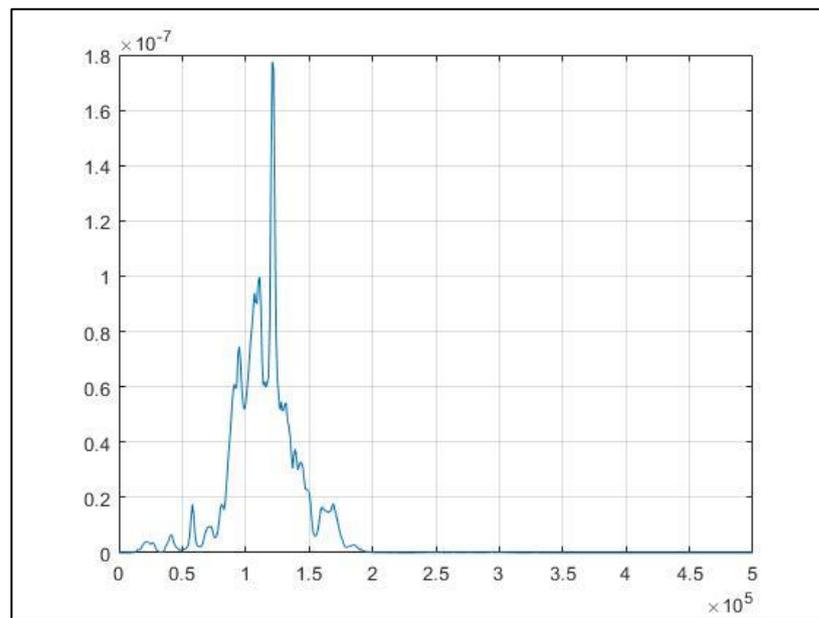Figure 3.12:Raw Signal From AE Data

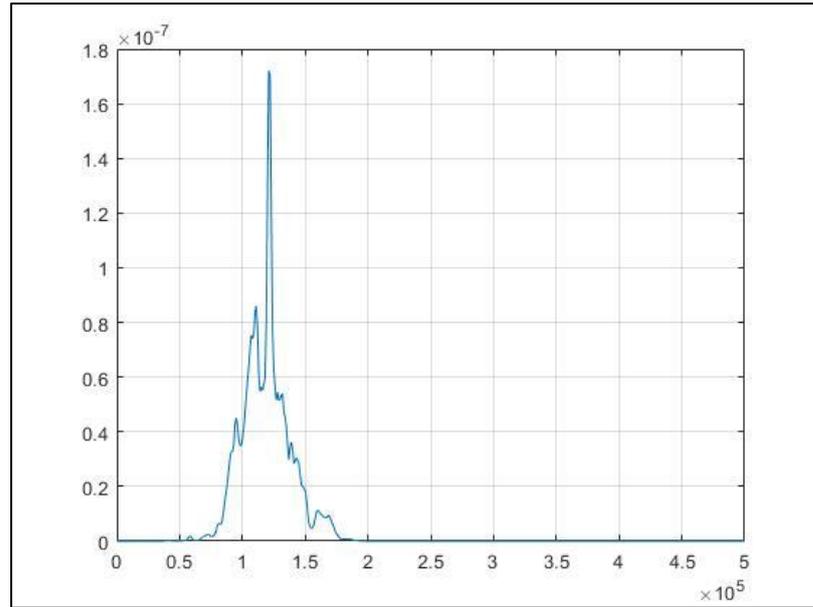Figure 3.13:After Applying Power Spectral Density Technique
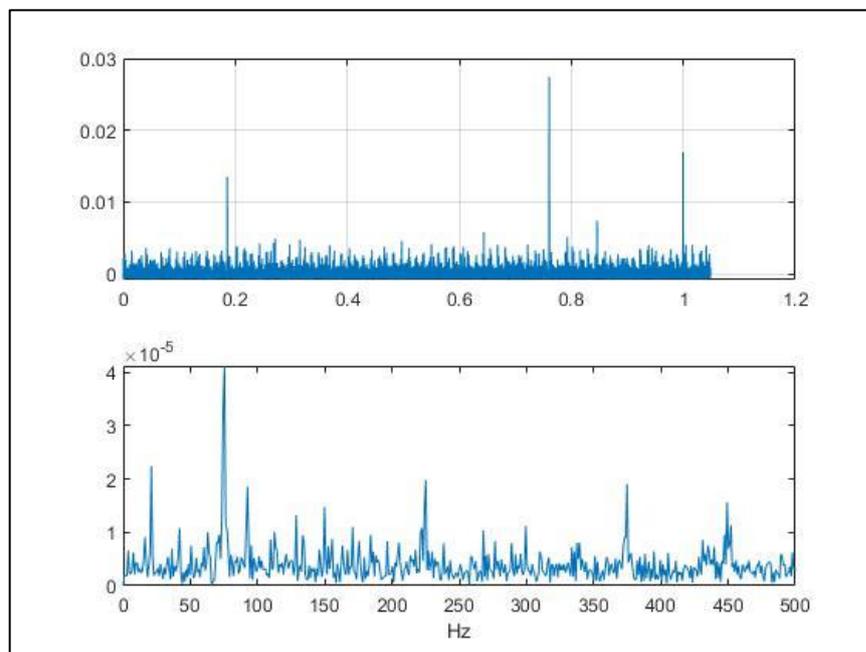
Figure 3.14:After Bandpass Filter

Figure 3.15:After applying Hilbert Transformation, FFT and Lowpass Filter

3.3.2 Feature extraction and Selection

3.3.2.1 Feature extraction

There are many feature extraction techniques, which were shared by researches over the world but the techniques used in this particular project are root-mean-square (RMS), kurtosis, skewness, crest factor and mean. There are also many other feature extraction techniques discovered by researches such as peak, time-domain kurtosis and so on.

These techniques allow user to view the amplitude of respective feature versus time. Each graph produced allow user to view the data from a different perspective. It is also important to note that it is a vital step to look into individual graphs, which allow users to identify trends, irregularity as well as non-uniform peak. By doing so, users are able to define a carefully analysed weightage in feature selection. The technique to distribute weightage of each feature in health indicator will be further discussed in the feature selection section. Below shows formula of RMS as well as kurtosis

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x(i)-\bar{x})^2} \tag{3.7}$$

$$Kurtosis = \frac{\frac{1}{N}\sum_{i=1}^{N}(x(i)-\bar{x})^4}{RMS^4} \tag{3.8}$$

$$Sk = \sum_{i=1}^{N}(\frac{\bar{X}_i - Mo_i}{s}) \tag{3.9}$$

$$C = \frac{|x_{i,peak}|}{x_{i,rms}} = \frac{\|x\|_\infty}{\|x\|_2} \tag{3.10}$$

$$Mean = \frac{1}{N}\sum_{i=1}^{N}(x_i) \tag{3.11}$$

Where

$\bar{x}$= Mean value of signal

Mo= Mode

N= Sample size

s= Standard deviation for the sample

x$_i$ = amplitude of signal at specific time

$x_{i,peak}$= Peak amplitude of signal at specific time

$x_{i,rms}$= RMS value of signal at specific time

3.3.2.2 Feature Selection

3.3.2.2.1 Brief Introduction and Comparison with Other Technique

Feature selection will then continue the task of feature extraction to produce a health indicator. Although there are many different features selection techniques available, neighbourhood component analysis (NCA) was chosen in this project.

First of all, NCA is classified in the category of "supervised machine learning (SML)" group. A SML indicates that the data will have both input variables as well as output variables. Unlike unsupervised machine learning (UML) which simply only have input variables without output variables. An example of UML technique, which is similar to NCA would be principal component analysis (PCA) Both techniques are equally vital in the industries depending on its working condition. A SML technique works in a way that user emphasizes on a certain algorithm that is able to learn the mapping function which the input was mapped to the output. By choosing the suitable algorithm which is able to relate a set of reasonable data, this technique can predict the new output variables when new input variables were plugged in. In other words, the algorithm emphasized by the user will learn from historical data, which is the input variables and iteratively predicts output variables. This process will continue until the percentage error of the technique is within acceptable range.

SML is a vital step in this project as support vector regression (SVR) was chosen as the prognosis technique. This is mainly because SVR is also classified in the SML group. SML outperform UML when it comes to classification problem and regression problem. Regression problem can be defined as problem, which its output variable is of real value.

3.3.2.2.2 Brief Theory of NCA

NCA is usually used for the purpose of distance metric learning. NCA is able to perform a reduction in dimension with the choice of learning k X d matrix rather than d X d matrix. K in this context simply imply to a new dimension of choice whereas d imply to the original dimensions. The uniqueness of NCA is simply because it can have higher adaptation to various kind of data and at the same time it is able to bring forward more degradation information. The general formula used in NCA is as follow:

$$Y = (A)x_1 + (B)x_2 + (C)x_3 + (D)x_4 + (E)x_5 \qquad (3.12)$$

Where:

A, B, C, D, E = Weightage assigned for each feature

$X_{1...5}$ = the amplitude of each feature

Therefore, once the user analyses the weightage of each features, NCA will then play its part in distributing weightage of each feature as well as combined them all together to produce a health indicator. If the weightage of a particular feature defined by user is higher than the other, the technique will simply emphasize more on that particular feature in its final product. In other words, feature that has higher weightage than the other will be highlighted in the health indicator produced.

3.3.2.2.3 Method of Analysing Weightage

The method used to analyse each weightage in the features will be Recursive Feature Elimination (RFE). This method is however suitable to apply in small samples. This method works in a way it strives to improve generalization performance through emphasizing on important features and eliminating weightage of unimportant features. Chen et al., (2014) claims that RFE has similar accuracy compared to Enhanced Recursive Feature Elimination (EnRFE) when the number of features is the same. It was also claimed that the accuracy of the RFE goes as high as 74%.

3.3.2.2.4 Brief Theory of How RFE Works

RFE serves three purposes in analysis weightage of each feature. The first purpose is producing a better performing model, which provides higher accuracy. The second purpose is easier to understand model, which helps machine-learning process faster, and better decision making of prognosis algorithm due to carefully analysed weightage. The last purpose however is to allow the model to runs faster. This can be explained in terms of computational time for prognosis process. With the right weightage distributed across the health indicator, it saves computational time, as the feature weightage does not contradict with the final data.

Others often misinterpret how this technique works as it was assumed that this technique assigned higher weightage to important features. In fact, this technique's primary objective is to eliminate weak features at each step followed by resequencing the importance of strong features, which were not eliminated. In other words, RFE removes weak features depending on the trails that particular feature picked up.

This technique could be performed in three simple steps which is first include all variables (feature) in the model. The second step is to drop the least useful variable which can also be said in terms of smallest fluctuation of amplitude in the model accuracy. The last step would be continuing the second step until a predefined criterion is satisfied.

However, some weak features do carry important details that would very much benefit the prognosis. This happens when strong features did not pick up specific trails on bearings characteristics whereas weak feature does. Although the amplitude of the weak feature might be low, but the details picked up were crucial. Therefore, removing weak features may sometime result in inaccuracy of classification.

The original formula used in RFE would be shown below:

$$w = \sum_{keSV} y_k a_k x_k \tag{3.13}$$

Where:

a_k = Lagrange coefficient

W= weightage

X = training data

Y is corresponding to the target where $y_i = \{\pm 1\}, i = 1, \dots, m$

However, this project deals with non-linear problem and the formula used above was modified to fit the condition. The formula used for non-linear problem would be as follows:

$$w_i = \frac{1}{2} a^T K \alpha - \frac{1}{2} a^T K(-i)\alpha \qquad (3.14)$$

Where:

K(-i) = kernel matrix by eliminating ith features in input x

The Lagrange coefficient could be found through cross validation of the 70/30 rule where 70 percent of the data were used as the training set and 30 percent of the data act as validation set. It is a parameter that could be generated through the collected data itself without users assigning value. However, the data to be applied must be of huge sample size to prevent overfitting problem or large variance. Figure 3.16 shows how RFE works.
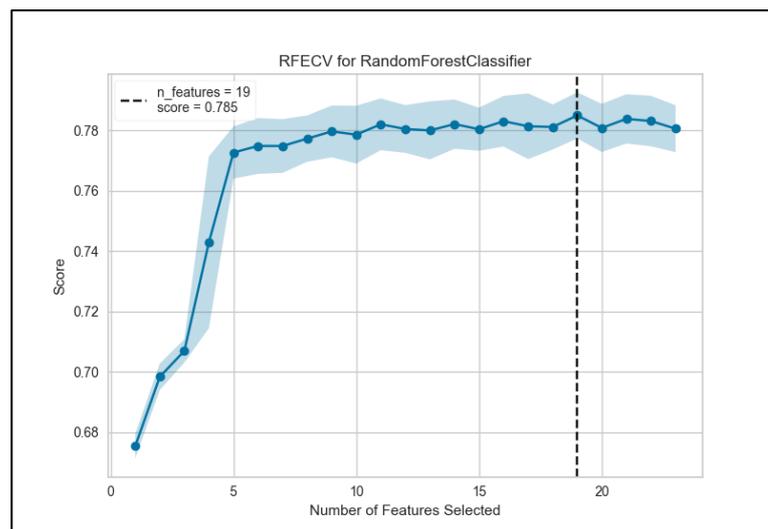

Figure 3.16: Sample of How RFE Works

3.3.3 Prognosis

In this project, two main objectives include developing a diagnosis technique as well as develop prognosis technique and algorithm. Diagnosis are as stated above where this technique captures the failure of bearing for which it could be further studied and

optimised in the prognosis process. Diagnosis can only be carried out when the set of data is complete where defective signals are present and adequate. After having the sets of defective signals as well as trends, prognosis technique could be applied to train the prognosis algorithm. The very initial step in prognosis is to apply the "run to failure test" data in order to train the algorithm recognises historical failing point. Once the training is complete where the test results are above acceptable range, the trained algorithm shall then be applied to the on-line bearing test. By carrying out on-line bearing test, the algorithm trained will then be above to make predictions of RUL of bearing as well as failing point of the bearing.

Prognosis can then be classified into three categories, which are data based prognosis, physics-based prognosis as well as hybrid-based prognosis which were discussed earlier in literature review section two. The technique chosen in this project to perform prognosis will be support vector regression (SVR), which lies in the multi-step time-series forecasting group. Multi-step time forecasting can be further classified into four different groups, which are direct multi-step forecast strategy, recursive multi-step forecast, direct recursive hybrid strategy and multiple output strategy. It is important to take note that SVR has the nature of direct recursive hybrid strategy, which will be further explained and compared in the following paragraph.

