DIAGNOSIS AND PROGNOSIS OF ROLLING ELEMENT BEARINGS AT LOW SPEEDS AND VARYING LOAD CONDITIONS USING HIGHER ORDER STATISTICS AND ARTIFICIAL INTELLIGENCE

by

Henry O. Omoregbee

A thesis submitted in (partial) fulfillment of

the requirements for the degree

Doctor of Philosophy

in the Department of Mechanical and Aeronautical Engineering

in the Faculty of Engineering,

Built Environment and Information Technology

at the

University of Pretoria

Supervisor: Prof. P. S. Heyns

(2018)

Summary

Condition monitoring (CM) is commonly used in determining the operational states and health of rotating machines. The rise in the complexity of modern machines have led to advances in CM technologies to increase and improve product reliability and reduced downtime. Vibration and acoustic emission signals generated by these modern and complex machines are often immersed in background noise making it difficult to detect faults. Extracting signal features that are sensitive to faults still attracts considerable attention to detect and identify faults in rotating machines. This is especially so for low speed machinery under varying load and speed conditions.

Such conditions are found in many industrial applications and include draglines in the mining industry and large rolling mills in many materials processing environments. Condition monitoring techniques for stationary systems are inadequate at accurately detecting and diagnosing faults under such conditions. Using acoustic emission transducers are better options to the use of accelerometers for data generation for the analysis under these conditions because of its higher sensitivity to detect low energy level response because of the low speed condition.

While skewness and kurtosis have been used extensively in the condition monitoring of bearings and gears, higher order statistical (HOS) techniques have not found wide application in machine condition monitoring. This is because if a process is Gaussian then HOS provide no additional information that can be obtained from the second or higher order statistics.

There is however reason to believe that these HOS techniques could play an important role in condition monitoring, provided appropriate care is taken. In problems that are non-Gaussian, non-minimum in phase, nonlinear in behavior and robust to additive noise, HOS techniques like the 6^{th} order statistical moment (hyperflatness) could play an important role in its condition monitoring (CM). By applying this method, processing of acoustic emission signal at low speed and varying load condition could be very useful by providing details about the signal which the conventional second order statistics cannot.

HOS techniques are extensions of the better-known concepts of correlation (in time or space) power spectra. Higher order spectra are higher order Fourier spectral representations of third and higher order correlations or moments. With this approach simple representation and interpretation of the online extracted information could be made possible. For example, different

colours of light emitting diode (LED) could be used to indicate the types of faults, such that a non-expert or a simple classification algorithm may interpret the result.

Empirical models include the fields of regression and classification which are also collectively referred to as supervised learning and is dependent on its construction or optimization on large sets of representative data. In classifying faults in rolling element bearings (REBs) at low speeds and under varying load conditions, support vector machines for regression and genetic algorithm (SVMGA) which is a supervised machine learning algorithm can be used for its classification or regression problems. Using a technique called the kernel, transforms the data and finds an optimal boundary between the possible outputs. It does some complex data transformations, and then separate the data based on the labels or outputs defined.

The Hidden Markov Model has also found application in the diagnosis of fault in rolling element bearings, as it has been proven to diagnose incipient faults better but often requires a large data set. For this reason, the Bayesian Robust New Hidden Markov Model (BRNHMM) will also be used here to diagnose faults of two different categories: debris induced fault on roller bearing and a fault induced on the outer race of a roller bearing.

The model setup in this work was formulated and validated with the use of data generated from an experimental test rig originally designed by Aye for his PhD research, and was further improved on as part of this work. Simulated data was also used to validate the result obtained from the test rig.

Fault diagnosis and prognosis is achieved with the use of eXtended Takagi-Sugeno (xTS) fuzzy and recursive least square algorithm (exTSFRLSA) and support vector data descriptive (SVDD) method in this work. The (exTSFRLSA) has many applications of which one is the prediction of the sequence of state change, based on the sequence of observations. SVDD belong to the statistical learning theory class which is used here in this work to show the remaining useful life (RUL) of the bearing under study.

Although many models have recently been applied with good generalization of results in predicting the RUL of bearings where they integrate the statistics (like the Kaplan-Meier estimator, Mahanalobis distance, principal component analysis (PCA) etc.) and artificial intelligent (AI) methods (e.g. artificial neural network (ANN), feed forward neural network (FFNN), recurrent neural network (RNN) etc.), it has been proven that using only the statistical method is not sufficient to give good predictions, but a hybrid of both methods often yield good results. exTSFRLSA is used on most occasions for tuning, adjusting parameters and for adaptation in the propagation model by comparing predicted and measured defect sizes as in, hence the instantaneous rate of defect propagation of the bearing can be captured despite defect growth behavior variation.

Acknowledgement

I thank the Almighty God, my rock and sure support for His loving kindness, unlimited love and favour upon my life. He guided me and put songs on my lips and I will never cease to sing it out all the days of my life.

Secondly, I would like to express my sincere gratitude to my supervisor, Prof. P. S. Heyns for forbearing and believing in me, his consistent encouragement, support and understanding shown during this program despite my temporal handicap. His extensive vision and creative thinking have been the source of inspiration for me throughout this work. I am also very thankful to Prof. Y.A. Hamam for his supportive efforts towards the successful completion of my work, also for his thorough guidance and time he sacrificed to attend to me, offering very useful suggestions.

I want to appreciate the entire staff of C-AIM laboratory especially Mr. G. Breitenbach, Mr. H. Booysen, Mr. P Matsaola. and a host of others for their friendly and ever open-hearted support shown to me during the carrying out of my experimental works there.

Finally, my thanks go to my parents, late Pa I. Omoregbee and my mum Mrs. M. Omoregbee, my siblings and their families; Dr. Mabel Olanipekun, Rev. and Mrs. J. Edeojonore, Ms. Joy Ivbanikaro, Mr. and Mrs O. Isaac-Ivbanikaro, and to Anna and David, also to my niece and nephew Esther Ayiseosa and Samuel Isaac-Ivbanikaro, for their support financially, morally and otherwise despite the difficult circumstances. To my friends, Rev. Nelson and Pastor Ajayi, Dr. and Dr. Abe, Dr. and Mrs. E. Eberechi and to all my fellow colleagues, I say thanks to all for the support.

List of Publications

Omoregbee H. O. and Heyns P. S.; "Low speed rolling bearing diagnostics using acoustic emission and higher order statistics techniques", submitted to Journal of Mechanical Engineering Research and Developments (resubmitted July 2018).

Omoregbee H. O. and Heyns P. S.; "Fault classification of low speed bearings based on support vector machine for regression and genetic algorithms using acoustic emission", submitted to Journal of Vibration Engineering & Technologies (Accepted February 2018).

Omoregbee H. O. and Heyns P. S.; "Fault detection in roller bearing operating at low speed and varying loads using Bayesian Robust New Hidden Markov Model", Submitted to Journal of Mechanical Science and Technology (Accepted May 2018).

Omoregbee H. O. and Heyns P. S.; "Bearing fault prognostics using eXtended Takagi-Sugeno fuzzy with recursive least square algorithms" submitted to Journal of Mechanical Engineering Research and Developments (submitted June 2018).

Table of Contents

Glossary

Abbreviations	**Description**
1NF	First Normal Form
ACP	Asset Care Plan
AE	Acoustic Emission
AFRLSA	Adaptive Fuzzy Recursive Least Square Algorithm
AI	Artificial Intelligence
ALE	Adaptive Line Enhancer
ANC	Adaptive Noise Cancellation
ANN	Artificial Neural Network
ASL	Average Signal Level
BCF	Bearing Characteristic Frequency
BRNHMM	Bayesian Robust New Hidden Markov Model
CBM	Condition Based Maintenance
CM	Condition Monitoring
DDM	Data Driven Model
DSP	Digital Signal Processing
DSS	Decision Support System
ERM	Empirical Risk Minimization
FFNN	Feed Forward Neural Network
FFT	Fast Fourier Transform
FRB	Fuzzy Rule Based
FRF	Frequency Response Function
HFIW	Hyper Flatness Information Wave
HFRT	High Frequency Resonance Technique
HMM	Hidden Markov Model

HOS	Higher Order Statistic
IHM	Integrity Health Management
KL	Kullback-Leibler
KLW	Kullback-Leibler Wave
KPCA	Kernel Principal Component Analysis
KW	Kurtosis Wave
LDA	Linear Discriminant Analysis
LDN	Load Demodulation Normalization
LED	Light Emitting Diode
LGE	Linear Graph Embedding
LLE	Largest Lyapunov Exponent
LMS	Least Mean Square
MBD	Model Based Diagnosis
NN	Neural Network
PCA	Principal Component Analysis
PSD	Power Spectral Density
RDA	Rotational Domain Averaging
REB	Rolling Element Bearing
RLSA	Recursive Least Square Algorithm
RM	Risk Minimization
RMS	Root Mean Square
RNN	Recurrent Neural Network
RTRRMS	Response Type Road Roughness Measuring System
RUL	Remaining Useful Life
SA	Synchronous Averaging
SANC	Self-Adaptive Noise Cancellation

SK	Spectral Kurtosis
SKW	Skewness Wave
S/N	Signal to Noise Ratio
SRM	Structural Risk Minimization
STFT	Short Time Fourier Transform
SVDD	Support Vector Data Descriptive
SVMGA	Support Vector Machines for Regression and Genetic Algorithm
TSA	Time Synchronous Averaging

List of Symbols

Symbols	Descriptions
l	A vector of length D
$C(n)$	Complexity ratio
b	Constant
X_{CF}	Crest factor
D	Data
$f(\mathbf{x})$	Gaussian process
$KL(P_1 P_2)$	Kullback-Leibler Probability density for P_1 and P_2 distribution
$l(p)$	Length of sequence
W	Margin
$X_{\max}$	Maximum of X
X_{median}	Median of X
$X_{\min}$	Minimum of X
$\|X\|_2$	Norm
n	Number of components
N	Number of data points
M	Number of data points in a small region
$S(\theta)$	Posterior distribution
$p(\theta)$	Prior distribution

$p(o \mid \lambda)$	Probability of the observation sequence given the model
X_{range}	Range of X
X_{rms}	Root mean square
f_s	Sampling frequency
X_{sum}	Sum of X
$F(\theta)$	The negative free energy
X_{var}	Variance
w	Weight vector
f_A	Analysis frequency
λ	Eigenvalue
ε	Error function
φ	Hyperparameters of the GP
∞	Infinity
$\phi(x)$	Kernel function
$\in$	Margin of tolerance
μ	Mean parameter
φ	Parameter vector
π	Pi
C_{opt}, γ_{opt}	Regularization optimum parameters
C, γ	Regularization parameters

α	Smoothening constant
σ	Standard deviation parameter

1.0 Chapter One Introduction.

1.1 Problem statement.

Slow rotating variable speed electric motors exposed to changing loads are typically found with slew bearing applications such as in mining machinery. CM is complicated by the fact that these bearings work in two directions.

These bearings are expensive to replace, and critical to the production capabilities of the machinery under consideration. There is therefore a need for CM techniques to be able to operate on more complex machinery like these which are under non-stationary operating conditions. CM has been the subject of extensive research for the past decades. It remains a rapidly expanding field that has gained considerable ground with research and application in different field of machinery like in bearings and gearboxes (Heyns, 2013). The changing operating conditions affecting these machines are the main influence of variation in the energy levels of measured response data.

The three mainstays of CM are fault detection, diagnosis and prognosis. Determining that damage has occurred to a bearing comprises detection which is often the early stage of fault diagnosis, while diagnosis in itself is a determination of the location and type of fault, prognosis involves estimating of the remaining useful life (RUL) of the damaged bearing and investigation of failure modes (Jammu and Kankar, 2011). The strategy where maintenance decisions are based on the condition of the machine is referred to as CBM and it has being proven to be generally superior to older and more established maintenance strategies including run-to-failure and the time based maintenance (Aherwar and Khalid, 2012).

Machines like draglines found in the mining industry are difficult to monitor due to the changing conditions affecting them. The common Fast Fourier Transform does not apply in analyzing the signals from these machines because these signals are not periodic and hence require an all-together different approach. While varying load tends to cause amplitude modulation of the vibration signal, fluctuating rotational speeds often induce frequency modulation (Stander and Heyns, 2006). It is considering this that the present research aims to offer an appropriate but simple method to diagnose and relate the prognosis of REB using acoustic emission.

For critical components such as bearings which are widely used in rotating machines in low speed and varying load applications as is often the case with draglines in opencast coal mines, the faults often initiate and accelerate failures of other components which may finally result in machine breakdown. Therefore, accurate fault diagnosis is required. Often localized and distributed defects are formed in bearing by wear, flaking, smearing, and corrosion, rough treatment during assembly.

To date, many ways of analyzing acoustic emission signals in rolling element bearings have emerged, such as the use of the time domain analysis, frequency domain analysis and the time frequency domain analysis. But there has been a difficulty in using higher order statistics (HOS) techniques of the time domain analysis due to the problem of outliers that tend to affect its signal analysis when used. In this work, ways by which HOS can be used for analysis will be considered.

For the minimization of machine downtime, lapses in production and human casualties, a sensitive and robust monitoring system is needed to detect faults which is diagnosed with the use of higher order statistics (HOS) and their combination with Kullback-Leibler divergence which helps to analyze large deviation results (outliers) that could be useful to the system, including the asymptotic rate of decrease of error probability binary hypothesis testing and continuous random variables problems, hence making it applicable for clustering purpose and for deciding whether the samples come from the same distribution.

Low speeds often require the need for an unconventional sensor approach, such as acoustic emission (AE) because the response from a low speed system is usually low. This gives rise to low amplitude modulated signals which require high sensitivity, high gain sensors like as found in acoustic emission sensors.

It is hard to effectively predict the future propagation trend of a particular fault during its early stages, due to the strongly stochastic characteristics of the failure propagation process (Li et al., 1999) despite the fact that a large number of features can be extracted to characterize the AE data. It has been shown by (Xi et al., 2000; Yu, 2011) that many features are only effective for a particular fault in a particular propagation phase.

Normalizing the data for the changing load and speed is the processing of the data which is often the next stage in CBM and this entails improving the quality of the signal and extracting diagnostic information from the signal. Extracting diagnostic information that is robust to operating and environmental conditions, intuitive for ease of fault detection and classification, while being compact in size for reduced computational cost, is the aim of normalization/signal processing techniques (Heyns, 2013). Processed data should also be represented in such a way that the understanding of the results is made simpler, which may serve for direct interpretation by the maintenance engineer (Vinson, 2014) .

The final stage is being performed by experienced and trained personnel, rule-based system, automated classifiers and change detection alarms. It is usually based on the correct interpretation of the information extracted which do often result in detecting and classifying of faults. The rule-based system, automated classifiers and change detection alarms are often supervised machine learning algorithms such as neural networks (NN) e.g. artificial neural network (ANN), recurrent neural network (RNN), neuro-fuzzy network, etc. which have gained popularity and success in the field of CBM.

Once the early faults are detected with the use of Bayesian Robust New Hidden Markov Model (BRNHMM) after the application of HOS for the determination of the energy coherent factor, the support vector machines for regression and genetic algorithm (SVMGA) are used for its classification because it reduces classification complexity. Their performance degradation assessment and their remaining useful life (RUL) estimation are then conducted next to maximize the life-time of the critical component (which is the bearing in this case).

Due to the complexity associated with the prognostics of components that operate under non-linear and varying conditions, the support vector data descriptive method and the eXtended Takagi-Sugeno fuzzy combined with recursive least square algorithm are used in this work to predict the remaining useful life (RUL) and the estimation of the defect growth of the bearing respectively, as they both simplify the mathematical calculations and make possible the life span prediction.

The challenge of monitoring the condition of REB as found in applications such as in the cutting head transmission of a continuous coal miners and draglines on opencast coal mines running at low speeds and varying load conditions is addressed in this thesis.

1.2 Literature survey.

The need to monitor the condition of REB in the cutting head transmission of a continuous coal mines and drag line on opencast coal mines running at low speed and varying load condition has attracted greater interest to improve asset management. There is the need for signal processing techniques which can indicate degradation in REB condition under variable load and speed especially at low speed, as an effective condition monitoring systems and strategies will assist the scheduling of optimal maintenance intervals thereby minimizing unnecessary down time of production equipment.

In this survey, we investigate condition monitoring of a variable load and speed system in rolling element bearing, data acquisition methods, data processing methods, signal processing methods as well as fault diagnosis and prognostics. This is necessary as the scope of work covered in this research work relates to all of these aspects and are revisited in later chapters. A schematic of the approach is illustrated in figure 1.1 in the next page.

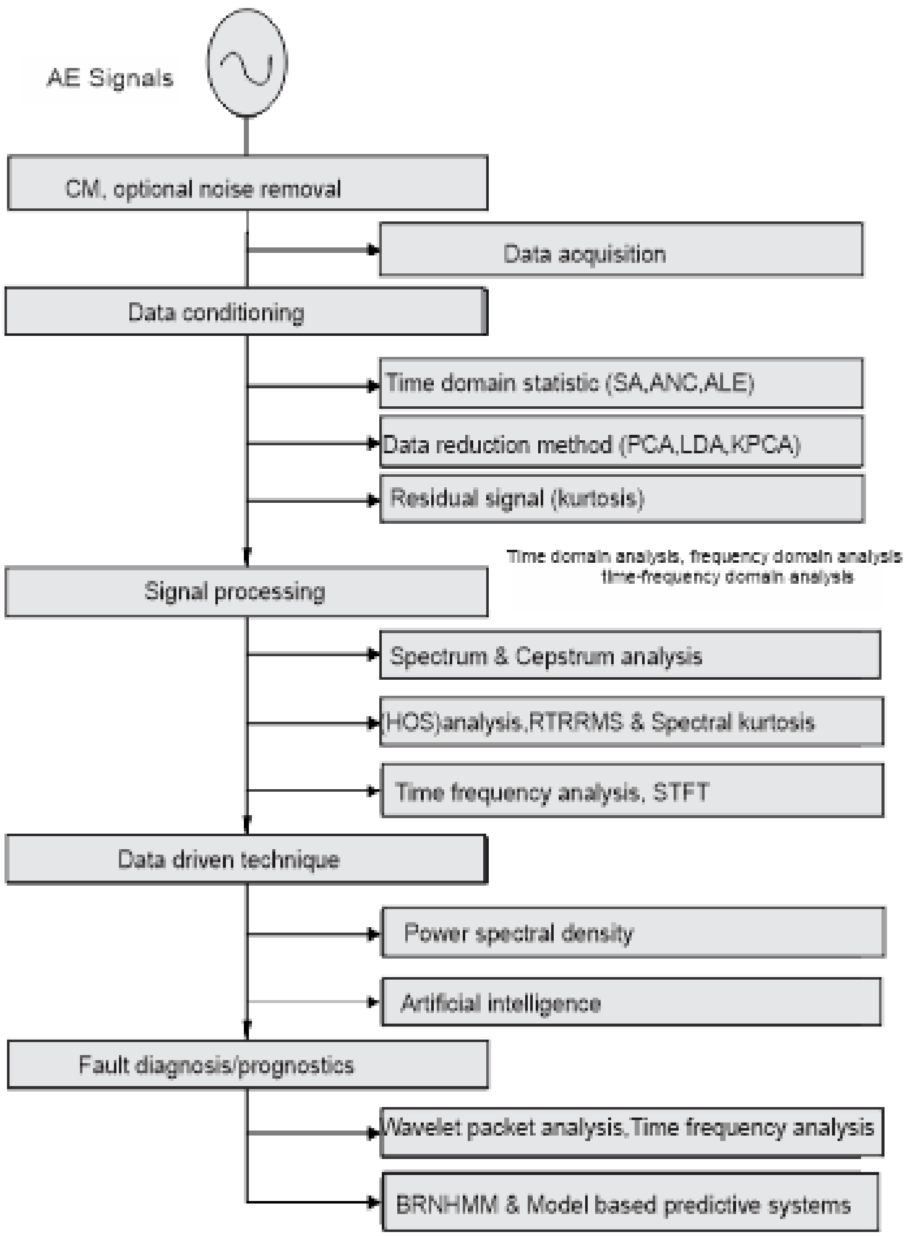

Figure 1.1, Schematic flow chart of literature review.