A direct multi-step forecast strategy revolves around creating unique model for each forecast time step. However, each unique model develops suggest no dependencies between predictions. A recursive multi-step forecast implements one-step model several times where further prediction could be made based on prior step input. However, predictions error will be built up or accumulated in a long run, which will affect the performance of forecasting. A direct recursive hybrid strategy is the combination of those two models mentioned above which can produce both their benefits. A multiple output strategy on the other hand develops one model, which is able to provide the entire forecast. However, it is highly complex which will further lead to the requirement of many data in order to refrain from over fitting the condition as well as slower training speed.

3.3.3.1 Support Vector Regression

3.3.3.1.1 Brief Explanation

SVR is a useful prognosis technique when applied in bearing fields as this particular method contains self-learning algorithm. For instance, the bearing experiment carried out in the lab has many variables which sums up to its failure. An alternative for self-learning algorithm will be developing different model or code for each variable which is not practical as well as time consuming. This is because each variable is specific and the time taken to develop each of them will be lengthy. Whenever a new variable arises, a new model must be developed which is also troublesome. Moreover, some of the variables were still lack of studies on that particular field which couldn't be developed properly and precisely. Although physics-based prognosis is well known for its accuracy when the number of models were sufficient to cover each variable and each model developed were accurate and precise which fits the condition of its experiment. However, it is usually advised to apply in a working condition where less equipment's were used and a working condition, which remains constant over a long period of time. This is mainly because whenever the working condition changes or whenever extra equipment were added, new model must be developed for each additional variable.

Self-learning algorithm however will learn each and every variable, which contributes to the bearing failure through historical data. By doing so, the historical data will provide SVR with examples of when bearing is going to fail as well as how the bearing is going to fail. This simply involves early indication of failures in the features as well as the lifespan of each bearing from past data. This can save up a lot of time and unnecessary development of new model. SVR might not be as accurate as a physics-based prognosis but it is able to cover many different fields and it is also highly flexible. It is suitable to apply in bearing field as it has a reasonable amount of variables.

3.3.3.1.2 Comparison with Other Techniques

SVR has very similar concept compared to support vector machine (SVM). The differences between SVR and SVM would be the slack variables used. In SVR, each training data point will be assigned with one slack variable whereas SVM assigns two slack variables to each training data point. SVM is mainly used in classification as well

as discrete categorial labels prediction whereas SVR uses a regression algorithm that predicts continuous ordered variables.

3.3.3.1.3 SVR's Basic Theory

SVR works in a way that every data point was plotted in several dimensional space. The number of dimensions highly depends on the number of features available. Each different feature will then be treated as a value of particular coordinate and classification will be performed to produce a hyperplane. A hyperplane can be considered as a boundary, which simply separates or differentiate between two different classes. Support vectors will then be introduced which can also be defined as data points that the margin pushes up against all points that are close to the opposing class. This technique focuses only on support vectors whereas training examples are literally ignorable. A boundary line can then be plotted based on the support vectors, which creates a margin. A boundary line has similar function as a hyperplane but it simply acts as a tolerance limit or threshold for data separation. Lastly, a very unique approach in SVR, which is a very important element in prediction, would be kernel. A kernel is unique in a way that it serves as a function to transform a lower dimensional data into higher dimensional. The number of dimensions can be infinite as users are able to define as many features as possible. By transforming from a lower dimension to higher dimensions, this technique is able to plot an optimum hyperplane, which separates classes in a clear cut. Figure 3.17 and Figure 3.18 show some basic working principles of SVR and margin.
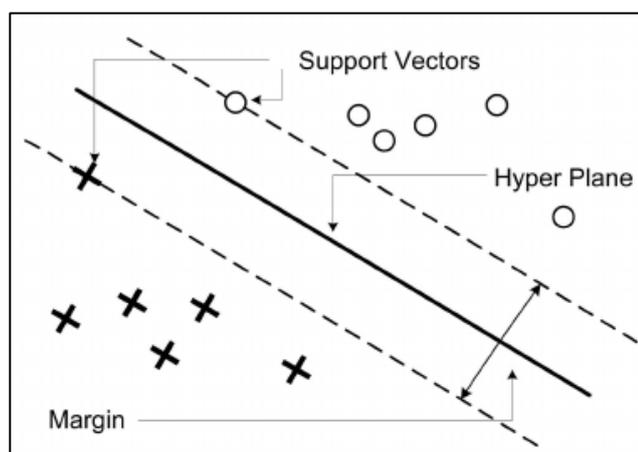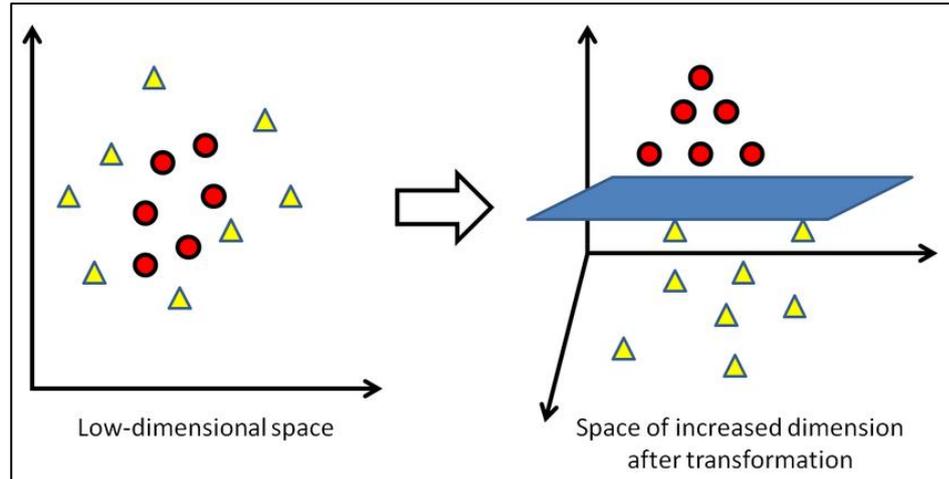

Figure 3.17: Label of Elements in SVR

Figure 3.18: Illustration of How a Kernel Function Works

3.3.3.1.4 SVR Technical Part

Technically, SVR consists of several main parts in order to generate results. The very first part would be providing a way in terms of training data's small subset to increase the computational advantages. This step could be further discussed through derivation for formula used.

$$f(x, \varpi) = \sum_{j=1}^{m} \varpi_j g_j(x) + b$$

(3.15)

Where:

X = input

m = dimensional feature space

b = bias term

$g_j(x), j = 1, ..., m$ denotes a set of nonlinear transformations.

$f(x, \varpi)$ denotes a linear model

In this particular formula, it works in such a way that the input will be mapped onto several dimensional features depending on the m value. Once it is completed, a linear model will be generated. It was noted that the bias term is usually zero and could be ignored.

The second formula, which plays an important part will be the loss function and it serves the purpose of measuring quality of estimation.

$$L_\in\left(y, f(x, \varpi)\right) \tag{3.16}$$

$$= \begin{cases} 0 & if \ |y - f(x, \varpi)| \leq\in \\ |y - f(x, \varpi)|-\in & otherwise \end{cases}$$

Where:

$\in$ denotes the insensitive loss

By revising the formula of loss function, empirical risk formula could be deduced as follow:

$$R_{emp}(\varpi) = \frac{1}{n}\sum_{i=1}^{n} L_\in\left(y_i, f(x_i, \varpi)\right) \tag{3.17}$$

Insensitive loss $\in$ were being used in higher dimensional space in order for SVR to perform a linear regression. However, reducing the complexity of the model is also important which can be performed through minimizing $\|\varpi\|^2$. Fortunately, a simple way of minimizing $\|\varpi\|^2$ would be adding on slack variables into measuring training sample's deviations beyond the insensitive loss zone. This is can be finalized into the below minimization functional:

$$\frac{1}{2}\|\varpi\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{3.18}$$

$$\min$$

$$\begin{cases} y_i - f(x_i, \varpi) \leq \varepsilon + \xi_i^* \\ f(x_i, \varpi) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1, \dots, n \end{cases}$$

Where:

C>0

If the problem is transformable, the formula applied in the dual problem shall be

$$f(x) = \sum_{i=1}^{n_{EV}} (\alpha_i - \alpha_i^*) K(x_i, x) \quad 0 \le \alpha_i^* \le C, 0 \le \alpha_i \le C$$

(3.19)

$$\underset{s.t.}{}$$

The main idea behind transforming an optimization problem to a dual problem is to maximize the objective function based on the dual variables assigned.

Lastly, one of the very main function in SVR would be the kernel function which has the following formula:

$$K(x, x') = \exp(-\frac{\|x - x'\|^2}{2\sigma^2})$$

(3.20)

Where:

$\Upsilon = \frac{1}{2\sigma^2}; \Upsilon > 0$

K = radial basis function (RBF) kernel function

X = vector

X' = point of reference vector

Exp () = math constant e raised to a power

$\|\quad\|$ means Euclidean distance.

σ is a free parameter which will be further discussed in the discussion of parameters

3.3.3.1.5 Parameters of SVR and How They Are Assigned

In the prognosis part, there were only two parameters that users should assign. These two parameters are the penalty parameter, which is C as well as the kernel parameter σ. The technique used in assigning the optimum parameters are cross-validation as well as grid-search.

In simple terms, cross validation could be explained in terms of splitting a set of training data into two or more subsets. A minimum of one subset will then be reserved to evaluate the model created through the other subsets trained (subsets that were not reserved). Grid search on the other hand could be explained in terms of trying out several combinations of parameters and determined pairs of parameters that

performed best. In order to determine which combination of parameters could perform best, cross validation could be done.

In this project, cross validation was done based on a 70/30 rule. 70/30 rule indicates that 70 % of the data will be trained whereas 30 % of the remaining data will evaluate the model created through the 70 % of the trained data. At the same time, grid search of this project was done based on different combination of C as well as ϒ. Rightfully, the combination of parameters should be systematically computed. However, systematically computing parameters would be exhaustive and time consuming (trying out each value would probably cost a week or more). Therefore, a generalized combination of parameters would be applied where the combination of the parameters would revolve around four values, which is one, ten, one hundred, as well as one thousand (for example C= 1, 10, 100, 1000, ϒ = 1, 10, 100, 1000). The combination with the highest accuracy would be chosen to compute the remaining steps. Hsu et al., (2016) claimed that a generalized combination of parameters could prevent lengthy computational time. It was also claimed that generalized combination of parameters could provide high accuracy if the number of parameters were limited (within two to five parameters).

3.3.4   Accuracy

In order to know how well the prognosis technique was progressing, determining the accuracy of the result is important. There are several ways of determining the accuracy such as calculating the standard deviation, mean square error, squared correlation coefficient and so on. In this project, the accuracy of the results was determined through the conventional accuracy formula as follows:

$$Accuracy = \left(\frac{(Y - X)}{X}\right) \times 100\%$$
(3.21)

Where:

X = real RUL of bearing

Y = predicted RUL of bearing

3.4   Precaution

The main reason of taking precaution is to prevent experimental set up from failing as well as avoiding random errors. Therefore, the precaution taken is mainly to assure the

experiment was carried out in a controlled condition where random error could be minimized. Precautions can be separated into two different sections. The two sections include precaution for carrying out experiment, as well as precaution for bearing removal and installation.

3.4.1   Precaution for Carrying Out Experiment

1. The bearing must be lubricated twice a day, which in other words mean lubricate once every 4.5 hours. This is to keep the bearing under controlled condition.

2. The experiment set ups such as bearing's housing should always be tightened properly to prevent unwanted vibrations.

3. AE sensor, thermocouple as well as accelerometer should always be intact with the bearing's housing to maximize signal detection process.

4. After each experiment, surface of AE sensor should be applied with a layer of Vaseline to maximize the contact surface between the sensor and the bearing's housing.

5. Load should only be applied once the bearing starts to rotate, at the same time load should be released before the test rig is shut down.

6. After each experiment, all software should be thoroughly checked to make sure the parameters set fulfilled the experimental condition.

7. Calibration of the load should also be done once a month to ensure the amount of load exerted onto bearing is accurate.

3.4.2   Precaution for Bearing Removal and Installation

1. Always seek for lab-staff permission before using hydraulic jack (sometimes the hydraulic jack might be under maintenance).

2. During the bearing removal process, the bearing should be removed slowly and at the same time constantly lubricated to prevent exerting a shock on the shaft which might further leads to shaft failure.

3. The force exertion of the hydraulic jack during bearing installation or removal should always be parallel to the axis of rotation of the shaft to prevent shaft failure.

4. The process of bearing installation and removal should always be done when the shaft is cooled down to room temperature. In other words, it is not

advisable to change the bearing right after the bearing fails. The reason behind is to prevent deformation of shaft because high temperature of shaft causes it to be prone to permanent deformation.

5. Support bearing should also be lubricated after every test bearing's failure. The reason not to lubricate the support bearing as frequent as the test bearing is because the support bearing has higher threshold compared to test bearing. However, lubrication of support bearing is still mandatory on a weekly basis.

6. Debris generated after each experiment should be cleaned thoroughly (especially in the bearing's housing). This is to make sure the experiment is under a controlled condition.

7. It is also important to prevent impurities from entering the test bearing during bearing installation process.

3.5   Summary

The entire methodology evolves around doing run-to-failure test repeatedly to acquire as much data as possible, performing signal processing and diagnostics to minimize noise as well as determine fault mode by high resonance frequency technique, performing feature extraction to detect irregularity in data through neighbourhood component analysis, distributing weightage of each feature through recursive feature elimination technique, combining all graphs from feature extraction to obtain health indicator and finally performing prognosis to predict remaining useful life of bearing using support vector regression. The steps above are then repeated consistently to understand and learn further on bearing's condition under certain circumstances.

CHAPTER 4

RESULT AND DISCUSSION

4.1   Bearing Lifetime and Causes of Premature Bearing Failure

Throughout final year project one and two, there were all together more than 10 sets of run-to-failure bearing test conducted. Out of all the test sets carried out, a minority of the test couldn't be presented due to early failure, which is not desired as well as not expected. In order to explain the early failures mentioned earlier, there is a high possibility of manufacturing defect, which serves as the main cause. It was suspected that some of the bearings used for test might contain flaws within it. This causes the bearing's shortened life span, which are not predictable as well as not desired. For example, test 18 shows that the bearing only lasted for less than a day before it fails. Do take note that all the parameters, precautions as well as procedures to carry out the run-to-failure bearing test remains the same for all test.