1.2.1 Condition monitoring of slow speed rolling element bearings.

Vibration of component parts occur because of metal to metal rubbing of each other. This often generates vibration signals which have been studied over time to carry useful information about the health state of machines after reasonable analysis has been done on the signal. But first these signals need to be collected with transducers like accelerometers. However, it has long been known from the literature that at low speeds the information carried by these signals are difficult to access because of the weak strength of the signals and the noisy environments that affect them. Later research reveals (Jamaludin, Mba and Bannister, 2001; EL Badaoui et al., 2001) that acoustic emission signals could be very useful for this sort of difficult conditions, and piezoelectric transducers are best suited for the signal collection. Acoustic emission signals are often used for fault diagnosis and prognostics for low speed studies. An in-depth literature on data acquisition is found in sub-section 1.2.1.1.

Despite the fact that response signals could be analyzed in the time, frequency and time frequency domains as indicated above, and many features could be generated from the vibration/acoustic data, extracting useful information from those original features as inputs to the diagnosis and prognostics systems and evaluating the performance degradation based on the extracted features becomes a big challenge due to its weak signal (Jianbo, 2011).

The main components of the rolling element bearing (REB) consist of the inner ring which is mounted on a rotating shaft, the outer ring mounted to a stationary housing, and the cage which separates the rolling element preventing contact between them during operation and the rolling elements which may be balls or rollers used for transferring load over a small surface (ideally, point or line contact) with the raceways (McInerny and Dai, 2003). These components often fail, and such failures generate a series of impact vibrations in short time intervals during diagnosis which occurs at bearing characteristic frequencies (BCF). The vibrations generated in the bearings are estimated based on the running speed of the shaft, the geometry of the bearing and the location of the bearing (Yam *et al.*, 2001).

Hence bearing failures can result from many sources which includes: mechanical damage, crack damage, wear damage, lubricant deficiency and corrosion, fatigue, debris contamination, misalignment, faulty installation, improper mounting, passage of foreign particles, excessive speed and inappropriate temperature. Wear damage results in gradual deterioration of the bearing component, leading to a loss of dimensioning and other associated problems while abrasive wear occurs when hard particles becomes entrained between the contact surfaces (McInerny and Dai, 2003).

Abrupt failure of bearings may cause malfunctioning of the entire system. Timely detection of these faults and estimation of the time to failure are areas of concern for researchers as this does often result in downtime for the system and economic loss to the customer. Soaring machinery repair and/or replacement costs could arise because of bearing failure which also has the potential to damage the machinery.

The following are type of rolling element bearing faults: rolling element line defect, inner race (case) line defect, outer race (case) line defect and case damage (EL Badaoui et al., 2001). Surface irregularities, pit and spalls on rolling surfaces, misaligned races, are often either localized or distributed defects (Tandon and Choudhury, 1999). Many condition monitoring (CM) techniques exist for acquiring data for bearing diagnostics and prognostics. These include AE analysis, vibration analysis, temperature data analysis, oil analysis and wear debris analysis.

It is often better to obtain data from rolling element bearings running at low speeds with acoustic emission transducers than with the use of accelerometer, as at low speed it becomes difficult to detect faults buried in noise especially when they are at their incipient stage, but with acoustic emission the signals are captured at very high frequency. Secondly, because acoustic emission transducers are very sensitive the weak signal generated (poor signal to noise ratio) from the low rotating bearing can be captured easily for analysis. Although signal discrimination can be very difficult with the use of acoustic emission, this work embeds the application of Kullback-Leibler divergence, the use of higher order statistics and statistic learning theories to solve this problem.

1.2.1.1 Data acquisition.

In this thesis, fault detection under variable operating conditions is covered. Unlike other work where the response data is acquired from a machine operating under constant and known operating conditions, a different approach is required with the case of variable operating condition. The most common approach used under constant and known operating condition is time synchronous averaging (TSA) a method which compensate speed fluctuation (Stander and Heyns, 2005). TSA is computed as an ensembled average of the time domain signal over many shaft revolutions and often an encoder or tachometer is needed. However, tacholess order tracking methodology has sprung up where accurate estimating of the phase of a shaft is made possible in the presence of large angular acceleration and noise (Schmidt et al, 2017).

There are types of data which can be acquired, and these are the incident data and the condition monitoring (CM) data. CM acquisition are usually automated through computer software and hardware such as LABView software and data acquisition card produced by National Instruments as found with this work. The data which captures events such as setting up, oil changes, slight overhauls, failures and repairs is referred to as incident data. The main purpose of data acquisition of acoustic emission signals is to measure the changes of the object such as the test bearing and the test environment, spall initiation and monitor of spall propagation detected with the help of the data.

The source of emission is closely associated with the dislocation accompanying plastic deformation with the initiation and extension of fatigue cracks in materials under stress, acoustic emission transducers are ideal for its use. The generation of transient waves during the rapid release of energy from localized sources within a material like in the case of metal to metal rubbing contact in a REB is referred to as acoustic emission with frequency range greater-than 100 kHz. As said earlier the piezoelectric transducers are examples of transducers used in the measure of acoustic emission signals and they often provide direct measure of failure mechanisms in action (Badaoui et al., 2001). The acoustic monitoring methods provide high sensitivity, non-invasive and localization of failure zone by time of arrival measurement. One of its drawbacks is in the fact that signal discrimination can be very difficult (yet extremely important for successful applications) because of noisy operating environments and a generation of weak signal.

1.2.2 Data conditioning.

A compact (low-dimensional) signal representation which reflects the presence of machine damage, usually free of signal errors and noise, and which is robust in time-varying operating conditions is often a good data (or signal). Representing the processed data in an intuitive manner, so that a physical understanding of the results can serve for good interpretation by maintenance engineer is what data conditioning is all about.

Signals in raw form can be very noisy, with a low signal to noise ratio and biased. There are several techniques for conditioning signal to make them useful for analysis. One of the often-employed techniques is to low pass filter the signal. To de-noise the acquired signal time synchronous averaging (TSA) is often used and the chosen cut-off frequency should be at least twice as high as the highest useful signal frequency.

Another efficient and well known algorithm-adaptive noise cancellation (ANC) which was proposed by (Gelle and Delaunay 2000) enhance the vibration to assist in bearing fault diagnosis and has been applied in rolling element bearing. Others like adaptive line enhancer (ALE) and self-adaptive noise cancellation (SANC) are other data conditioning methods but they have some drawback for example SANC comes with a delayed version of the measured signal itself and difficulty in parameter setting at adaptation phase.

Data normalization is all about reducing and eliminating data redundancy, it is also a process in which data model attributes are organized to increase the ideas of the entity types. Methods of normalizing/conditioning data includes synchronous averaging (SA), adaptive noise cancellation (ANC), data reduction methods, and residual signal analysis but due to their drawbacks as will be explained latter these methods cannot apply well to the data obtained from varying loads and speed condition as what is found in this research. When an entity type contains no repeating groups of data then it is in its first normal form (1NF). Data normalizations goal is to guaranty the quality of the data before being introduced into any learning algorithm. It speeds up training time by beginning the training process for each feature within the same scale and is often used alongside with artificial intelligent algorithms (Nayak, Misra and Behera, 2014). The subsections below present examples of normalization methods and describes where they are best suited for.

1.2.2.1 Synchronous averaging (SA).

An effective means of removing background noise from a measured signal and extracting the periodic components from the vibration/acoustic signal is through the means of synchronous averaging (SA). When the amplitude of a signal is synchronously averaged at a fixed angular increment for many shaft revolutions, phase information for the angular resampling can be obtained from a tachometer/shaft encoder which provides a certain number of pulses per revolution. This method cannot be applied in all situations (especially with varying load and speed conditions) since not all diagnostics signal content of interest is synchronous with the rotation of the shafts in a rotating machine especially in REB where the rotation of the rolling elements is not synchronous with the rotation of the shaft. In a work related to synchronous averaging (Bechhoefer and Kingsley 2009; Mba 2003) calculated that non-synchronous noise is reduced by the inverse of the square-root of the number of revolutions, but however, it is still not applicable in the condition of varying load and speed as found in this work.

1.2.2.2 Data reduction methods.

Data with high dimension is often difficult to deal with. By reducing the data into fewer principal components using data reduction techniques (like the principal component analysis (PCA), linear discriminant analysis (LDA), linear graph embedding (LGE), kernel principal component analysis (KPCA) etc.) which capture all the vital features helps to solve the problem associated with high dimensional data. Techniques like PCA seem very useful but can only be performed on a set of observations that vary linearly hence are being a drawback when variations are non-linear (Aye S.A, 2014). The Mahanalobis distance (was used for varying load and speed conditions) which very much does the same function as the PCA. It was used here in this work to reduce the principal components thus helping SVDD to describe the dataset by using a hypersphere with minimized radius in the feature. Detailed work on this can be found in chapter 5.

1.2.2.3 Residual signal analysis.

This is a method which aims at the removal of normal (or healthy) components from a signal so that only those components which are damaged related are considered. It is closely related to digital filtering technique. Under this method is example like self-adaptive noise cancellation (SANC) which is based on the least mean square (LMS) adaptive filter (Antoni and Randall, 2002) and the discrepancy analysis method. Residual signal analysis is said to be an effective tool in the detection and diagnosis of gear faults and the discrepancy method is now gaining more ground in bearing analysis.

1.2.2.3.1 Discrepancy analysis.

This analysis is like the residual signal analysis. It is a process where all healthy vibration components, such as the meshing frequency and its harmonics are removed from the signal, leaving only the damaged related vibration components. The envelope of the filtered signal is referred to as a discrepancy transform. The discrepancy signal tends to be simpler and smoother than the original vibration waveform and may thus be resampled using less accurate reference signal than would be required to resample the original waveform (Heyns et al., 2012 and Heyns et al., 2016). The discrepancy analysis is an on-going work that has seen significant breakthroughs by the Centre for asset integrity management (C-AIM) at the University of Pretoria, South Africa, for condition monitoring of gear system and REB.

1.2.3 Signal processing.

Signal processing is an enabling technology that encompasses the fundamental theory, applications, algorithms, and implementations of processing or transferring information contained in many different physical, symbolic, or abstract formats broadly designated as signals. It uses mathematical, statistical, computational, heuristic, and linguistic representations, formalisms, and techniques for representation, modeling, analysis, synthesis, discovery, recovery, sensing, acquisition, extraction, learning, security, or forensics.

Digital signal processing (DSP) refers to various techniques for improving the accuracy and reliability of digital communications. The theory behind DSP is quite complex. Basically, DSP works by clarifying, or standardizing, the levels or states of a digital signal (Moura, 2009).

Vibration/acoustic analysis of machine components is commonly used as a fault-detection technique employed in rotating bearing element systems (Karacay and Akturk 2009). These techniques can be classified into three domains: time domain analysis, frequency domain analysis and time frequency domain analysis (Jianbo, 2011), more detail about these techniques is found in the subsequent section.

1.2.3.1 Time domain analysis.

Statistical methods of analysis are often used to perform diagnostic on measured signals and it is as old as the science of measuring the signals. It entails calculating the root mean square (RMS), crest factor, kurtosis and the peak value of a signal. While the RMS gives an indication of the continuous or steady state amplitude in a time varying signal, the peak value gives half the difference between the maximum and the minimum values in the signal. The crest factor is the ratio of the peak value divided by the RMS of the signal which gives an indication of impulsiveness in the signal. Kurtosis is the normalized fourth order statistical moment which indicates also the impulsiveness in the signal. These parameters also known as the overall vibration parameters do not give nor provide any diagnostic information however, they are easy to implement in low cost online monitoring equipment. Depending on the quality and sensitivity of the extracted features used in estimating the bearing condition, the accuracy of the fault can be determined however, since each feature has known merits and demerits it is crucial to develop a systematic method that incorporates all the merits of the features extracted thus making the method more sensitive and robust to effect detection (Malhi and Gao, 2004).

The K-S test and the T^2 test are statistical based analysis which comes under statistical models. The K-S test was applied by (Andrade et al., 2001) to detect early fatigue cracks in gears. (Baydar and Ball, 2000) utilized the multivariate statistics in combination with PCA to detect localized faults in a two-stage helical gearbox, principal components were calculated for the normal condition and T^2 and square predictor error were used to calculate for new measurements. Although the time domain technique does not clearly give diagnostic information, the non-parametric statistical techniques are good indicators (when complemented by other method like the Lempel-Ziv complexity) of prevailing faults especially under the varying loads and speed condition as found in this work.

1.2.3.2 Descriptive metrics.

One of the earliest CBM systems is the descriptive metrics used to describe certain phenomenon from the time domain measurement. Parameters such as peak to peak values, skewness, shape factor, impulse factor, clearance factor, energy ratio etc. are often used in describing characteristics from the time domain under descriptive metrics. Descriptive metrics also tend to have a long history in the context of haul road monitoring where they are referred to as response-type road roughness measuring systems (RTRRMS) but on their own they are poor indicators and diagnostic measures cannot be obtained when used for the varying loads and speeds conditions.

1.2.3.3 Higher order spectral analysis.

Condition monitoring (CM) of rolling element bearings (REB) is commonly used in determining the operational state and health of machines to detect early stages of component degradation (Yan and Gao, 2004; Hong and Liang, 2009). At low speeds however, little energy is generated and conventional spectral based vibration monitoring becomes unsuitable (Jamaludin, Mba and Bannister, 2001).

Time domain parameters like crest factor, skewness and kurtosis have been used for many years in vibration monitoring. These parameters are however not sensitive enough to detect early faults, even with AE. It is therefore tempting to use higher statistical orders which would emphasize the effects of slight irregularities in the race even more and thus solve the problem of multiple events which sequentially occur at localized measurement points during testing in AE signal processing.

(Zaeri et al., 2011) suggest that the ability of skewness and kurtosis for fault detection decreases at low speeds. Since these indicators do not perform well at low speeds, there is a need to formulate better indicators using other higher order statistical parameters to be able to propose methods other than the traditional ones that could detect faults at lower speeds.

When calculating the statistical moments of signals measured in machines, the presence of outliers which result from noise generated by moving parts within and around the machine renders the estimated moments unstable and this could become severe for higher order statistical moments like skewness, kurtosis, hyperskewness and hyperflatness. For heavy tailed distributions, this implies that these estimates have high variance and are generally too unstable to capture the properties of the distribution (Niu, 2005). Higher orders expose one to the effects of outliers (measurement errors) which cause significant variance and no longer captures the correlation with damage.

Being nonlinear functions, the utilization of higher order statistical (HOS) techniques allow the analysis of the systems operating under the influence of random inputs, where the processes deviate from Gaussianity to indicate that there is a fault developing. This stems from the property of Gaussian processes to have zero higher-order spectra. HOS techniques provide high signal to noise ratio domains in which one can perform detection, signal reconstruction, if the time domain noise is spatially correlated. HOS parameters which are defined in terms of higher-order moments of the data (orders greater than 2), contain much information if one were to look beyond the power spectrum domain (Petropulu, 2000; Dube, Dhamande and Kulkarni, 2013; Ulus and Erkaya, 2016).

Higher order spectral (HOS) analysis is used to detect phase coupling due to non-linear interaction between the frequency components. The most commonly used HOS analysis is the bispectral analysis which is used to detect quadratic phase coupling. (El Badaoui et al., 2001) applied bi-spectrum analysis to synchronously sample helicopter transmission data. The higher order statistical moment like the hyper-flatness, hyper-skewness etc. are rarely used (but might prove to be good indicators if data are well conditioned) because of its demerit of being subtle to interference with outliers in data which could lead to wrong diagnosis in analysis. Here in this research they have successfully been applied for the case of varying conditions, non-linearity and non-stationary situation, more detailed work on this can be found in chapter 2.

1.2.3.4 Frequency analysis.

This analysis involves the transformation of the time domain signal to the frequency domain using fast Fourier transform (FFT). By extracting features from the frequency spectrum, it enables the CM of the machinery thereby focusing on the frequency of interest from the obtained frequency spectrum. This is often applied where there are harmonics in the periodicity of the signal which opposes the situation in this work which is under varying loads and speeds of which signal generated is highly non-periodic. It entails the conversion of a time domain signal to the frequency domain using fast Fourier transform (FFT). The term spectrum refers to amplitude representation versus the positive frequency range of the time signals' Fourier transform. The fact that in using spectrum analysis each discrete frequency can be monitored in contrast to the over-all amplitude is an advantage over the time domain analysis.

Defect frequency is often referred to as the frequency at which a certain defect on a component will cause an increase in the amplitude of the spectrum.

The most powerful bearing diagnostic techniques depend on detecting and enhancing the impulsiveness of the signals (Randall and Antoni, 2011), (Antoni and Randall, 2002) combine time and frequency analysis with the concept like spectral kurtosis, they also discussed "high frequency resonance technique" (HFRT) for the purpose of shifting the frequency analysis from very high range of resonant carrier frequencies to the much lower range of the fault frequencies, so that they could be analyzed with good resolution however, it should be noted that with AE very high frequencies are involved and the fault characteristic frequencies caused by the defective bearing and its harmonics are difficult to detect in the corresponding spectrum of the diagnosis by the conventional FFT-based envelope analysis, as it occurs within a narrow band spectrum. Several impact tests are also required by HFRT to determine the bearing resonance frequency, this makes it more computational expensive (Nelwamondo et al., 2006). The HFRT technique is usually applied under constant load and speed condition which negates its application in this work.

1.2.3.5 Cepstrum analysis.

Cepstrum analysis is a technique in which the spectrum of a logarithmic spectrum is calculated. It is a frequency analysis of a frequency analysis which is used to detect a series of harmonics or sidebands and to estimate their strength. Where various harmonics in a conventional spectrum is reduced to one peak in what is referred to as the quefrency domain, the periodicity in the conventional spectrum is therefore detected (Stander, 2005).

(El Badaoui et al., 2001) stated that cepstrum analysis is insensitive to the phase variations in the transmission path of a gear system. The product of the power spectrum of the source function with the squared amplitude of the frequency response of the transmission path of a gear system is said or known to be power spectrum of a signal measured at an external point on the casing of a rotating machine of a gearbox.