Another explanation for early bearing failure would be flaws in the technique of bearing installation. Due to cost effectiveness, the changing of test bearing does not require high technology machineries but a conventional hydraulic jack instead. The bearing installation process also did not involve any sensors or any means of measurement for the sake of its alignment and depth (depth of bearing to shaft) installed. The bearings were installed manually based which precision standards were based on naked eye examination. With the given condition, there are several possibilities that might lead to premature bearing failure. The first possibility that can be concluded will be a misalignment between the shaft and the bearing, which is also further shown below in Figure 4.1.
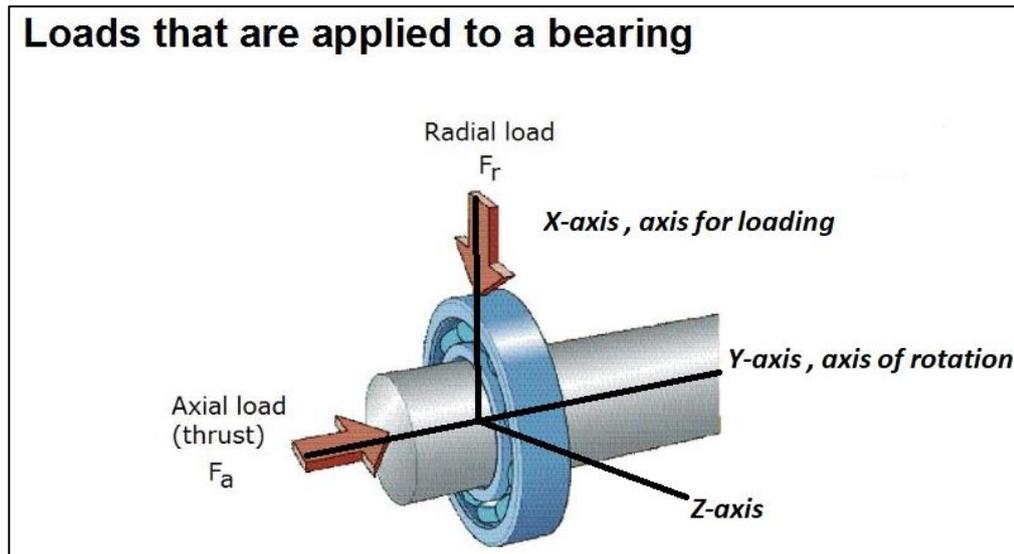
Figure 4.1: Application of Loads on Bearing

In order to simplify this issue, the alignment of the bearing to shaft can be classified into two different axis which is X axis as well as Y-axis. The X-axis is mainly for loading purposes whereas the Y-axis serves the purpose of rotation. However, the experiment conducted in real world condition can never be in ideal state. In fact, there is several more axis of motion, which contributes to premature failure. To simplify this, it is assumed the practical case have only one extra Z-axis motion. By doing so, the problem could be minimized as other axis motions were classified into only Z-axis. The Z-axis motion is the main cause to premature bearing failure as it promotes the motion of wobbling. This is mainly because bearing wasn't designed to withstand any other forces aside from X and Y-axis. In short, the force from Z-axis induces more friction, heat as well as asymmetrical force distribution. Therefore, the bearing fails in a shorter period of time.

The second possibility that could lead to premature bearing failure would be over exertion of force during bearing installation. As mentioned above, bearings were installed manually where the amount of force exerted was based on a rough estimation. By right, the bearing should be force fitted slightly onto the shaft, which might cause minimal damage on the bearing's outer casing. This will not actually harm the bearing and even if it does, the actual lifetime of bearing wouldn't differ much. However, the scenario changes drastically when over exertion of force during bearing installation occurs. When the force exerted exceeds the threshold of what the bearing could absorb, the bearing's internal as well as external cage would deform and this can directly affect

the ball movement within the bearing. This creates a lot more friction, heat as well as asymmetrical force distribution within the bearing. Thus, the bearing will soon experience premature failure.

There is also a possibility that the bearing experienced premature failure due to manufacturing defect. Although the brand of the bearing used in MISUMI, which is a trustworthy brand, however, manufacturing defect does happen sometime. Table 4.1 shows the table of test and its experimental lifespan. Note that all tests have constant load as well as constant speed of shaft, with a load of 3 kN and speed of 1250 RPM.

Table 4.1: Test Number and Lifespan

Test	Lifespan (minutes)
17	1770
18	185
19	3995
20	1150
21	825
22	785
23	1757
25	1055
26	1040

4.2   Load of Bearing

This section will be discussing why the load of bearing throughout this experiment was fixed at 3 kN and this section. First of all, the bearing used throughout the experiment is MISUMI B6002 and this particular bearing has a basic load of 2.83 kN. The term basic load stated above simply implies the optimum-loading load relative to its lifespan. In other words, the bearing can perform best under 2.83 kN in terms of load effectiveness.

In order to verify the above details provided by MISUMI dealers, several tests were conducted. The very first test's parameters were fixed with 1250 RPM as well as 2 kN bearing load. The test lasted for more than a month and it is not favourable. This is because the time is of essence and this project focuses more on shorter duration

experiments. In order to reduce the lifespan of bearing, the load applied to the bearing should be higher than its basic load and this will effectively shorten its lifespan. Therefore, a new parameter of 3 kN and 1250 RPM was implemented. As a result, the bearings lifespan was greatly reduced to a desired level.

4.3   Technique Used to Deal with Self-Healing

Self-healing is a process where bearing's health state shows improvement without any external aid. First of all, when bearing is put to work, there are several factors, which causes the bearing's health state to degenerate. The very main cause of most machines as well as bearings fails is due to the friction force that cannot be eliminated. Although bearing may reduce a tremendous amount of friction, however, not all friction was eliminated. Therefore, friction produces heat that further degenerate the parts within the bearing such as its cage, ball, inner ring as well as outer ring. If the bearing works for a reasonable amount of time, the heat generated will be sufficient to cause permanent deformation in the bearing, which directly affects the bearing's health state. However, if there is an idle time in between the experiment, this will prolong the RUL of the bearing. In other words, this is also known as self-healing of bearing.

As this project was carried out in the lab of university and the university policy does not allow non-stop run-to-failure test, there is approximately 15 hours of idle time in between each session of nine-hour experiment. The idle time allows the bearing to cool itself down which prevents it from further permanent deformation that will lead to degrading health state. These phenomena are not desire in this project, as it doesn't tally an industrial environment (industrial machineries run 24/7). On the other hand, this situation might also cause the results to be inaccurate as there is a constant fluctuation not only in the AE signals emitted, but also fluctuating health state of bearing.

In order to overcome this problem, the data collected must be filtered. In other words, the self-healing data will be eliminated to obtain much more accurate results. Riaz et al., (2017) claims that root-mean-square feature and crest factor feature has been popular when diagnosing bearings failure. Therefore, the trimming of the data was based on the RMS feature plotted and the standards set to alter the data is to reduce the constant fluctuation in the RMS feature. For example, the RMS of the bearing starts from zero and reaches two in the first day, the RMS of the bearing starts from one

during the second day due to self-healing characteristics of the bearing. Therefore, user should not include the second day's data before it reaches two (two is the amplitude where the bearing left off on the first day). This allows a linear graph to be plotted, which is also much alike compared to non-stop run-to-failure test. Graphs of RMS before and after self-healing technique are shown in Figure 4.2 and 4.3.
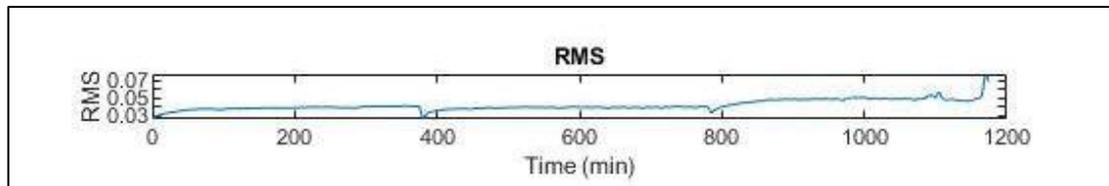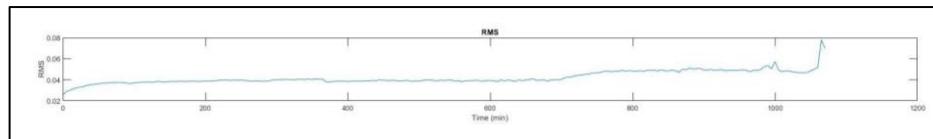

Figure 4.2: Graph Before Self-Healing Technique

Figure 4.3: Graph After Self-Healing Technique

4.4   Impurities

Impurities mentioned in this project simply imply certain contents, which is not desired in this experiment. By having impurities, it causes deviations in results, as the condition of the experiment would have a higher degree of differences compared to ideal conditions. Moreover, impurity is a general term, which may have many variables. Therefore, it is not possible to include all the impurity variables into consideration nor generate any mathematical models, which fits impurities precisely into it.

The common impurities, which can be distinguished, would be dust particles, iron debris, as well as contaminated lubrication. For example, dust which are trapped within the bearing's ball, iron debris that is situated between the metal housing and bearing's outer ring, as well as contaminated lubrication all over the bearing. These impurities will shorten the lifespan of a bearing, as it will cause an increment in friction generated and therefore inducing more heat. With this condition, it promotes deformation of inner and outer ring as well as wear and tear on the ball and cage.

The impurities were simply shown in all the graphs presented. It might have produced a slight fluctuation in the results, however it causes deviations in the final results. The

amount of impurities was usually low in each set of data as clearing impurities is part of the standard operating procedure. However, if the impurities level exceeds the acceptable range, it will cause a drastic reduction in bearings lifespan.

4.5 Diagnosis

4.5.1 Feature Extraction

Although there are many features within this experiment to be extracted, the five features shown in Figure 4.4 are emphasized.
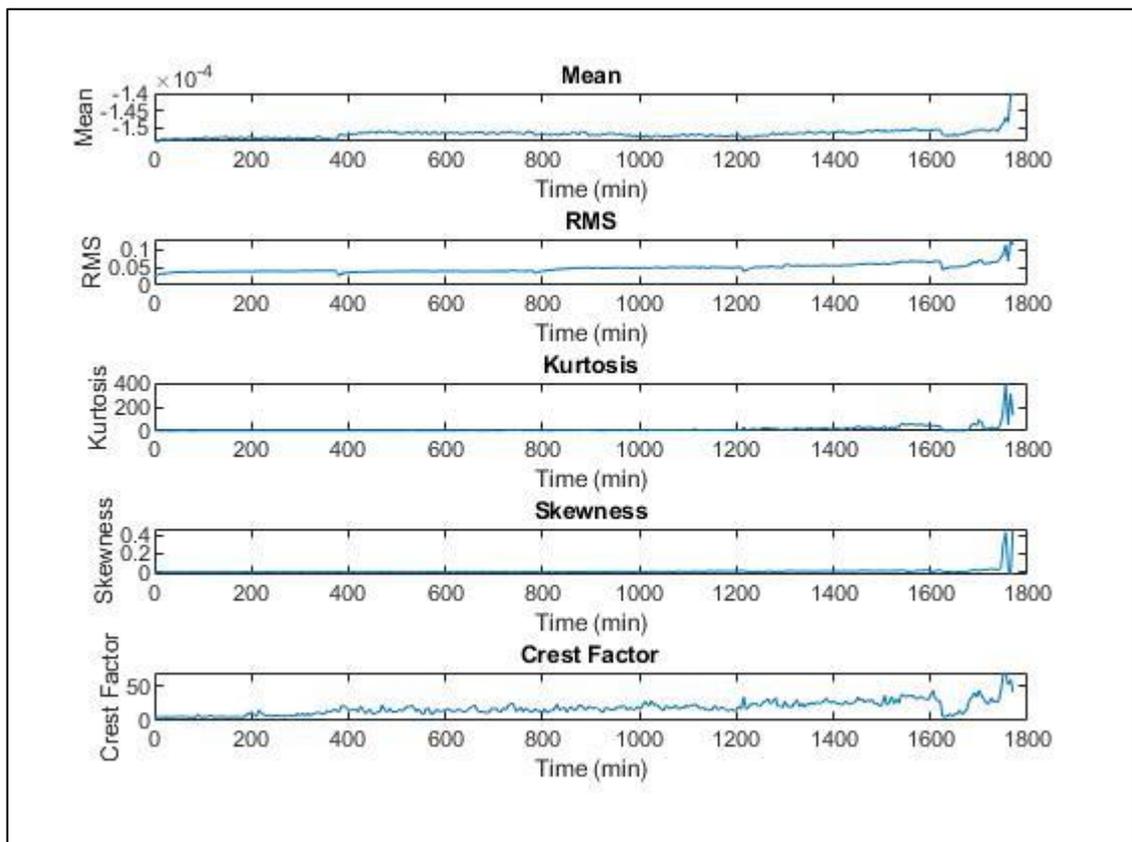

Figure 4.4: Types of Features

These five different features consist of mean, root-mean-square, kurtosis, skewness as well as crest factor. By having several different features rather than one, allows the user to analyse the fluctuation within each feature. As presented above, it can also be seen that each graph has almost similar curve but yet it is different. This is mainly because, some features pick up the trial of failure earlier, whereas some features pick up the other characteristics contained within the running bearing.

The feature extraction can be divided into two different sections for discussion. The very first section will be the plotted features before 1600 minutes. This is because the earliest indication of failure can be deduced from crest factor feature at 1600 minutes. When comparing all 5 different features, what can be further deduced is that the feature of kurtosis and skewness remains very smooth before 1600 minutes. Mean on the other hand shows a minimal fluctuating graph with a small amplitude before 1600 minutes. Root-mean-square however shows a smooth graph before 1600 minutes but yet having higher amplitude. Lastly, crest factor shows a slightly fluctuating graph with high amplitude before 1600 minutes. By just classifying all the features before 1600 minutes, user is able to tell that kurtosis and skewness will not be able to provide sufficient information, as the smooth graph does not have any details. Mean on the other hand might be able to provide information. Crest factor and root-mean-square however shows high potential in providing information which user desires. Riaz et al., (2017) claims that root-mean-square feature and crest factor feature has been popular when diagnosing bearings failure. This is because root-mean-square and crest factor are very powerful and specific approach in determining bearing and rotary machinery imbalance.

The second section that is worth discussing will be the plotted features after 1600 minutes. As it can be seen that the very first feature that shows earliest indication of bearing failure would be crest factor. It simply shows a drastic reduction in amplitude followed by an increment at 1700 minutes. The other features however exhibit almost similar characteristics where a huge increment happens approximately at 1775 minutes. Although all features reveal a drastic fluctuation before the failure occurs, crest factor will be emphasized in this case as it presents early indication of failure approximately 180 minutes beforehand. The other features however present early indication of failure approximately 20 minutes beforehand which is considered too short to be useful.