(El Badaoui et al., 2001) proposed a technique where a moving cepstrum integral was used to detect and localize tooth spalls in gears. This technique utilizes a moving window to isolate the tooth gear fault which enables the detection and localization of local tooth spalls on gear teeth. Cepstrum has been said earlier to be insensitive to phase variation which is one of its drawbacks even though it cannot be applied to non-stationary signals and this makes it not applicable under the varying loads and speeds as it becomes less sensitive especially under the low speed condition as signals under the varying and low speed condition are often weak.

1.2.3.6 Spectral kurtosis.

Spectral kurtosis (SK) is a technique used for the detecting and localizing the presence of non-Gaussian signal characteristics. It is used to detect frequency bands which are prone to impulsive characteristics which could be implemented as a filter to recover randomly occurring signal components which occur in the presence of stationary additive noise (Antoni, 2006).

SK is a useful signal processing tool (for signals that are stationary) in CM as many rotating machine faults e.g. bearing tend to generate impulse responses which can successfully be detected. SK was used by (Barszcz and Randall, 2009) to experiment the detection of tooth crack in a planetary gear of a wind turbine SK proved superior to the popular synchronous averaging in detecting this fault. It cannot be used for conditions that are varying and non-stationary as is the case in this research.

1.2.4 Data driven modelling techniques for CBM of rolling element bearing.

The data driven approach to condition monitoring consist of feature extraction and fault classification. The transformation of measurement data to expose patterns and signatures within the data thereby giving insight into the condition involves feature extraction. The process of feature extraction exposes fault frequencies in the signal data that are synchronous with the rotating bearing through a series of signal processing techniques consisting of digital re-sampling, power spectral density (PSD) computation and feature reduction.

They could either trace the cause of fault(s) responsible for a system tending towards failure or which had failed (for which they are basically referred to as diagnostic system whose inputs are sets of symptoms), or they could be diagnostic system diagnosing the system of interest from a model(s) set to have the correct behavior. Having made to learned from this model(s) could relate faults and symptoms inherent in them.

The model based diagnostic system have suffer some setbacks in the past because often the physical system is often modeled from a linear perspective, but real-life conditions are non-linear and unpredictable in nature, as most of the models bear their roots from linear models and hence precise system behavior are often not well modeled. But with the advent of neural networks, the model based diagnostic system has overcome this set back. Unlike the rule based diagnostic system that makes decision(s) from rules adopted from standard fundamentals practiced over time, model based diagnostic systems deal with models describing the behavior of system mathematically, logically etc.

Neural networks are networks of neurons made to mimic how the human brain works. It simply follows the pattern of the structure of cerebral cortex portion of the brain and thus does reasoning or manipulate information the way the brain does. With the advent of neural network Artificial

Intelligence (AI) has been able to classify input patterns in predefined classes or to categorize the patterns by grouping them into their similarity (Huang et al., 2007).

1.2.4.1 Data-driven modelling approach.

Data driven model devices are used to capture the relationships between the relevant input and output variables and do not represent the physics of a modelled process. Since they are based on objective information (i.e. the data) such devices could be more accurate than process models and the latter may often suffer from incompleteness in representing the modelled process. Data driven approaches are an attractive option to integrity health management (IHM) since it does not require complex models and can be applied to many types of systems. They use real data obtained from data acquisition system and track features revealing the components degradation and to forecast the global behavior of a system. Data driven techniques include wavelet analysis, statistical analysis, neural network analysis, and frequency domain analysis (Marton et al., 2013).

Manually classified training data or fault histories are not required by data driven model. Data driven models represents the machine response signal as a function of instantaneous state space variable, including the machine state (e.g. GPS coordinates, angular position), and the operating conditions (e.g. speed, load, temperature).

It becomes possible to reconstruct the machine response under different conditions, with all the causalities in the data adequately represented by the data driven model. Reconstructing the machine response is possible at a specified steady state operating condition (e.g. constant shaft angular speed) so that it subsequently becomes possible to apply spectral analysis. The machine response may also be reconstructed at a standardized operating condition so that it may be compared in a consistent manner with other measurements (Heyns, 2013).

1.2.4.2 Data-driven prognostic methods.

System behavior monitoring using regularly collected condition data instead of using comprehensive system physics or human expertise are often modeled by data-driven methods (Heng et al. 2009). Data-driven approaches are classified into two categories in general. These are statistical and machine learning approaches. While statistical approaches build or make models by fitting a probabilistic model to the available data, machine learning approaches seek to make out complex patterns and make intelligent decisions based on empirical data. The degradation patterns of sufficient samples representing equipment failure progression is been used by both the statistical and machine learning methods which is a requirement that makes the major challenge in data-driven prognostics since it is often not possible to obtain samples of failure progressions.

In industrial systems plants are not allowed to run until failure due to its consequences especially for critical systems and failure modes. But quality and quantity (sample size) of system monitoring data has a high influence on data-driven methods. Prognostic datasets sample sizes in the literature range from 10 to 40 (Gebraeel, Lawley and Ryan, 2005). Defining a reasonable failure threshold is difficult, especially when limited historical failure data is available (Bolander et al., 2009).

1.2.5 Fault diagnosis of slow rotating REB.

At the tail end of section 1.2.3.1 we discussed that statistical techniques are good indicators (when complemented by other method like the Kullback-Leibler divergence and Lempel-Ziv complexity) of prevailing faults especially under the varying loads and speed condition as found in this work and so also are other techniques like the support vector machine for regression and genetic algorithm, support vector data descriptive (statistical learning theories) and Bayesian robust new Hidden Markov model (a non-parametric statistical method) as will be revealed sub-section 1.2.5.1 and 1.2.5.2 and in the later chapters of study in this work.

Statistical moments tend to describe the shape of the amplitude distribution of vibration/acoustic data collected from a bearing and are sensitive to the repetitive impulses in fault detection. As discussed earlier in previous sections many other techniques have sprout up for fault diagnosis which include the frequency domain techniques and the time-frequency techniques, but these methods have some drawback as it relates to varying load and speed, especially at low speed conditions. In previous research on the application of statistical moments to condition monitoring in rolling element bearings, the third and fourth normalized central moments which restrain the selective range of statistical parameters have been considered and these moments are expected to increase no matter what the orders are (Niu, 2005). Outliers in HOS are seen to have irregular results that are not well correlated with magnitude of fault as it relates to their use hence, it becomes necessary to do something about the outliers.

The root mean square (rms), peak value, crest factor and kurtosis have been combined with high frequency resonance techniques and an adaptive line enhancer to detect and localize the damage in rolling bearings (Williams, 2001). (Tandon and Choudhury, 1999) showed that probability density function is correlated with bearing defects. Kurtosis, the fourth order moment is a statistical measure used to describe the distribution of observed data around the mean as it measures the relative peakedness or flatness of a distribution as compared to a normal distribution. It defines the degree to which a statistical frequency curve is peaked (Zaeri et al., 2011) and has been proven effective and useful when combined with other algorithms like Kullback-Leibler divergence (Yang, Mathew and Ma, 2003; Wang and Chen, 2009).

A good indicator for identifying damage in low speed machinery is the statistical index kurtosis which must be used with global rms values and the time signal. Bearing condition is categorized as normal if it has a kurtosis value below three, and un-healthy if the kurtosis exceeds three. As mentioned earlier, on their own they are sensitive to interference from undesired signals and noise and they are not effective for detection of incipient defects (Behzad, 2011; Graney and Starry, 2012).

It was concluded by (Dube, Dhamande and Kulkarni, 2013) that kurtosis is not a good indicator of faults on the outer race of a bearing. The most powerful bearing diagnostic techniques depend on detecting and enhancing the impulsiveness of the signals (Randall and Antoni, 2011), Randall and Antoni combine time and frequency analysis with spectral kurtosis, they also discuss “high frequency resonance technique” (HFRT) for the purpose of shifting the frequency analysis from very high range of resonant carrier frequencies to the much lower range of the fault frequencies, so that they could be analyzed with good resolution. However, it should be noted that with AE very high frequencies are involved and the fault characteristic frequencies caused by the defective bearing and its harmonics are difficult to detect in the corresponding spectrum of the diagnosis by the conventional FFT-based envelope analysis, as it occurs within a narrow band spectrum.

Model based diagnosis (MBD) systems refer to an area of artificial intelligence. They could either trace the cause of fault(s) responsible for a system tending towards failure or which had failed for which they are basically referred to as diagnostic system whose inputs are sets of symptoms or they could be diagnostic system diagnosing the system of interest from a model(s) set to have the correct behavior and having made to learned from this model(s) could relate faults and symptoms inherent in them. Unlike the rule based diagnostic system that makes decision(s) from rules adopted from standard fundamentals practiced over time, model based diagnostic system deals with models describing the behavior of system mathematically, logically etc.

The neural network (NN) is a powerful building tool for a wide class of complex nonlinear systems especially in the control strategy application. Two classes of control application are in the open loop identification and closed loop feedback control (Chin-min et al., 2011). Chin-Min Lin et al. made it clear that the basic problem in neural network closed loop feedback control is the providing of an on-line learning algorithm that does not require preliminary off-line training.

The artificial intelligent system is a system which can mimic humans in reasoning and taking action but their structural design are based on or focus from three different categories as proposed by (Jardine, Lin and Banjevic, 2006).

“Rule-based diagnostic system”: Under this system, rules are outlined strictly for diagnostic purposes and adhered to. These rules could be basically structured in relation with each possible fault that can be identified to each component’s state of condition under use. For this system to be effective, the rules must be made by experts of the field of consideration who have got vast experience and knowledge relating to each possible fault that could arise.

“Case-based diagnostic system”: This is a system made to learn from the history of the machine condition to diagnose fault arising. This sort of system could be very effective but will often require precise history of the machine state when in good performance and will need to update itself on the life cycle of the machine as it comes in use. They are utility systems that often require large data base capacity and could sometimes raise or make wrong decisions due to the non-linearity inherent in most machine working condition.