4.5.2   Signal Processing and Diagnosis

As discussed, and illustrated as above in the methodology section based on test 17 results, signal processing and diagnosis process includes several steps such as PSD, BPF, HT as well as LPF. The final outcome from LPF will allow users to diagnose the fault mode. This simply includes frequency where a particular element went faulty, as

well as when a particular element went faulty. In order to do so, there are several particulars, which must be calculated such as BPFO, BPFI, BSF as well as FTF, as shown in Table 4.2.

Table 4.2: BPFO, BPFI, FTF, BSF

Part	Frequency, Hz
BPFO	74.11 Hz
BPFI	112.2 Hz
FTF	8.012 Hz
BSF	99.15 Hz

Formula and labelling to calculate BPFO, BPFI, FTF, and BSF is depicted in Figure 4.5 and Figure 4.6 respectively, and parameters used to calculate BPFO, BPFI, FTF, and BSF is listed in Table 4.3.

Formulas to Calculate Bearing Frequencies
(inner race rotating and our race stationary)

$$FTF = \frac{S}{2}\left(1 - \frac{Bd}{Pd}\cos\emptyset\right)$$

$$BPFI = \frac{Nb}{2}S\left(1 + \frac{Bd}{Pd}\cos\emptyset\right)$$

$$BPFO = \frac{Nb}{2}S\left(1 - \frac{Bd}{Pd}\cos\emptyset\right)$$

$$BSF = \frac{Pd}{2Bd}S\left\{1 - \left(\frac{Bd}{Pd}\right)^2(\cos\emptyset)^2\right\}$$

RPM - Revolutions Per Minute
S - Revolutions per secord or relative speed difference between inner and outer race (1)
FTF - Fundamental Train Frequency
BPFI - Ball Pass Frequency of Inner ring
BPFO - Ball Pass Frequency of Outer ring
BSF - Ball Spin Frequency
Bd - Ball or roller diameter
Nb - Number of balls or rollers
Pd - Pitch diameter
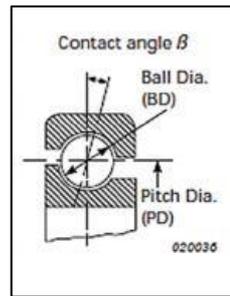Ø - Contact angle

Figure 4.5: BPFO, BPFI, FTF and BSF Formulas

Figure 4.6: Labelling of Bearing to Calculate Bearing Frequencies

Table 4.3: Parameter to Calculate BPFO, BPFI, BSF, and FTF

Parameters	Values	Units
Number of Balls, Z	9	-
Shaft Frequency, f_s	21.01	Hz
Ball Diameter, d	4.75	mm
Pitch Diameter, D	23.5	mm
Contact angle, α	0	°

The above table was the calculated value for BPFO, BPFI, BSF and FTF for test 17. The calculated value was based on the formula given on Figure 4.3 as well as the parameters that was practiced on the test rig machine.

The below results shown are the frequency spectrum results after applying HRFT based on the end results of each day's test. At the same time, a set of figures was also provided to show the changes in raw signal day by day.

4.5.3   Day 1
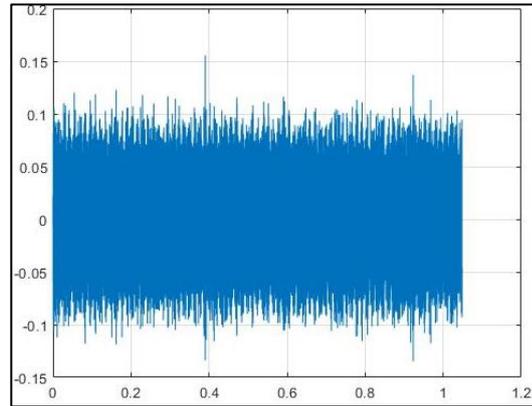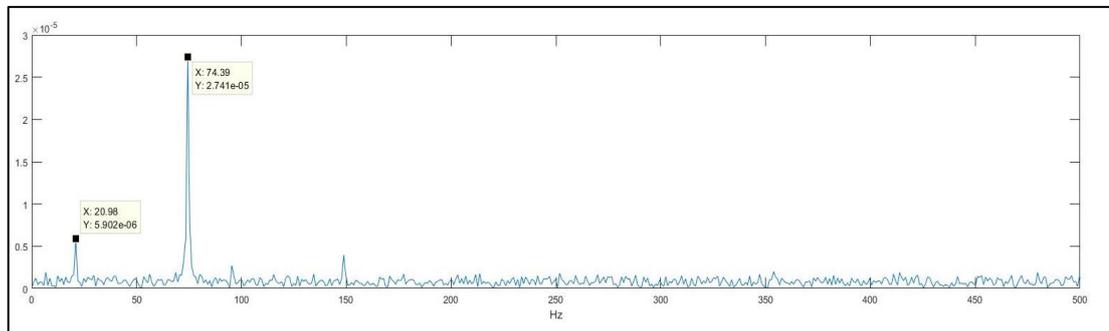

Figure 4.7: Day One Raw Signal (24 hours)

Figure 4.8: Day One Frequency Spectrum Result

As shown in Figure 4.7 and 4.8, the second data tip indicates the frequency of 74.39 Hz, which is very close to the BPFO calculated (74.11 Hz). Therefore, it can be deduced that the bearings outer ring was undergoing some permanent defect. This can be further explained by the possibilities of premature failure discussed above. There might be a high possibility that the force exerted during bearing installation was indeed too much and causes the outer ring to deform slightly. Therefore, this slight deformation causes the outer ring to fail on the very first day.

4.5.4   Day 2
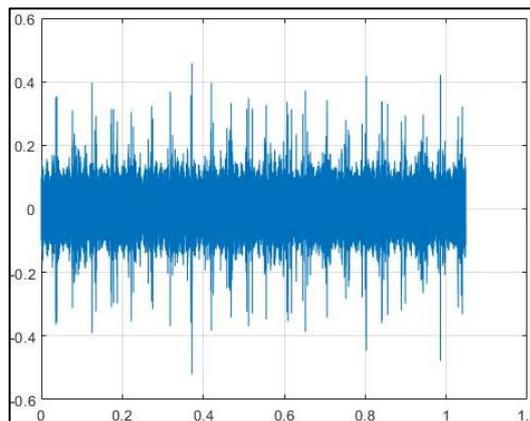
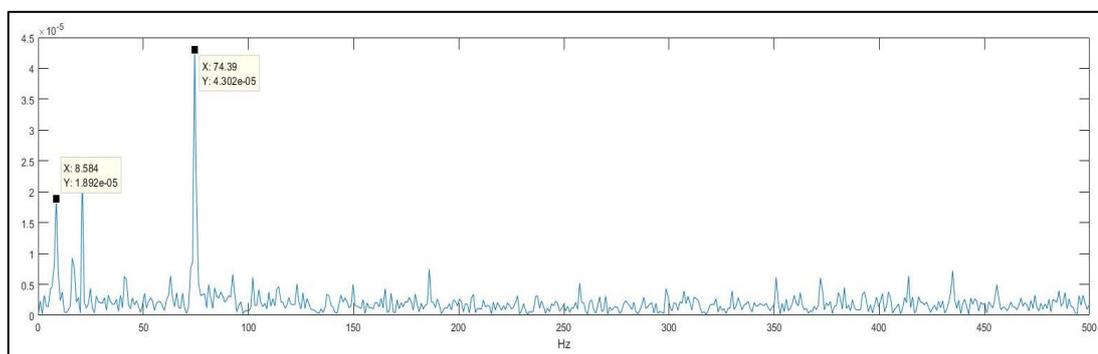Figure 4.9: Day Two Raw Signal (48 hours)

Figure 4.10: Day Two Frequency Spectrum Result

As shown in Figure 4.9 and 4.10, the second day of the test continues to show defects on the outer ring. However, there is also another peak in this graph that can be discussed. The first data tip indicates a frequency of 8.584 Hz. This drastic fluctuation indicates a possibility of failure on the bearings cage as the lowest frequency calculated was 8.012 Hz, which is also known as the FTF. However, users could not deduce this as failure of the bearing's cage as the amplitude isn't sufficiently high but this could allow users to safely assume a high degree of degradation experienced by the cage.
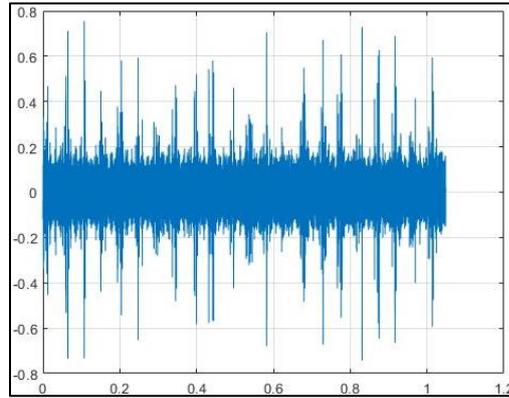
4.5.5   Day 3

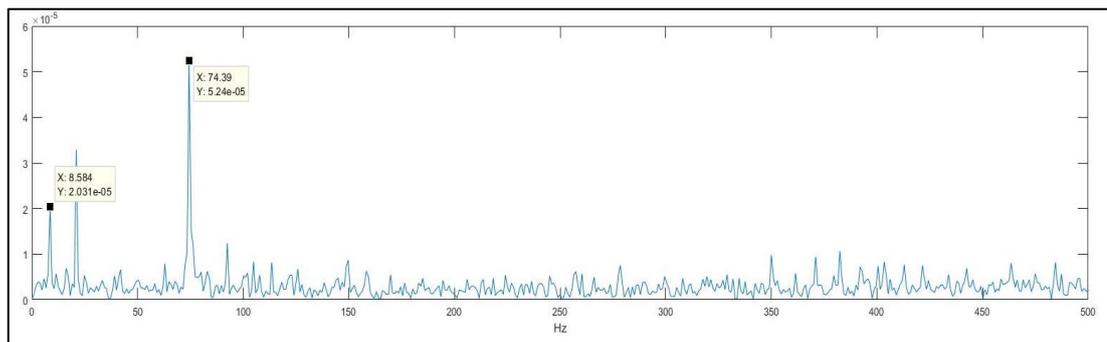Figure 4.11: Day Three Raw Signal (72 hours)

Figure 4.12: Day Three Frequency Spectrum Result

As shown in Figure 4.11 and 4.12, the third day of the test continues to show defects on the outer ring as well as possibility of bearings cage failure as discussed on the second day of test. It can also be observed that the fluctuation of the graph is of higher degree compared to the second day of the test. This also shows that the lifespan of the bearing is decreasing gradually.
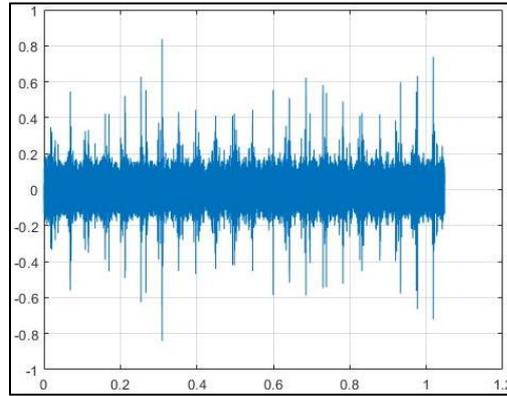
4.5.6   Day 4
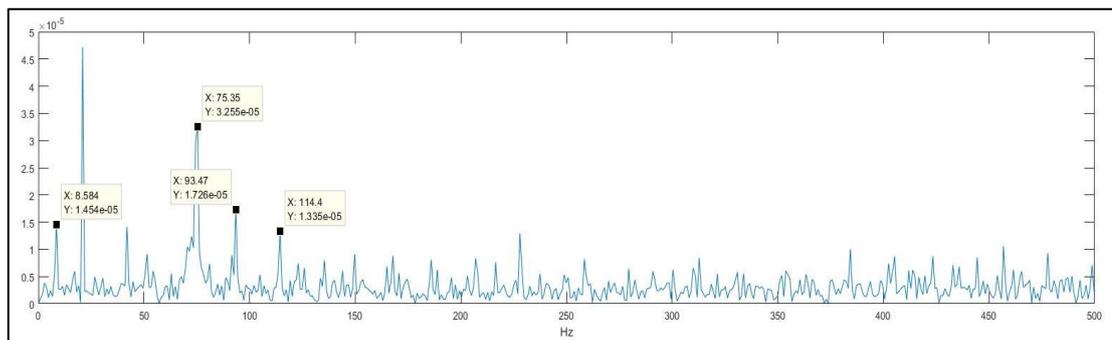
Figure 4.13: Day Four Raw Signal (96 hours)

Figure 4.14: Day Four Frequency Spectrum Result

As shown in Figure 4.13 and 4.14, the fourth day of the test continues to show defects on the outer ring as well as possibility of bearings cage failure as discussed on the second day of test. However, new peaks were revealed as 93.47 Hz as well as 114.4 Hz. These 2 peaks were close enough to the BSF, which is 99.15 Hz as well as BPFI of 112.2 Hz. The value of the data tip of 93.47 Hz is not greater than the BSF calculated, therefore users could interpret the peak as an early indication of bearing's ball failure. The data tip of 114.4 Hz however indicates the degradation of inner ring's health state.
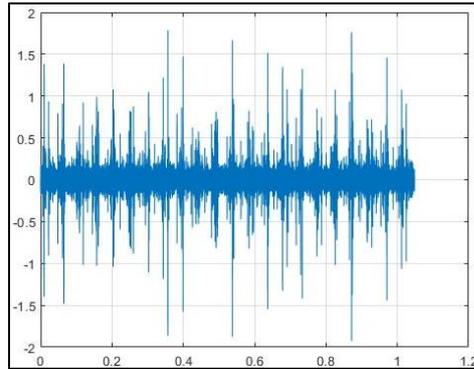
4.5.7   Day 5
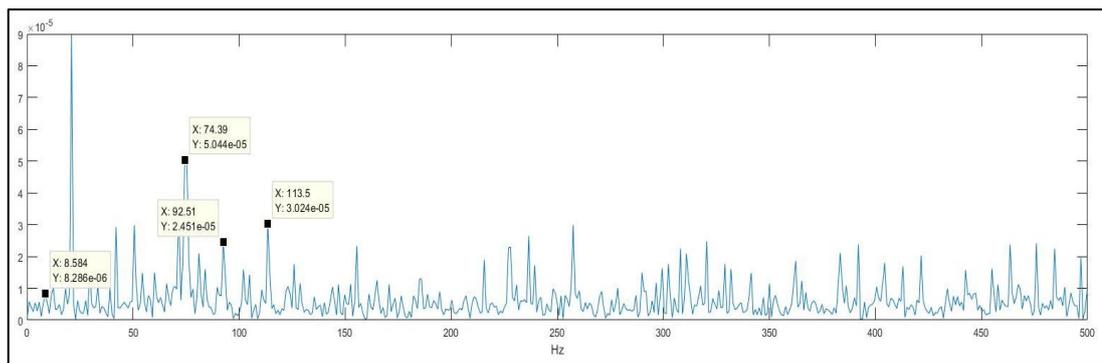
Figure 4.15: Day Five Raw Signal (120 hours)

Figure 4.16: Day Five Frequency Spectrum Result

As shown in Figure 4.15 and 4.16, the fifth day of the test continues to show defects on the outer ring as well as possibility of bearings cage failure as discussed on the second day of test. It also shows degradation of health state on the inner ring as well as early indication of bearing's ball failure as discussed on the fourth day of the test. Although there might be some changes in data tip values, users are safe to assume the interpretation of graph before this still stands. However, the graph presented on fifth day shows higher degree of fluctuations compared to the fourth day, which also indicates a gradually decreasing lifespan of bearing.

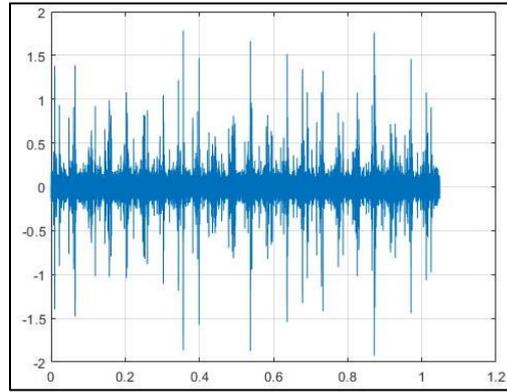**4.5.8   Day 6 (Day of Bearing Failure)**

Figure 4.17: Day Six Raw Signal (144 hours)

It can be seen through the changes of the raw signal amplitude from day one till the very last day where the test bearing failed. It changes from a uniform amplitude towards a high degree of fluctuation in its amplitude. It was also noticed that the raw signal of on the sixth day, as seen in Figure 4.17 and 4.18, showed a trend which its amplitude is "spikey". This characteristics shows instability as well as degenerating properties in it's health state.
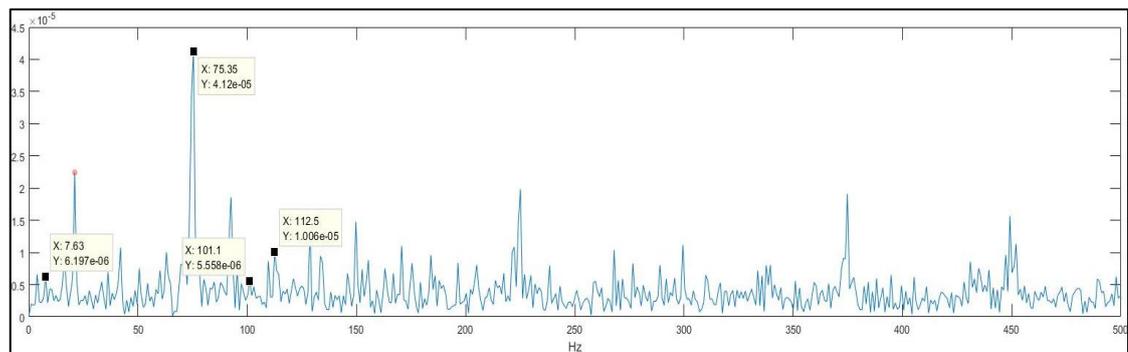

Figure 4.18: Day Six Frequency Spectrum Result

As shown in Figure 4.17 and 4.18, the sixth day of the test continues to show defects on the outer ring as well as defect on the inner ring as mentioned earlier. The peak of 7.63 Hz remains low, which couldn't assure failure on the cage but only the degrading health state of it. A new peak of 101.1 Hz appears which tally the frequency of BSF but it has relatively low amplitude. Therefore, it can only be deduced that the health state of ball is degrading.

The graph also represents the data of bearing when it fails. As the BPFO shows the highest amplitude compared to the other frequencies, user could deduce that the failure of outer ring is the main reason of the bearing failure.

4.6 Prognosis

4.6.1 Feature Selection Technique

In order to combine all the features with the right amount of weightage, NCA technique was chosen as explained in the methodology section. RFE was also used to assign weightage to each feature.

Figure 4.19 shows the health indicator of test 17 derived from NCA technique.
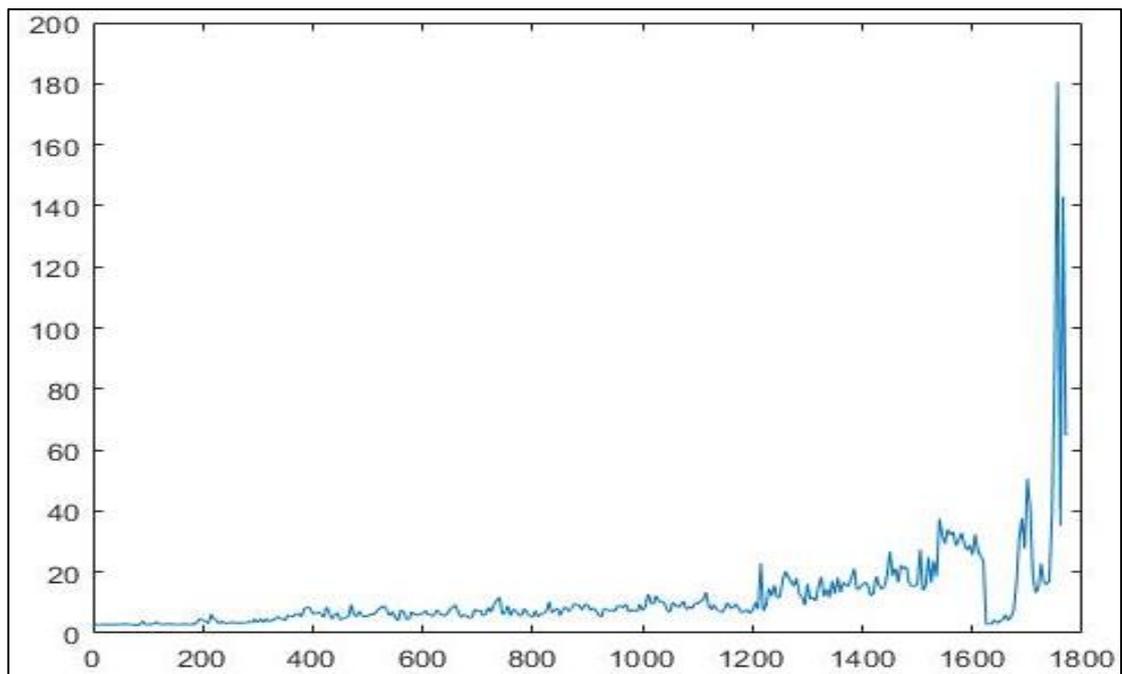

Figure 4.19: Health Indicator of Test 17

The formula used in NCA is:

$$Y = (A)x_1 + (B)x_2 + (C)x_3 + (D)x_4 + (E)x_5 \qquad (4.1)$$

The formula used in RFE is:

$$w_i = \frac{1}{2}\alpha^T K \alpha - \frac{1}{2}\alpha^T K(-i)\alpha \qquad (4.2)$$

As discussed earlier in feature extraction, all five features inhibit different trends, which might be useful in determining bearings RUL. Therefore, RFE was used to determine the weightage of each feature. Below will be the explanation on how RFE assigns weightage.

The very first feature to be discussed will be skewness. As skewness remain almost smooth all the while without fluctuation up till 1775 minutes, the weightage distributed to it will be zero (0). This is because it doesn't show any form of indication to bearing failures before 1775 minutes, and the fluctuations after 1775 minutes is almost similar to kurtosis. Therefore, it can be classified as weak feature although the trails picked up after 1775 minutes is important. This is because it would be a better choice to rely on kurtosis to extract the failing indication after 1775 minutes. In other words, this feature will be removed from weightage distribution.

The second feature to be discussed will be crest factor. Crest factor shows constant fluctuations throughout the entire graph and it is the feature, which presents the earliest indication of bearing failure. Therefore, this feature was assigned with the weightage of 0.3 because it is considered as a strong feature. The reason why this feature wasn't assigned higher weightage than 0.3 is due to the constant fluctuation. Although it shows the earliest indication of bearing failure, which is desired by user, the constant fluctuation might cause inaccuracy in final result.

The third feature to be discussed will be RMS. RMS shows a smooth graph before 1600 minutes but with higher amplitude. However, the degree of the fluctuation of the graph after 1600 minutes isn't high enough to indicate a bearing failure. In short, the trails picked up by this feature before 1600 minutes is desired but not the trails after 1600 minutes. Therefore, this feature is considered as a intermediate feature and the weightage assigned to this feature would be 0.2.

The fourth feature to be discussed will be mean. Mean shows a smooth graph with slight fluctuations before 1600 minutes and a gradual increase in the graph after 1600 minutes. Although the indication of bearing failure beforehand happens at 1640 minutes, the degree of fluctuations wasn't high enough to assign higher weightage. Therefore, the weightage assigned to this feature would be 0.1 as it is a weak feature. The last feature to be discussed will be kurtosis. This feature was assigned with the weightage of 0.4, which is higher than any other features. The main reason behind this is due to the trails picked up at 1775 minutes. At 1775 minutes, it shows a drastic

change in amplitude, which confirms the failure of bearing and this particular characteristic was only presented in kurtosis as well as skewness. As mentioned above, the rightful weightage of skewness will be assigned to kurtosis to emphasize on the trails picked up at 1775 minutes. Moreover, kurtosis shows an early indication of failure at 1520 minutes. Although the amplitude wasn't high enough to deduce as an early indication of bearing failure, the information is useful and important later on in prognosis. Therefore, kurtosis is considered as a strong feature.

As a result, the health indicator generated with the weightage did carry all the important information needed to predict the RUL of bearing. As discussed above in the feature extraction section, the two main features, which are able to provide information on degrading health state of bearing before 1600 minutes would be RMS as well as crest factor. RMS carries the amplitude whereas crest factor carries the degree of fluctuation. After assigning a carefully analysed weightage based on RFE to NCA technique, the important information of health degradation state was carried to the health indicator. This can be seen before 1600 minutes of the health indicator, it shows a certain degree of fluctuations with reasonable amplitude which justifies the degrading health state of bearing.

The characteristics after 1600 minutes were also successfully captured. The main characteristics, which are desired after 1600 minutes would be the early indication of crest factor, as well as the failure, captured at the end of kurtosis. From the health indicator, it was seen that both the characteristics mentioned were emphasized.

4.7   Support Vector Regression

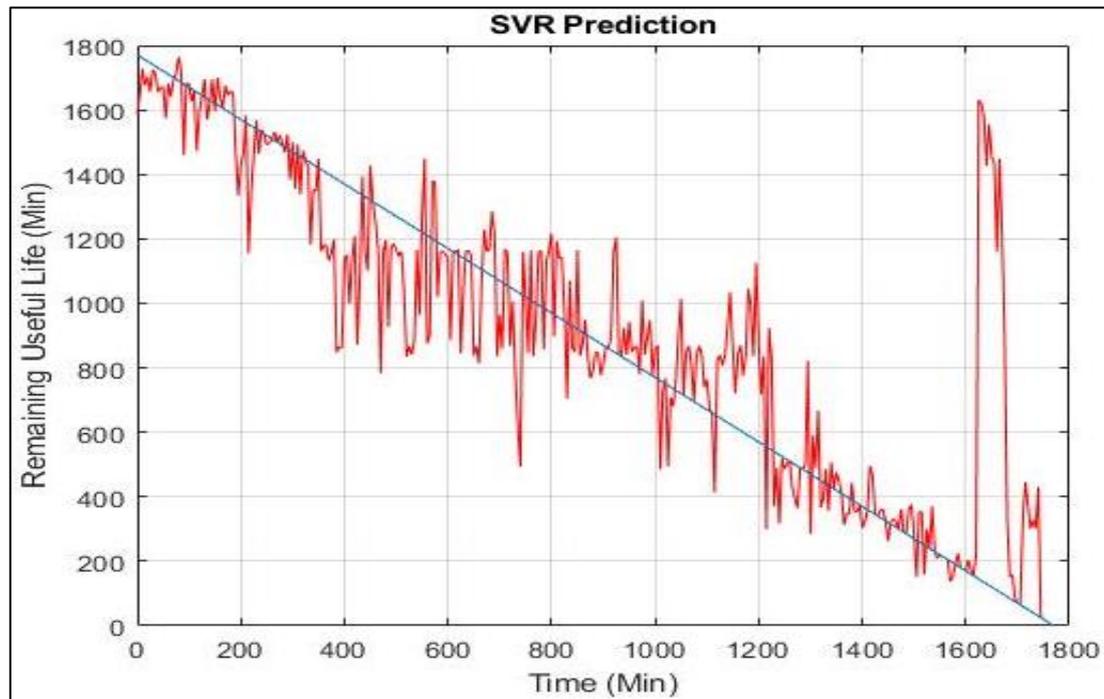**4.7.1   Individual Training Set**

Figure 4.20: Test 17 Predicted RUL

Figure 4.20 shows the predicted RUL of test 17. The blue linear line shows the real RUL of the bearing whereas the red fluctuating lines shows the predicted RUL of bearing. Note that the more interception between the real RUL vs the predicted RUL, the higher the accuracy of the prediction.

The above figure could be further explained in terms of the bathtub curve. First of all, the curve could be divided into three different stages, which can be classified by time. The first stage would be from zero to approximately 400 minutes. The second stage would be from 400 to approximately 1200 minutes. The last stage however would be from 1200 to approximately 1800 minutes.

The first stage is also known as the infant mortality stage in terms of the bathtub curve. As it can be seen, the fluctuation of the graph is minimal although there is a spike at approximately 200 minutes. There are several explanations, which could satisfy the appearance of the spike, which are impurities or unsuitable weightage used in NCA. However, the weightage distributed in NCA was justified above and user could safely assume the spike was due to impurities. The small degree of amplitude

fluctuation indicates the stable condition of the bearing. In other words, the bearing's elements can be assumed to be in good condition initially. This stage also indicates the decreasing failure rate of the bearing as the bearing starts to settle into the experimental condition. It was noted that the first stage is not stable at the beginning as the bearing's condition haven settle down.

The second stage is known as the normal life stage. As the bearing proceeds from the first stage to the second stage, the bearing reaches stability and the AE signal emitted will be more uniform. Although the degree of fluctuation at second stage is significantly higher, it was noted that the trend is consistently in a horizontal line. This is the stage where the bearing's condition is most stable and its health state does not fluctuate. In fact, the second stage is considered much more stable compared to all the other stages as it has low constant failure rate. At approximately 900 minutes, it can be seen in the figure above that the predicted health state starts to change a little in the sense that the signal no longer exhibits the trend as before. This indicates the bearing in about to enter the third stage.