“Model-based diagnostic system”: This is another intelligent system that is highly efficient for fault diagnosis and prediction. Such systems are often made to comprise different mathematical, neural network and logical methods to aid in diagnostic reason. With a model based diagnostic system fault prediction or diagnosis are made from the comparison of the real monitored condition of the system and the model design system. The model-based diagnostic system often outperforms the other two diagnostic systems because a decision is made from best practice analysis.

1.2.5.1 Identified need.

There is however the need for more supportive algorithms based on statistical non-parametric methods to aid the HOS method. This allows for online fault diagnostics and prognostics as the derived formula achieved in this work based on HOS, only relates determining the energy coherent in the signal at low speed to which the algorithms derived in this thesis can be augmented to include other aspects of statistical learning approach. At such low speed, the energy generated from bearing defects might not show as an obvious change in signature and thus become undetectable using conventional vibration measuring equipment. A few of these supportive algorithms to further help in fault detection and useful life estimation are discussed below.

1.2.5.2 The Bayesian robust new Hidden Markov model.

Vibration and acoustic sensor signals are usually measured and compared to reference measurements to determine bearing conditions. Several approaches can be used for the analysis of these signals, which include time domain analysis, frequency domain analysis and time-frequency analysis. Of these, frequency domain analysis is the most commonly used because of the simplicity of application of the Fourier transform and its ease of interpretation (Li et al., 2000). Frequency domain methods however do require that the bearing defect frequencies must be known or estimated, and interpretation becomes more difficult when the signal to noise ratio is low. Frequency domain methods also tend to average in transient vibrations and therefore becomes sensitive to background noise (Ocak and Loparo, 2001; Nelwamondo, Marwala and Mahola, 2005).

(Li et al., 2000) used neural networks to diagnose motor rolling bearings by combining it with the time-frequency domain analysis for the diagnosis of a fault. (Lin et al., 2011) used the radial basis function (RBF) neural network and the robust adaptive controller (RAC) to analyze the stability of a non-linear system. In another instance (Hariharan and Srinivasan, 2009) used a new approach for the classification of rolling element bearing faults as they combined the RBF network and the probabilistic neural network (PNN) to achieve this. With these investigations and some others, it was found that fault classification during the incipient stages is very difficult to achieve, especially under varying load and speed conditions. This led to researchers exploring statistical approaches which are very effective for fault diagnosis under varying load and speed conditions, especially for faults during their incipient stages (Derouiche et al., 2012).

The Hidden Markov Model (HMM) is a non-parametric statistical method which has a strong capability of pattern classification and is suitable for dynamic time series of signals that are non-stationary, has poor repeatability and reproducibility. HMMs are often referred to as the ‘gold standard’ for the difficult task to perform speech recognition (Baruah and Chinnam, 2005; Li et al., 2005). (Nelwamondo, Marwala and Mahola, 2005) used HMM combined with the use of Mel-frequency cepstral coefficient, fractal and Gaussian mixture models for early classifications of bearing faults. (Geramifard, Xu and Kumar, 2013) used a HMM based semi-non parametric approach for fault detection and diagnosis in synchronous motors. (Soualhi et al., 2012) used HMM and combined it with neural network for the detection and diagnosis of fault in induction motors, but in all these works HMM was not as successful as expected for online fault detection, as large data sets are usually involved.

To address this (Dorj and Chen, 2013) were the first to use a Bayesian Hidden Markov Model-based approach for detecting anomalies in electronic systems.

The fact that signature of a defective bearing is spread across a wide frequency band and often masked by noise poses great difficulty of fault detection in bearings, especially in the early identification of the defects. Hidden Markov models (HMM) have proven to be very effective in the diagnosis of faults in bearings especially at the incipient stage. It is a stochastic signal model being referred to as Markov sources or probabilistic functions of Markov chains which is a random process of discrete-valued variables that involves many states. The actual sequence of states is not observable hence the name hidden Markov. The Hidden Markov Models are grouped into three categories namely; discrete, continuous and semi-continuous HMMs. The fault detection in bearing using vibration/acoustic signal finds its grouping under the continuous

HMM. The difference between these three groups of categories is defined by the use of different output production probabilities (Nelwamondo et al., 2006).

In this work, a Bayesian robust new HMM (BRNHMM) will be used and compared to neural network pattern recognition (as shown in chapter 3) for fault classification and detection under varying load and speed conditions, because it has proven very effective in modeling both static and dynamic signals and is suitable for online fault detection, since small data samples are used in its diagnostics. The Kullback-Leibler divergence was effectively used to access the divergence to the probability function of the BRNHMM and used to find its lower bound approximation.

1.2.5.3 Introduction to support vector machines for regression for fault classification.

Fault classification is critical to detecting faults, because quick classification and corrective measures lead to minimized plant downtime and increased throughput. In some applications process, trial and error by way of inspecting is one way of finding the cause of fault, which is usually slow and tedious leading to less fault classification. Another way of classifying faults is through the exploration of sensor and process data to gain engineering intuition to what is the cause of the fault. This method could be said to be a primitive form of pattern recognition which is also slow since the engineer has to excogitate through a large amount of data, and relies heavily on empirical engineering understanding of the process (Goodlin et al., 2002).

Other new methods have focused on automating the fault classification with automatic pattern recognition. The principal component analysis (PCA) approach is an example of automatic fault classification, where the dimensionality of the input features is reduced for both supervised and unsupervised classification purposes. When referring to classification we focus on two major issues; the defect classification which refers to the identification of defective components and severity classification which refers to the differentiation of defective bearings based on defect size, which can be estimated by the magnitude of a representative feature. Since no guidelines are generally available as to which feature is more sensitive for a particular defect condition, developing an effective feature selection scheme is essential to improving the accuracy of defect severity classification (Malhi and Gao, 2004).

The operational state and health of a machine is often determined through condition monitoring. Several factors may affect the operation of a machine and these may include machine speed, load, lubrication, alignment, etc. and often significantly affect the machine life. Researchers often distinguish different stages during the bearing life cycle, which are traditionally referred to as “infant mortality, useful life, and wear out”, “good, damaged, failure imminent” or “pre-failure, failure, near catastrophic” (Hamadache and Lee, 2014). Different vibration techniques have been developed to detect the faults in rolling element bearings (REB). This has been done, with two main purposes: firstly, to separate the bearing related signal from other components and to minimize the noise that may mask the bearing signal. The widely used techniques for this purpose are self-adaptive noise cancellation (SANC), adaptive filtering, synchronous averaging and discrete random separation. The second purpose is to identify the status of the bearing, thereby distinguishing the good and the faulty bearings to indicate the defective component (Hamadache and Lee, 2014).

An area that has gained much recognition in handling AE and vibration data in fault classification and recognition is Artificial Neural Networks, which has the capability of mimicking human experience gained in pattern recognition. Neural Networks may be designed to classify input patterns in predefined classes or to categorize the patterns by grouping them according to their similarities. They may also be designed to respond in real time to the changing system state descriptions provided by continuous sensor inputs. Artificial-intelligence-based fault diagnostics methods have the potential to tackle problem without human experience, with a method which aims at recognizing different machine health conditions via the features extracted from the AE signals, with the accuracy of the identification of these conditions been further enhanced through classifiers that exhibit good performance (Shen et al. 2014; Hagan and Beale, 1997; Lu 2007).

Support vector machines (SVMs) on the other hand are based on statistical learning theory that are special for solving learning problems of a smaller sample numbers that provide better generalization than ANN and guarantee the local and global optimal solution to be the same. SVMs are hence introduced into rotating machinery for fault diagnosis with high accuracy and good generalization for smaller sample numbers (Yang, Yu and Cheng, 2007). SVMs represent a machine learning approach that is widely used for data analysis and pattern recognition. The algorithm was developed by Vapnik and the current standard incarnation was proposed by Cortes and Vapnik. SVMs have well defined formulations which are consistent with mathematical theory(Yang et al. 2005; Yang et al. 2007).

1.2.5.4 Support vector machines for regression and genetic algorithms for fault classification.

This work further provides more insight into sub-section 1.2.5.3 where it is critical to detecting faults, because quick classification and corrective measures are needed to minimized plant downtime and increased throughput. SVMGA is a technique of mono-class classification, which tries to solve the problem of best features selection by applying the principles of evolution by optimizing the statistical components selected for the SVM for regression training solution. By optimizing, it solves the problem of outlier' detection and reduces the dimensionality of the data.

The choice of features however can affect the performance of classification as features generated are often refined to try to achieve the desired level of performance. However, developing features manually can be time consuming; generated features should have the ability to identify complex relationships within large data dataset where the mapping from data to class labels is often obscure (Guo, Jack and Nandi, 2005).

Many research works dealing with support vector machine and regression for fault diagnosis have been reported in the recent literature. A few examples of these include fault diagnosis of a rolling bearing based on feature extraction and neural network algorithm (Unal et al., 2013), bearing fault diagnosis based on KL transform and support vector machine (Lu, 2007), and an investigation for fault diagnosis based on a hybrid approach using wavelet packet and support vector classification (Li, Jiang and Xiang, 2014). However, none of this work deals with the preferred selection process of features for best performance. This technique (SVMGA) which aims at enhancing bearing fault diagnostics is developed by the fusion of multiple statistical features through SVMGA another dimension to what we have seen in chapter 3, where a selection of best features from experimental data is done at the classifier training stage.