The third stage is known as end of life stage. In other words, it is also known as wear out stage where the bearing is about to fail and operator needs to take extra care to prevent unscheduled break down. Although the degree of fluctuation is not high, but the predicted RUL plotted is steep. The degree of steepness plotted in the graph above indicates the increasing failure rate of the bearing. It was also noted that at the end of 1600 minutes, there is a drastic change in amplitude, which contributes to a highest peak throughout the entire graph. There are several explanations for this situation such as impurities, self-healing, or improper technique applied during experiment. As discussed, impurities such as debris of failed bearing may cause inaccurate results generated, not able to conduct experiment during the weekend due to university policy allows the bearing to self-heal, and improper technique in terms of overlooking loading force (loading force lower than 3 kN) may cause the AE signals emitted to reduce which further causes an inaccuracy in final results. The third stage is also a phase where the bearing is not in stable condition and operators should take extra care about the situation.

The accuracy of this predicted RUL is 75.43 %. As it can be seen in the graph presented, there are several regions when the predicted RUL went off-track and it contributes to a lower accuracy. Besides, the highest peak at approximately 1600

minutes as discussed above also causes the accuracy to deviate. However, the accuracy is within acceptable range and extra care should be taken in future attempts.

Table 4.4 will be the other individual training set with brief explanations.
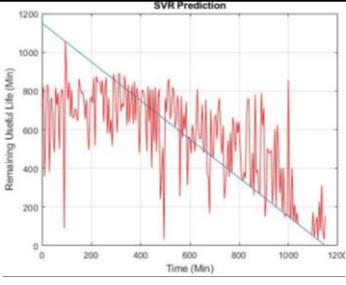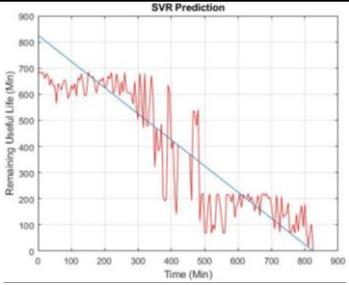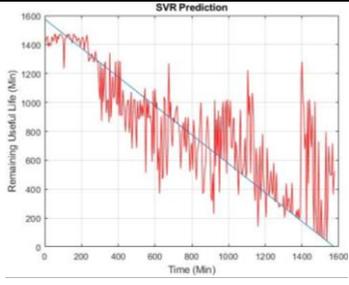
Table 4.4: Training Set Description

Test 20	Test 21	Test 23
The accuracy of test 20 is 70.56 %. It was noted that the early prediction of the test was not on-track. This causes a certain amount of deviation in the accuracy. At the same time, it can be noticed that there were lots of random peak which might be caused by impurities. This can also cause deviation in final results.	The accuracy of test 21 is 81.59 %. Test 21 only lasted for approximately 800 minutes. As explained in methodology section, this SVR method handles smaller data better and have better prediction results.	The accuracy of test 23 is 70.32 %. It was noted that the beginning of the prediction was quite well. However, random peaks start to appear at the middle as well as at the end of the bearings RUL. This can be deduced that debris was created and accumulated throughout the experiment which further causes the bearing's condition to be debris saturated. As a result, the prediction starts to deviate after the middle of bearing's RUL.

Table 4.5: Prediction Accuracy of Tests

Test	Prediction Accuracy (%)
17	75.43
18	-
19	52.37
20	70.56
21	81.59
22	71.48
23	70.32

4.7.2   Combined Training Set

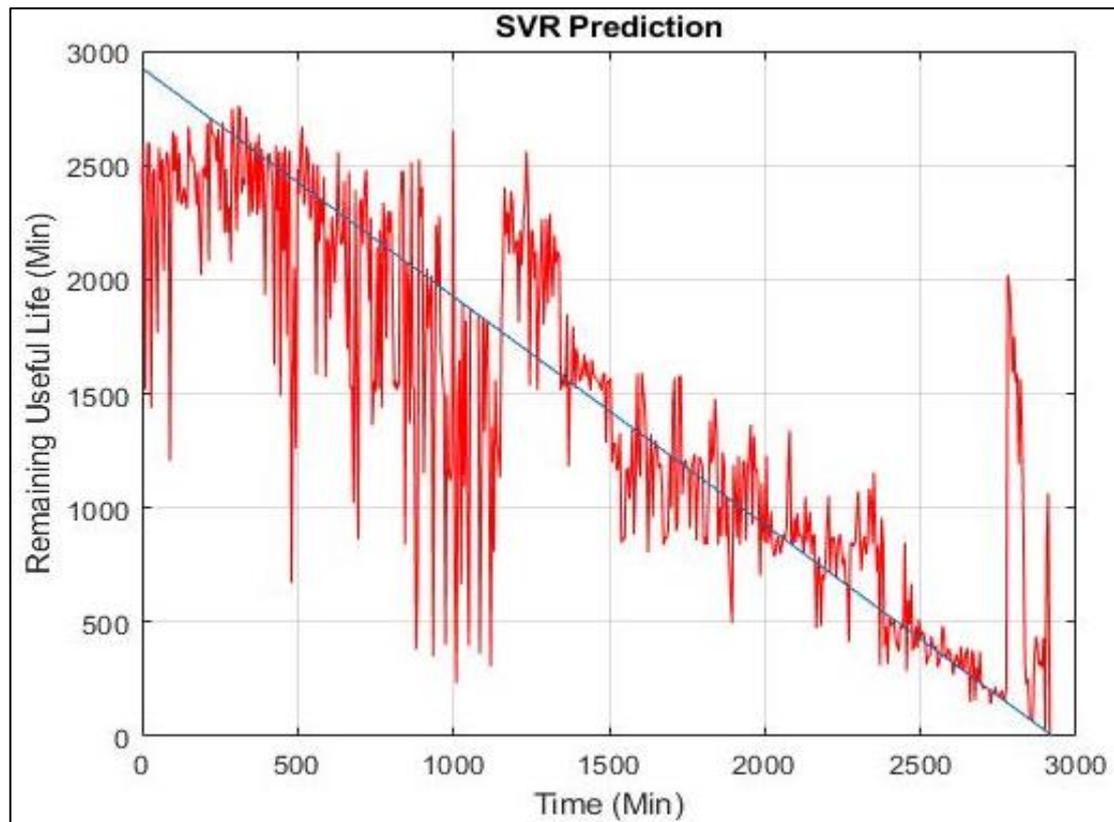
Figure 4.21: Test 17 and Test 20 Training Set

Figure 4.21 shows a training set, which includes test 17 and test 20. The accuracy of this training set is 72.59 %. The accuracy is considered reasonably high, as it is a combination of predicted RUL from two sets of tests. As it can be seen from the above figure that this particular training set has characteristics of both test 17 as well as test

19. The pros of this training set simply include the captured failure possibilities as well as reasonable decision making of the SVR. On the bad side, the inaccurate predicted characteristics were also included within this training set. This involves the peak at approximately 2750 minutes from test 17 as well as the inaccurate early prediction of test 20. Although there is a certain degree of fluctuation throughout this particular training set, but the interception of real RUL and predicted RUL is frequent. The trade off between the pros and cons of this technique is considered worth it.

There are also several training sets, which were attempted during this project. Figure 4.22 will illustrate the training data results, which was generated.

Figure 4.22: Test 17 and Test 23 Training Set

This combined training set of tests 17 and test 23 achieved an accuracy of 60.98 %, which is relatively low. This could be explained in terms of the RUL trend of both training test. It was noted that the trend of RUL for test 17 is not similar when compared to test 23. The degree of fluctuations in both data is significantly different. When both training sets exhibit different RUL trends, this causes an inaccurate result generated.
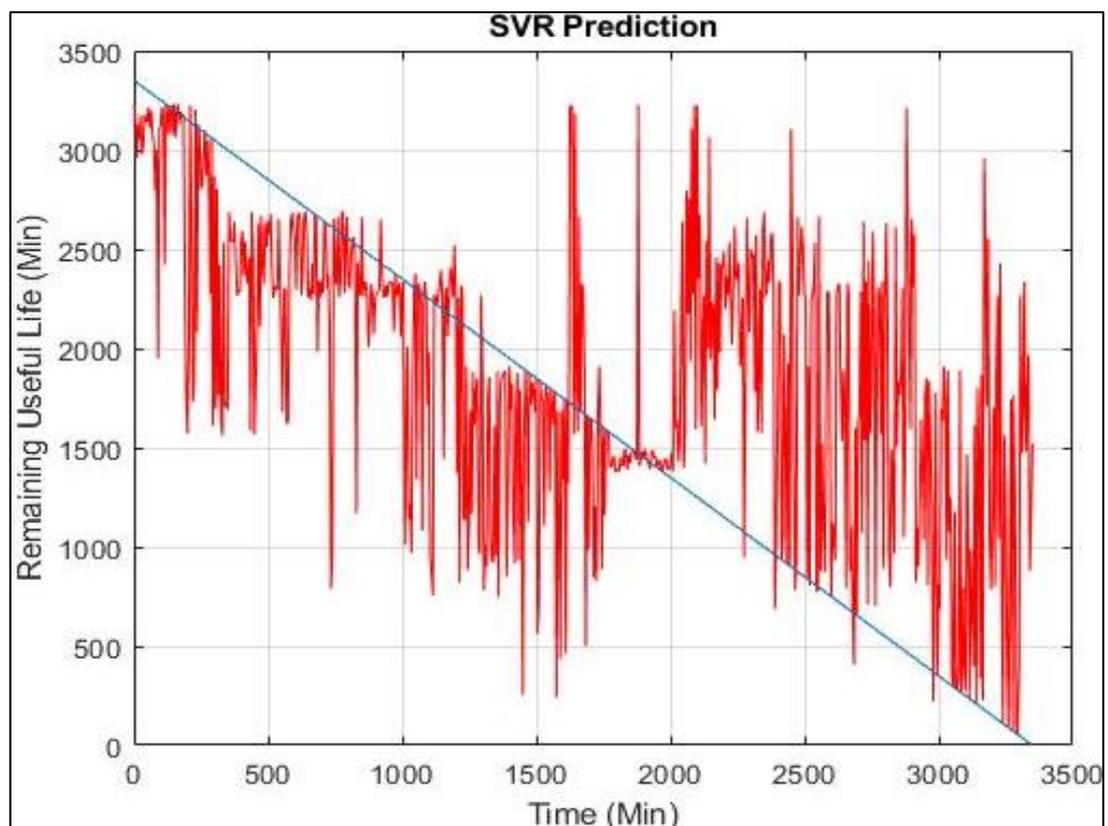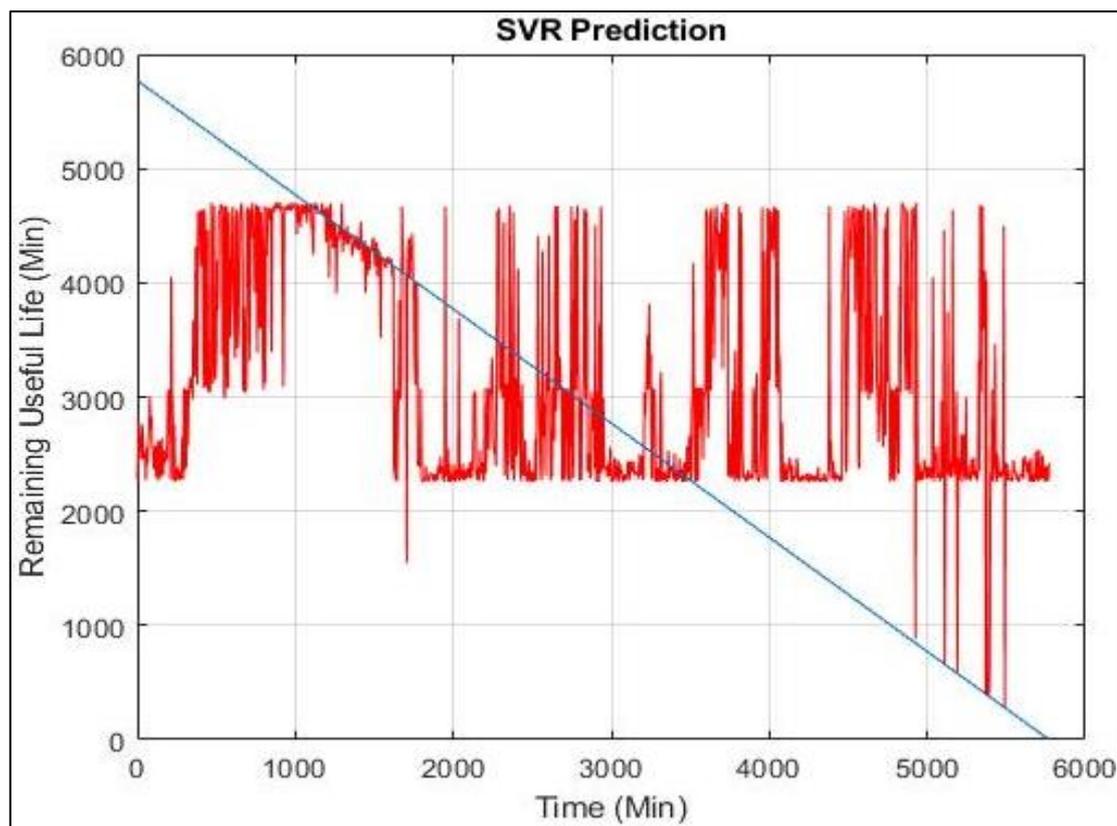
Figure 4.23: Test 17 and Test 19 Training Set

Figure 4.23 combines training set of tests 17 and test 19 achieved an accuracy of 43.47 %, which is very low. This could be explained in terms of the RUL trend of both training test as well as accuracy of individual test. It was noted that the trend of RUL for test 17 is not similar when compared to test 19. Test 17 failed gradually with the desired curve and trend whereas test 19 lasted very long before failures and its curve is simply a constant fluctuation. The trend of RUL for both training sets are completely different and there is a very huge gap of RUL between both tests. The degree of fluctuations in both data is also significantly different. Lastly, the accuracy of test 17 is 75.43 % whereas accuracy of test 19 is only 52.37 %. Therefore, the poor performance in test 19 prediction affects the overall prediction performance of the combined training sets.

Table 4.6: Prediction Accuracy of Tests

Test	Prediction Accuracy (%)
17 + 20	79.18

17 + 23	60.98
17 + 19	43.37

4.7.3   Training Set Online Testing

From the entire combined training test, the training test with the highest accuracy was chosen to perform the online testing. This is because a good prediction can be done only with good characteristics picked up from high accuracy training test. Therefore, low accuracy training sets were not included in online testing. Online testing was done on test 22 and test 23, as shown in Figure 4.24.