Unlike the target value of a SVM which can only be used to handle a binary problem, the target value of a Support Vector Machine for Regression (SVMGA) is continuous, having shown great potential in time series prediction and can also establish a stable nonlinear relationship between inputs and outputs. Outputs of small deviations from their target values are formed when the feature vectors extracted from the samples belonging to the same class are fed to the trained SVR (Lu, 2007; Widodo et al., 2009). In the article by (Li & Chen W., 2014), a complex dynamic model for aligning roller bearings was established, thereby studying the problem of surface damage, preload and radial clearance. (Li and Chen, 2014) worked on a physical model and one-class support vector machine for rolling bearing fault diagnosis.

SVM was originally designed for binary classification (Sloukia, Bouarfa and Medromi, 2013). Some binary classification problems do not have a simple hyper-plane as a useful separating criterion. For SVMs problems, there are variants of mathematical approaches that retain nearly all the simplicity separating hyper-plane such as "one-against one", "one-against all" like in the classification of health states.

SVMGA have distinct advantages over SVM and other AI algorithms like ANN. because it is used for classifying faults which is over two classes that the SVM supports and for the fact that SVMGA uses high dimensional input space that does not necessarily depend on the number of features, because they have the potential to handle large feature spaces (Unal et al., 2013; Gunn 1998). This has been shown in applications such as in text categorization; biological studies etc. This has however not been applied to low speed bearings which are sinusoidally loaded along axial and radial directions. This emulates real life scenarios where bearings in functional equipment may be loaded with varying cyclic loads at frequencies that might be unrelated to the equipment rotational frequency and its harmonics, due to auxiliary equipment (such as pump) mounted on the machine and running at its own rotational frequency. Application of SVMGA to low speed bearings potentially holds distinct advantages for the following reasons:

- High training speeds.
- Good classification accuracy.
- Unlike the target value of SVM which is only used to handle binary problem, the target value of SVMGA is continuous.
- Feature vectors extracted from the samples belonging to the same class when fed to the trained SVMGA produces outputs of small deviation from the target values (Changqing Shen et al., 2014).

Therefore, these properties are more attractive to be used for solving multi-class problems to when compared with SVM based classifier or other AI algorithms. (Li, Jiang and Xiang, 2014) used a hybrid method that combines wavelet packet transform (WPT) and support vector classification (SVC) to deal with the difficulty to obtain large number of fault samples under practical condition for mechanical fault diagnosis. (Xiang and Zhong, 2016) developed a novel personalized diagnosis methodology and used it to investigate shaft unbalance, misalignment, rub-impact and their combinations. The method looks promising and the probable tools needed for it are numerical simulation (including finite element method), big data analysis or a combination of the two approaches. The associated draw-back with this method is its cumbersomeness and the large data involved, which is not appropriate for online solution as is the case found in this work. (Tong et al., 2017) used a Redundant Second Generation Wavelet Packet Transform (RSGWPT) and Local Characteristic-Scale Decomposition (LCD) to detect and extract fault feature from vibration signals, this method is based on the energy ratio in obtaining a set of desired intrinsic Scale Components (ISCs) while (Fatima, Mohanty and Kazmi, 2016) did a fault classification and detection in a rotor bearing rig by using SVM and a compensation distance evaluation technique for selecting two sensitive features from the twelve time domain features of the measured vibration signals from each of the accelerometer used.

1.2.5.5 Wavelets and packet analysis.

Wavelets are used to compute the time-scale representations rather than the time frequency representations and are used to convolve different dilations and translation of the mother wavelet with the signal of interest. It is found that at low scales (high frequencies) the wavelet enables a good time resolution while the reverse is the case at high scales (low frequencies), as good scale resolution is often obtain at the expense of poor time resolution (Rafiee et al., 2009).

For features extraction, wavelet analysis has been used a lot (Peng and Chu, 2004), also in the detection of gear and roller bearing damage in the presence of time-varying operating conditions and for the monitoring of the condition of bridges based on the detection of singularities in the vibration response of the bridge when traversed by a vehicle. Wavelet is complex to implement as there is no general method to select the wavelet function for a monitoring task (Yen and Lin, 1999)**,** and for the fact that we are looking for something easy to implement on online monitoring the wavelets and packet analysis was not chosen**.**

1.2.5.6 Time-frequency analysis.

Whereas the frequency domain analysis emphasizes the periodicity in signals and is therefore not well suited to study non-stationary waveform signals the time-frequency analysis proves useful in this regard. Many time-frequency techniques have been devised to better analyze non-stationary signals. This is often done by plotting the signal energy (or power) as two dimensional functions of both time and frequency causing the induced fault signal pattern to be seen better (Heyns, 2013). Some of the popular time-frequency transforms used for CM of rotating machines include the short-time Fourier transform (STFT), the Wigner-Ville distribution and the Choi-Williams distributions (Bartelmus, 2003). But despite the potential inherent to these techniques, their application is still limited in practice owing to the difficulty of quantitative interpretation.

1.2.6 Prognostics.

Faults in bearings often result in severe vibration of rotating machinery and timely detection of faults goes a long way to reduce downtime for the system and financial loss for owners. The ability to detect and sometimes isolate faulted components and its failure condition is known as diagnosis while prognostics is defined as the capability to predict the progression of a fault condition of a component failure and to estimate the remaining useful life (RUL) (Baruah and Chinnam, 2005). According to the International Standard Organization (ISO), failure prognostics is said to be the “estimation of the time to failure and the risk for one or more existing failure modes” (Tobon-Mejia *et al.*, 2012).

Traditional maintenance actions could either lead to blind proactive (where preventive maintenance is done without current input from the component) or it could be purely reactive i.e. fix a unit after it fails. Condition Based Monitoring (CBM) is well suited for finding faults related to vibration monitoring and equipment’s available to perform these are the sensors, digital signal analyzer etc. Vibration signals can be collected and well analyzed to detect faults (diagnostics) and their location and predict failure time for such system to fail in the nearest future (prognostics), giving rise to easy development of models and other unique processes for use like the adaptive recursive least square algorithm (Yen and Lin, 1999; Nelwamondo, Marwala and Mahola, 2005; Benkedjouh et al., 2012).

An effective prognostics program often gives sufficient time to schedule a repair and to acquire ancillary components before catastrophic failures occur. Bearing prognostic activity aims at anticipating the failure date by predicting the future health state of the bearing system and its Remaining Useful Life (RUL). Numerous tools and method can be used to predict RUL, these methods are classified into two principal approaches: model-based prognostics also known as physics of failure prognostics and the data-driven prognostics.

While model-based prognostics is involved with the prediction of the RUL of critical physical components where mathematical or physical models of degradation are used such as crack by fatigue, wear, corrosion, etc., the data-driven prognostics aims at transforming the sensored data into relevant models which can either be parametric or non-parametric (Gebraeel, Elwany and Pan, 2009; Medjaher, Tobon-Mejia and Zerhouni, 2012).

Many models have been applied with good generalization of result in predicting the remaining useful life (RUL) of bearings where they integrate statistical means (like the Kaplan-Meier estimator, Mahanalobis distance, Principal Component Analysis (PCA), Hidden Markov Model (HMM) etc.) and artificial intelligent (AI) methods. (Gebraeel et al., 2004) applied neural network for residual life predictions from vibration based degradation signals and latter (Gebraeel et al., 2005) used Bayesian approach for estimating the residual-life distributions from components, (Baruah and Chinnam, 2005) in their work used Hidden Markov Model (HMM) for the diagnosis and prognosis in machine process. Support Vector Data Description (SVDD) was later used by (Benkedjouh et al., 2012) for bearing fault prognostics.

This work provides an approach for modeling the degradation feature of bearings thereby giving useful information on the RUL of the bearings by using the eXtended Takagi-Sugeno fuzzy with recursive least square algorithm (RLSA) for prognostics purpose. The exTSFRLSA is a method that applies the principle of evolving fuzzy and Paris law to captures crack growth and is applicable to online data capture (hence its importance to this work) as less data is required to do its analysis for the degradation phenomenon of the corresponding RUL of the system.

The exTSFRLSA as used in this work is used for tuning, adjusting, and adapting the parameters involved in the propagation model by comparing predicted and measured defect sizes as in (Li et al., 1999; Angelov and Zhou, 2006), hence the instantaneous rate of defect propagation can be captured despite defect growth behavior variation and also to increase processing time of computation. SVDD which belongs to the statistical learning theory class is used here in this work to show the remaining useful life (RUL) of the bearing under study. An in-depth study of these methods is found in chapter 5.

1.2.6.1 Case-based predictive system.

A case-based predictive system is a system that is made to learn from the history of the machine condition to predict future faults arising. This sort of system could be very effective but will often require precise history of the machine state when in good condition and will need to update itself on the life cycle of the machine as it comes in use. They are utility systems that often require large data base capacity and could sometimes make wrong decisions due to the non-linearity inherent in most machine working conditions.

When using any Neural Network (NN) algorithm on a set of data to predict the remaining useful life (RUL) of the bearing or system, it is important to do data preprocessing (i.e. check if the set of data is "reasonable") as the accuracy of the model is improved by detecting trends and outliers.

The International Standard Organization state that, failure prognostics is the "estimation of the time to failure and the risk for one or more existing and future failure modes". This implies that the field of prognostics is not only interested in predicting the effects of known failure modes on asset life, but also how these may initiate other failure modes (Sikorska, Hodkiewicz and Ma, 2011).

The implementation of prognostics system results in an optimal maintenance schedule and it is the art or act of predicting future conditions based on present signs and symptoms. Bearing prognostics aim at anticipating the failure date by predicting the future health state of a bearing system and its RUL (Jammu and Kankar, 2011).