Figure 4.24: Online Test 23 with Training Set for Test 17 and 20

The training set of tests 17 and test 20 were applied on test 23 online and it achieved an accuracy of 70.66 %. As shown on the above section, combined training set of tests 17 and test 20 has accuracy of 79.18 %, which could be explained in term of their similarity between RUL trend. However, when the training set was applied online on test 23, the accuracy reduces by 8.52 %. The accuracy reduces is mainly due to the differences in trend between test 23 compared to trend of test 17 and test 20.
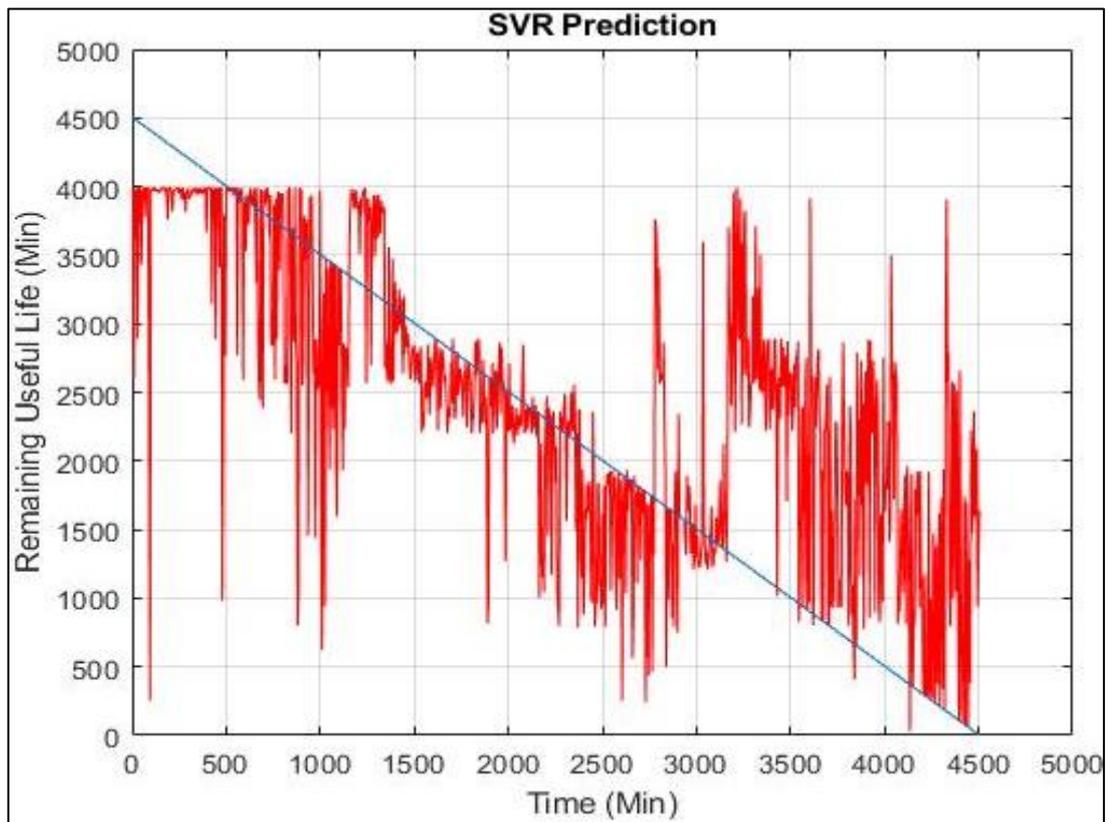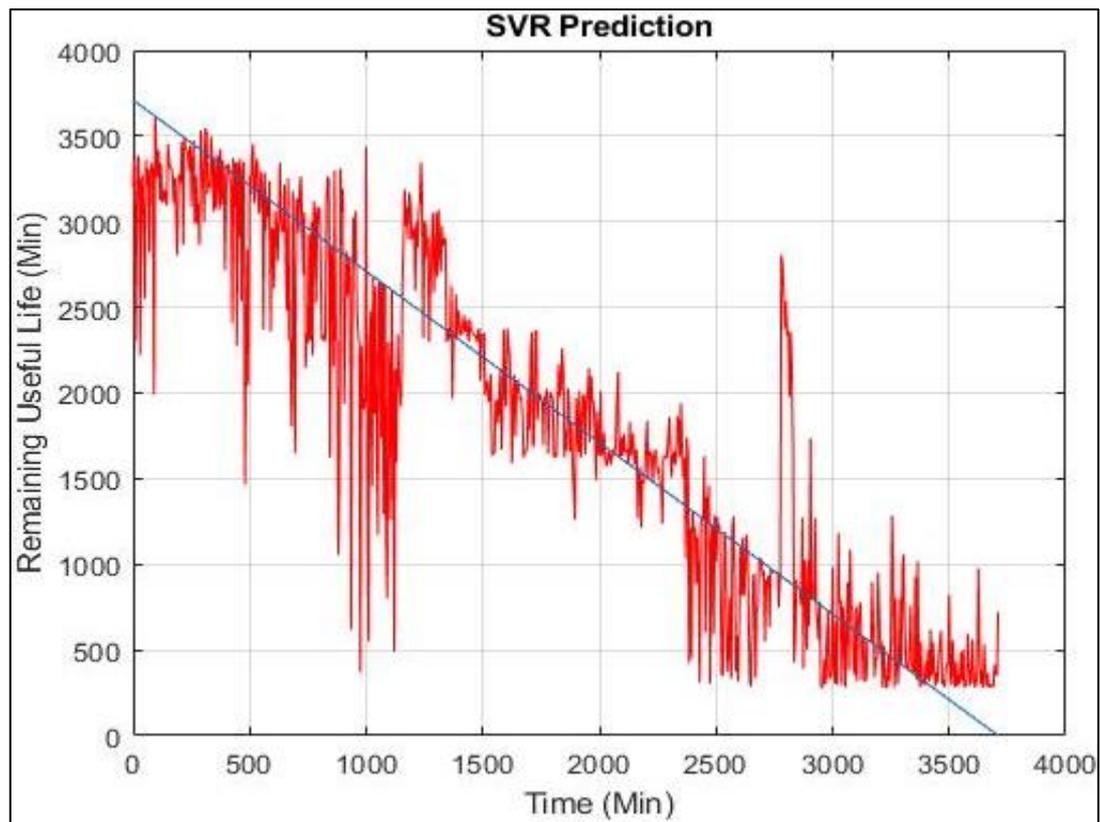
Figure 4.25: Online Test 22 with Training Sets of Test 17 and Test 20

The training set of tests 17 and test 20 were applied on test 22 online and it achieved an accuracy of 81.95 % as shown in Figure 4.25. As shown on the above section, combined training set of tests 17 and test 20 has accuracy of 79.18 %, which could be explained in term of their similarity between RUL trend. When the training set was applied on test 22, the accuracy increases by 2.77 %. This is mainly because the trend of the entire test in the result generated inhibits similar RUL trend. Therefore, a high accuracy was achieved.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATION

5.1 Conclusions

From the aim and objectives of this project, which is summarized into "developing signal processing, diagnosis as well as prognosis algorithm" was achieved throughout this study.

For signal processing, the main objective of it is to reduce the noise level within the received signals in order to verify fault signals emitted by bearing. The noise was emitted through the vibration of experimental set up application parts. Several steps were taken in order to achieve a final result. These steps consist of PSD, BPF, HT, FFT and finally LPF. The specific series of approach is also known as HRFT. It is important or user to understand that the noise can never be eliminated but reduced. Although it might still affect the accuracy of final result, but the difference would be insignificant.

The main purpose of diagnosis is to determine what are the main reasons of the bearing's failure. By applying the formula of FTF, BPFO, BPFI and BSF, the bearing's fault frequency could be determined which could further tally with the frequency spectrum generated through HRFT. Fault frequency with highest amplitude among the four-fault frequency stated above would indicate the main reason of bearing failure. From diagnosis, user could also analyse the condition of the experimental setups such as the shaft's condition. For instance, the push belt of the test rig was once worn out during this project and was determined through the diagnosis results. From the diagnosis result, the fluctuation of the graph indicates an inconsistent speed of shaft.

For prognosis, it was mainly to predict the RUL of the bearing based on historical data. The feature selection technique chosen in this project is NCA and the technique to distribute weightage would be RFE. Out of 10 sets of tests carried out during this project, all the individual test results were considered reasonable which their accuracy ranges between 70.56 % to 81.59 % (except for test 18). However, when training several sets of data together, the ranges of the predicted accuracy are between 72.59 % to 15.94 %. It was noted that low accuracy of one test will affect the accuracy of the combined training set. At the same time, results of different RUL trend trained together in combined training sets would also results in low accuracy. The online test however yields accuracy between 50.67 % to 81.95 %. It was noted that similar RUL trends between online test and training test would produce high accuracy.

In conclusion, the combination of signal processing, diagnosis as well as prognosis is able to predict the RUL of the bearing which fulfils the objective of this study. These combinations of techniques can also be applied to bearings of different type, dimension as well as working condition. The technique used is considered reliable and produces reasonably high accuracy.

5.2   Recommendation

5.2.1   Real life application

Bearing is a very common element, which exist around our environment which plays a very important role. As this project's aim is to predict bearing's multistage failure, it could be applied in many fields to help increase efficiency, to prevent casualty and so on.

For example, this project could be implemented in manufacturing firms, which uses rotary machines. As most manufacturing firm depends on rotary machines to produce goods, bearing serves as an element, which reduces friction of the rotary machine. This simply enables the manufacturing firm to progress under controlled condition with higher precision and accuracy.

However, when the manufacturing firm face an unplanned breakdown of the machines due to bearing failure, it could cost a lot of losses not only in terms of money but also at the same time customers trust. Although there is a lot of breakthrough in technology in the past few decades, many manufacturing firms are still relying on human's judgement (by looking, by hearing, and maybe even the scent), which is based on their experience. This could lead to either an unplanned breakdown or wastage of bearing's RUL (maintenance while the bearing is still in good condition). This could be avoided through implementation of this project where the condition of the bearing is constantly supervised.

Manufacturing firm could deploy this project by installing sensors and investing in a super computer that could run the signal processing, diagnostics and prognosis algorithm smoothly. It is clear that the number of bearing in a manufacturing firm could be in hundreds or even thousands, and by installing sensors in each of them could be expensive.

Therefore, it is only wise to classify each the bearings into several categories based on their working condition such as bearing types, load exerted onto each bearing,

temperature of each bearing as well as load threshold of each bearing. After each bearing was categorized, users could choose to implement this project onto the bearing, which is going to fail earliest among its category which can be done through examining historical report. As each category inhibits similar characteristics, it is safe to assume that these bearings have approximate RUL. Therefore; the predicted RUL of the supervised bearing could illustrate the predicted RUL of the others among its category. Therefore, when the supervised bearing indicates the need to be maintained, user should maintain all bearing within that particular category. By doing so, the cost of implementing this project would be greatly reduced and at the same time avoiding unplanned breakdown.

There is also another method, which is cost effective and simple at the same time. This particular method emphasizes only on a certain area, which is considered crucial. To implement this method, user must first define the machines, which could possibly affects the manufacturing firm most. The word "affects" stated above depends upon the company's best interest and this usually includes the cost as well as safety. For example, user could determine the machines that could possibly cause a lot of losses if unplanned breakdown occurs as well as machines that could possibly impose a threat in terms of safety. By doing so, user could focus only on important machineries that is of interest with a fairly lower cost.

5.2.2   Further study on this project

In fact, bearing is installed in many fields and each field has different requirement. For example, the bearing installed in rotary machines has lower safety factor compared to the bearing installed in aerospace applications. This is because each element in aerospace applications needs to be very precise and accurate in order for it to function and any errors could cause casualty. Therefore, extra care was taken when pursuing aerospace application.

However, the principles and mechanics of bearing remains the same for both rotary machines and aerospace application. The only difference between them is the technique of keeping the bearing in controlled condition as well as the quality of the bearing. This project could be further expanded into aerospace application.
This can be done by testing on aerospace grade bearing and implementing a different technique in keeping the bearing in controlled condition. For example, the current test

bearing used is MISUMI B6002, which is recommended for rotary machine used whereas examples of aerospace grade bearing are NTN J3. Instead of using lubricant, aerospace grade bearing is advised to apply high quality grease which, was claimed to be more efficient and effective. Besides, as the working environment is different, the technique used in signal processing, diagnosis as well as prognosis should also improvise.

5.2.3   Ways to improve project overall results by improving techniques

The project could be further improvised not only in terms of signal processing, diagnosis, and prognosis technique but also bearing installation technique.

As discussed above in the result and discussion section, the bearing installed was based on naked eye examination of users to ensure the shaft alignment relative to the bearing as well as depth. Misalignment of shaft relative to bearing could result in premature bearing failure due to asymmetrical force distribution whereas excessive depth of bearing relative to shaft could result in permanent deformation of bearing's outer ring.

This could be solved by investing in laser sensors, which could assure the alignment of the shaft relative to bearing. At the same time, an automated hydraulic jack could also avoid excessive depth of bearing relative to shaft.

Besides, the technique used in FFT could also be further improved. As this project uses "eight-point DFT" which is a general formula that is applicable to real data, other algorithms, which were much more specific in signal processing could be studied and further enhanced. Although "eight-point DFT" produces a relatively high accuracy, there are still other algorithms, which could provide higher accuracy. For instance, Liu et al., (2012) claims that window interpolation algorithm applied in harmonics measurement achieves frequency error lower than 1×10^7 Hz, phase error lower than 0.0001% as well as amplitude error lower than 1×10^6 %.

Furthermore, the approach of this project is data based prognosis which the decision making in prognosis is fully dependent on the historical data. As this project is rich in historical data and the condition of the bearing is well under controlled, another alternative of prognosis technique that can be applied will be hybrid-based prognosis. Hybrid-based prognosis is a technique where both data and physics theory contribute to the decision making of prognosis results. The main reason why hybrid-

based prognosis wasn't pursued in this experiment is due to the time constraint as hybrid-based prognosis is much more complicated and coding required for hybrid-based prognosis is also much complex. It is suggested to prolong the project time frame to allocate more time for hybrid-based prognosis implementation.

Additionally, it was also discussed in the result section that there was still noise even after signal processing was carried out. This does not imply that the signal processing technique was not effective but the background noise generated by the bearing's housing is inevitable. Therefore, user could install damper at specific locations within the test rig to reduce the level of unwanted noise generated. However, user should be extremely careful on the location of the damper installed as it might also mask important signals generated from the bearing. The suggested location of damper installation would be the outer surface of bearing's housing, the base of the test rig, as well as parts that involves bolts and nuts tightening. It was noted that the suggested location for damper installation are the experimental parts that are prone to vibrate as well as emit signals (the parts stated are not in a piece but a combination put together through bolts and nuts). However, the signals produce by these parts is the noise indicated in the results above. This is because the important signal, which is desired was only emitted from the test bearing itself.

REFERENCES

Altmann, J. & Mathew, J., 2001. Multiple Band-Pass Autoregressive Demodulation For Rolling-Element Bearing Fault Diagnosis. *Mechanical Systems and Signal Processing*, 15(5), pp.963–977.

An, D., Kim, N.H. & Choi, J.-H., 2014. Practical options for selecting data-driven or physics-based prognostics algorithms with reviews. *Reliability Engineering & System Safety*, 133, pp.223–236.

Anon, A new procedure for using envelope analysis for rolling element bearing diagnostics in variable operating conditions. *Mechanical Systems and Signal Processing*, 38, pp.23–35.

Antoni, J., 2004. The spectral kurtosis: a useful tool for characterising non-stationary signals. *Mechanical Systems and Signal Processing*, 20(2), pp.282–307.

Benkedjouhb, T., Medjaher, K., Zerhouni, N., & Rechak, C. (2013). Remaining useful life estimation based on nonlinear feature reduction and support vector regression . *Engineering Applications of Artificial Intelligence , 26* (7), 1751-1760.

Dong, M. & He, D., 2006. A segmental hidden semi-Markov model (HSMM)-based diagnostics and prognostics framework and methodology. *Mechanical Systems and Signal Processing*, 21(5), pp.2248–2266.

Duque, O., Perez, M. & Morinigo, D., 2005. Detection of Bearing Faults in Cage Induction Motors Fed by Frequency Converter using Spectral Analysis of Line Current. *IEEE International Conference on Electric Machines and Drives, 2005.*

Ebersbach, S. & Peng, Z., 2008. Expert system development for vibration analysis in machine condition monitoring. *Expert Systems with Applications*, 34(1), pp.291–299.

Fernández-Francos, D. et al., 2012. Automatic bearing fault diagnosis based on one-class ν-SVM. *Computers & Industrial Engineering*, 64(1), pp.357–365.

Granitto, P., Furlanello, C., & Gasperi, F. (2006). Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *Chemometrics and Intelligent Laboratory Systems , 83* (2), 83-90.

Immovilli, F. et al., 2009. Detection of Generalized-Roughness Bearing Fault by Spectral-Kurtosis Energy of Vibration or Current Signals. *IEEE Transactions on Industrial Electronics*, 56(11), pp.4710–4717.

Jammu, N.S. & Kankar, P.K., 2011. A Review on Prognosis of Rolling Element Bearings . *International Journal of Engineering Science and Technology (IJEST)*, 3.

Kang, M. et al., 2015. Envelope analysis with a genetic algorithm-based adaptive filter bank for bearing fault detection. *The Journal of the Acoustical Society of America*, 138(1).

Liang, M. & Bozchalooi, I.S., 2010. An energy operator approach to joint application of amplitude and frequency-demodulations for bearing fault detection. *Mechanical Systems and Signal Processing*, 24(5), pp.1473–1494.

Liu, J., Wang, W. & Golnaraghi, F., 2008. An Extended Wavelet Spectrum for Bearing Fault Diagnostics. *IEEE Transactions on Instrumentation and Measurement*, 57(12), pp.2801–2812.

Liu, Z. et al., 2015. A Fault Diagnosis Methodology for Gear Pump Based on EEMD and Bayesian Network.

Loutas, T., Roulias, D., & Georgoulas, G. (2013). Remaining Useful Life Estimation in Rolling Bearings Utilizing Data-Driven Probabilistic E-Support Vectors Regression. *EEE Transactions on Reliability , 62* (4), 821-832.

Ming, Y., Chen, J. & Dong, G.M., 2010. *Weak fault feature extraction of rolling bearing based on cyclic Wiener filter and envelope spectrum*. Mechanical Systems and Signal Processing, 25.

Nataraj, C. & Harsha, S., 2006. The effect of bearing cage run-out on the nonlinear dynamics of a rotating shaft. *Communications in Nonlinear Science and Numerical Simulation*, 13(4), pp.822–838.

Nguyen, P.H. & Kim, J.-M., 2015. Multifault Diagnosis of Rolling Element Bearings Using a Wavelet Kurtogram and Vector Median-Based Feature Analysis. *Shock and Vibration*, 2015, pp.1–14.

Önel, I.Y. & Benbouzid, M.E.H., 2008. Induction Motor Bearing Failure Detection and Diagnosis: Park and Concordia Transform Approaches Comparative Study. *IEEE/ASME Transactions on Mechatronics*, 13(2), pp.257–262.

Peng, Y. & Dong, M., 2010. A prognosis method using age-dependent hidden semi-Markov model for equipment health prediction. *Mechanical Systems and Signal Processing*, 25(1), pp.237–252.

Purushotham, V., Narayanan, S. & Prasad, S.A., 2005. Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition. *NDT & E International*, 38(8), pp.654–664.

Rafsanjani, A. et al., 2008. Nonlinear dynamic modeling of surface defects in rolling element bearing systems. *Journal of Sound and Vibration*, 319(3-5), pp.1150–1174.

Rai, V. & Mohanty, A., 2006. Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform. *Mechanical Systems and Signal Processing*, 21(6), pp.2607–2615.

Randall, R.B. & Antoni, J., 2010. Rolling element bearing diagnostics—A tutorial. *Mechanical Systems and Signal Processing*, 25, pp.485–520.

Samanta, B. & Al-Balushi, K., 2001. Artificial Neural Network Based Fault Diagnostics Of Rolling Element Bearings Using Time-Domain Features. *Mechanical Systems and Signal Processing*, 17(2), pp.317–328.

Soualhi, A., Medjaher, K., & Zerhouni, N. (2014). Bearing Health Monitoring Based on Hilbert–Huang Transform, Support Vector Machine, and Regression. *IEEE Transactions on Instrumentation and Measurement , 64* (1), 52-62.

Sreejith, B., Verma, A. & Srividya, A., 2008. Fault diagnosis of rolling element bearing using time-domain features and neural networks. *2008 IEEE Region 10 and the Third international Conference on Industrial and Information Systems*.

Stack, J., Habetler, T. & Harley, R., 2005. Fault signature modeling and detection of inner race bearing faults. *IEEE International Conference on Electric Machines and Drives, 2005*.

Tobon-Mejia, D.A. et al., 2011. Hidden Markov Models for failure diagnostic and prognostic. *2011 Prognostics and System Health Managment Confernece*.

Tobon-Mejia, D.A. et al., 2012. A Data-Driven Failure Prognostics Method Based on Mixture of Gaussians Hidden Markov Models. *IEEE Transactions on Reliability*, 61(2), pp.491–503.

Sloukia, F.E. et al., 2013. Bearings Prognostic Using Mixture of Gaussians Hidden Markov Model and Support Vector Machine. *International Journal of Network Security & Its Applications*, 5(3), pp.85–97.

Soualhi, A. et al., 2014. Prognosis of Bearing Failures Using Hidden Markov Models and the Adaptive Neuro-Fuzzy Inference System. *IEEE Transactions on Industrial Electronics*, 61(6), pp.2864–2874.

Silva, J. & Cardoso, A., 2005. Bearing failures diagnosis in three-phase induction motors by extended Parks vector approach. *31st Annual Conference of IEEE Industrial Electronics Society, 2005. IECON 2005*.

Wang, T. et al., 2013. Rolling element bearing fault diagnosis via fault characteristic order (FCO) analysis. *Mechanical Systems and Signal Processing*, 45(1), pp.139–153.

Williams, T. et al., 2001. Rolling Element Bearing Diagnostics In Run-To-Failure Lifetime Testing. *Mechanical Systems and Signal Processing*, 15(5), pp.979–993.

Xiao, Y. & Ishikawa, T., 2005. Bearing strength and failure behavior of bolted composite joints (part I: Experimental investigation). *Composites Science and Technology*, 65(7-8), pp.1022–1031.

Yan, K., & Zhang, D. (2015). Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sensors and Actuators B: Chemical*, *212*, 353-363.

Yu, Y., Yudejie & Junsheng, C., 2005. A roller bearing fault diagnosis method based on EMD energy entropy and ANN. *Journal of Sound and Vibration*, 294(1-2), pp.269–277.

Zhou, H. et al., 2015. Detection and diagnosis of bearing faults using shift-invariant dictionary learning and hidden Markov model. *Mechanical Systems and Signal Processing*, 72-73, pp.65–79.
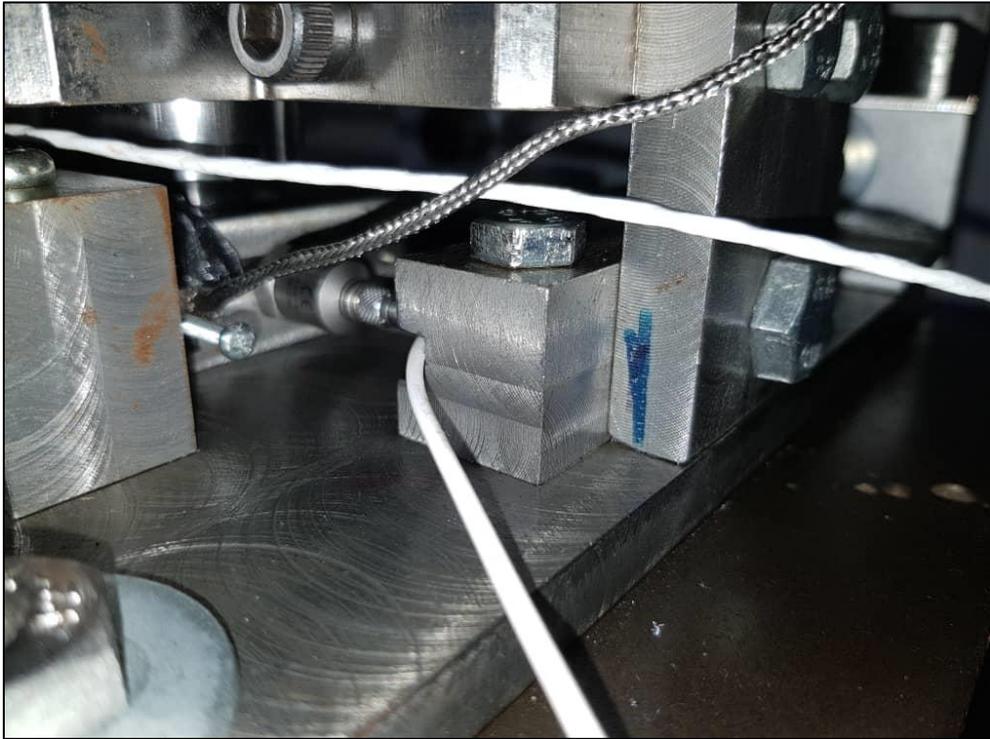
APPENDICES

Appendix 1: Disassembled Test Rig for Bearing Replacement
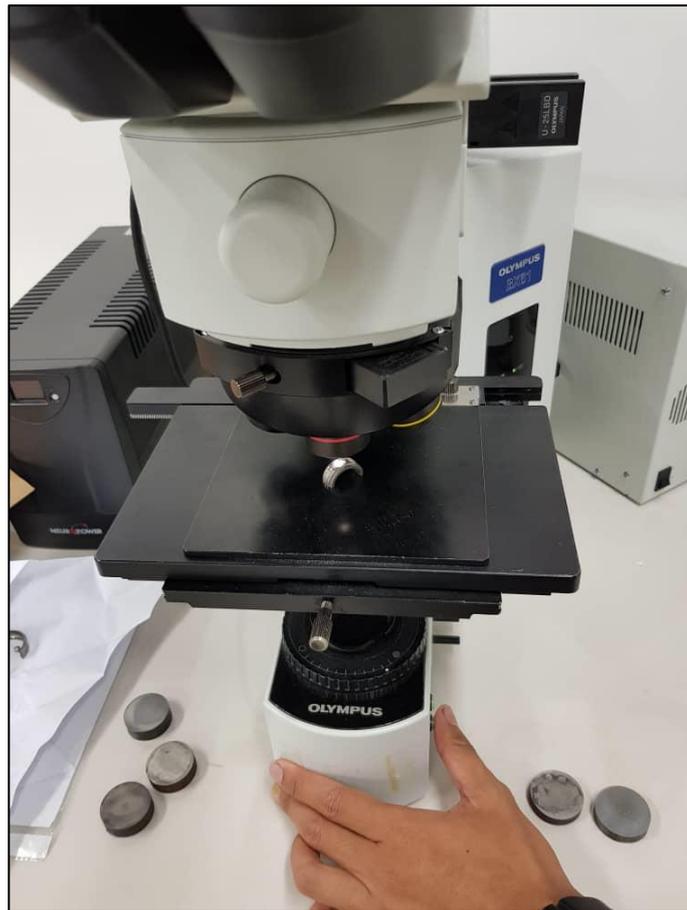

Appendix 2: Force Application Mechanism
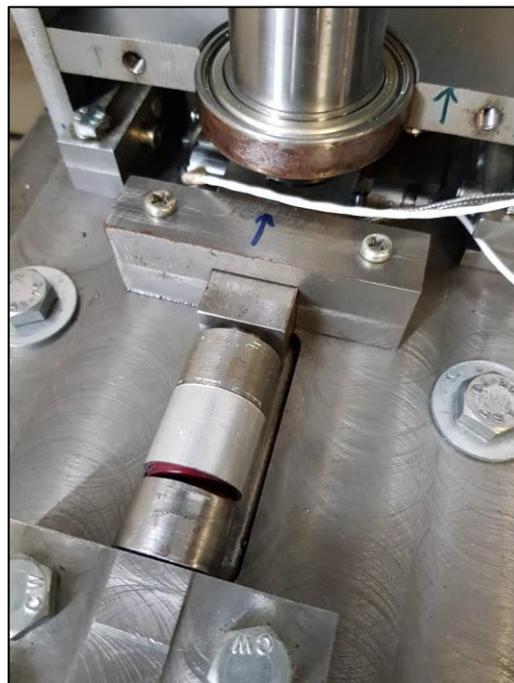
Appendix 3: Thermocouple Installation

Appendix 4: Electrical Discharge Machine Cutting of Bearing

Appendix 5: Observation of Bearing's Fault Mode Under Microscope

Appendix 6: Microscopic View of Bearing's Ball Flaking

Appendix 7: Purchase Order of New Bearings

Appendix 8: Push Belt Replacement