The approach of prognostics utilizing statistical survival models to analyze previous failure histories (failure data), thereby incorporating both event data and recorded CM data is called statistical failure data analysis. Useful decision-making information for the development of asset care plans (ACP) is achieved by turning vast amounts of data collected from industry through failure data analysis. Information such as asset time-to-failure and the type of failure is included in event data. The duration that an asset was operational up to the occurrence of a failure is known as time-to-failure. An indicator variable could arise because of the type of failure included in event data. Suspensions and failures are distinguished by their indicator variables. Suspensions happen when an asset is taken out of service before failure occurs. Failures in turn can be categorized as a maintenance action taken to influence the asset survival time or a breakdown, when a predefined failure condition is reached. For a predefined failure condition, a vibration level can be an example. Suspensions and failures are referred to as events (Guerlain, Brown and Mastrangelo, 2000)**.**

The deterioration process of a piece of production equipment can be viewed as a two-stage process, with the first stage being the normal operation stage and the second the potential failure stage. The area of interest in prognosis is when the second stage starts and how it develops (Wang, 2007). With prognostics, significant reduction in expensive downtime spares inventory, maintenance labour cost and hazardous conditions can be achieved. However, prognostics is a new area of research which is yet to receive prominence when compared to other areas of CBM. A problem of pattern recognition can be fault diagnosis. Where failure prognostic aims at assessing the current health condition of a machine thereby predicting the future to an estimation of its RUL (Heng et al. 2009; Zhang et al. 2010; Benkedjouh et al. 2012).

(Derouiche et al., 2012) applied neural networks for monitoring mechanical defects of rotating machines and the novelty in their work is in the using of generated previous values from data generated to predict into the future the likely trend to be obtained. To go about this, they explored various classical regression models with NN models and compared their predictive capabilities.

1.2.7 Introduction to model description.

It is common in systems, to consider features like inputs, system states and control actions including the system output. Artificial intelligence is a very useful tool in systems that are nonlinear (Mahamad, 2010). This suggests that intelligent control systems could be a good alternative to model- based controller or for systems with combined traditional approaches when creating hybrid control system.

(Guerlain, Brown and Mastrangelo, 2000) shed some light on algorithms that constitute important areas of engineering endeavour, like system engineering which includes data mining, data fusion and decision analysis. They strongly supported the notion that, decision support system (DSS) reasoning should be intelligent from a system perspective.

The neural network (NN) is a powerful building tool for a wide class of complex nonlinear system especially in the control strategy application. Two classes of control application are in the open loop identification and closed loop feedback control (Chin-Min, et al. 2011). Chin-Min, et al. 2011 made it clear that the basic problem in neural network closed loop feedback control is the providing of an on-line learning algorithm that does not require preliminary off-line training.

The Radial Base Function (RBF) helps in solving this problem as it helps to approximate the system dynamics and the adaptive laws which are derived to on-line, and which tunes the parameters of the neural network so as to achieve favorable estimated performance (Chin-Min, et al. 2011).

A major problem in approximating the system dynamics is to be able to make the approximation error between the neural network approximation and the unknown dynamic function for stabilizing the closed loop system

(Derouiche et al., 2012) also applied neural networks for monitoring mechanical defects of rotating machines and the novelty in their work is in using generated previous values from the data generated to predict into the future the likely trend to be obtained. They tried various classical regression models with NN models and compared their predictive capabilities. They were also of the opinion that among other parameters, the amplitude of vibration may be able to reveal progressive deterioration of a machine and that the rate of increase is proportional to the degree of damage hence the RMS value was used as the indicator to test the various model.

1.2.7.1 Model-based description.

In recent literature on computational intelligence, especially in the field of machine learning, the area of empirical modeling which encompasses the known field of data driven modeling (DDM) has greatly expanded its capabilities. Data driven modeling is based on analyzing the data of a given system thereby sourcing for links between the system state variables of the input and output variables without the knowledge of the physical behaviour. Where other models may suffer from being not robust enough, data driven models are often more accurate because they are based on objective information (i.e. the data) from the system being considered.

In condition-based maintenance systems many types of analyses as well as models have been employed for its diagnostic or prediction of machine health, using the model based diagnostic system approach, but the most prevalent is vibration analysis. Model based diagnostic system approach often encompasses data collection, data analysis and then structuring of data collected from which decisions are taken to either detect or predict the action of the situation of the system.

The data collection involves the use of transducers in collecting data from the monitored system and often they are required to be mounted in areas where true data reflecting the condition of the system under diagnosis can be collected and this requires skill as the sensors are often very sensitive due to their structure and design. So, placement of the transducer in the right position often requires skill.

The analysis of the data is very crucial with computer application as the computer will only perform the task given it to do, hence the choice of machine learning in its best practice are often chosen to affect the best decision for the system.

A special approach in data analysis with the model base diagnostic system is the use of Artificial Intelligence (AI) which is based on the biological learning process of the human brain, leading to the formation of neurons combined in a network which could mimic humans and do what the conventional computers do poorly. They have the capabilities of recognizing pattern via classifying and clustering (Widodo, et al., 2009).

1.2.7.2 Model-based diagnostics system.

This is another intelligent system that is highly more efficient for fault diagnosis and prediction. They are made to often comprise of different mathematical, neural network and logical methods to aid in diagnostic reasoning. With a model based diagnostic system fault diagnosis is made from the comparison of the real monitored condition of the system and the model design system.

Data driven approach for condition monitoring can broadly be used for diagnostic and prognostics purpose. The data driven diagnostic methods as well as the data driven prognostics method both comprise the artificial intelligence, statistical and other approaches. Model based diagnostic system whose acronym is MBD infer to an area of artificial intelligence.

Typical approaches that apply in fault diagnosis/prognostics are based on the theory methods and these approaches are grouped into two main categories, which are the model based approaches and the data-based approaches.

The model-based approach is often based on mathematical models (sets of algebraic or differential equations) to represent the system behavior, including the degradation phenomenon. These mathematical models constitute statistical and physical models. Statistical models are developed from the collected input/output data, which do not include not recorded conditions, while physical models are useful in accounting for all operating situations. Similar to model based techniques is the approach that operates directly on signal data without the use of mathematical models which is the data driven which uses signal processing techniques to expose patterns/signatures in signal data that give insight into machine condition (Jardine, Lin and Banjevic, 2006).

1.2.7.3 Model-based prognosis system.

To estimate the remaining useful life (RUL) of an asset, a physical understanding of the system is needed which is usually referred to as the model or physics-based approach. Model-based methods to prognostics imply precise expertise and theory relevant to the monitored machine. In industrial applications physics/model-based approaches are usually not the most important solutions as the defect type is often peculiar from one asset to another and is difficult to be identified without meddling with the operations. Hence the physical rules within the system should be well known in detail even though samples of failure degradation are not essential in physics-based prognostics. Utilizing residuals that represent the dispersion of sensed measurements from their expected values of healthy systems is the first phase in physics-based prognostics (Luo, and Pattipati, Kawamoto, 2003; Aye, 2014) while the second phase requires mathematical modeling of failure degradation.

The advantage of data-driven methods over the physics-based prognostic approaches and model-based prognostic approaches is that in industry applications, obtaining dependable data (data-driven) is easier than building physical models (model-based), while on the other hand, the generated behavioral models (data-driven) from real condition data leads to more accurate predictive results than those obtained from measured data (experience/physics-based) (Jammu and Kankar, 2011).

1.3 Research objectives.

Varying load and speed effects usually result in conditions of smearing in the frequency content and amplitude modulation which becomes most detrimental to the system, thereby complicating the diagnosis of a fault in the system. As reviewed in the literature there are three approaches to processing machine vibration signals namely: time domain analysis, the frequency analysis and the time-frequency analysis (Tandon and Choudhury, 1999). A combination of these analysis techniques could be possible.

Since traditional frequency analysis can only be applied if the signals are periodic, this is not appropriate for varying loads and speeds that produce signals that are highly non-periodic. Time-frequency analysis is still limited in practice with regard to this work owing to the difficulty of quantitative interpretation inherent with their use. Rotary machine diagnostics in the presence of varying operating conditions remains a challenge.

The time domain technique includes the use of statistical measures like kurtosis, skewness, means, standard deviation, etc. On their own they are poor indicators and cannot diagnose fault effectively (Wang and Chen, 2009).

The higher order statistics (HOS) have been difficult to use due to the problem of outliers that tend to affect its signal analysis when used and hence, has less commonly used in the past, but they hold promise when well applied. Other factors influencing vibrations, in the case of rotating elements, are varying load and speed. As explained in Chapter 2, a variation in these factors produces some difficulties in recognizing the presence of fault in a signal. HOSs were used in this work to cluster faults together while separating them into categories.

While skewness and kurtosis have been used extensively in condition monitoring of bearings and gears, higher order statistical (HOS) techniques have not found wide application in machine conditioning monitoring. This is because if a process is Gaussian then HOS provide no information that can be obtained from the second or higher order statistics (Niu, et al., 2005). However, the Kullback-Leibner divergence theory was used in this work to overcome the problem of outlier which would have been a problem to the use of HOS, much as been said about it in the literature study.

In this work, the use of HOS is investigated concentrated and worked upon for use as fault indices under the varying load and speed conditions in chapter 2 and some non-parametric statistical methods was used for fault diagnostics and prognostics as shown below in other chapters.

The BRNHMM is a classical statistical and a machine learning model and performs when the features are independent of each other. The SVMGA is also a classical statistical machine learning model and they both have the natural ability to reason under uncertainty and both often handles missing data better. Each model can be linked with each other in a particular manner which is characterized by similarities and particularities as they are formalized based on machine learning methods.

Each classification method uses a decision function f whose parameters are determined in the training stage and then used for classification purposes. They both need no user interaction and can adapt many kernel parameters to given data without having to sacrifice training cases for validation. The BRNHMM theorem performs when the features are independent to each other while SVMGA performs using the radial basis function kernel and both are more likely to perform better as they can handle non-linearities in the dataset.

Support Vector Machines (SVMs) are a set of related supervised learning methods used for classification and regression that belong to a family of generalized linear classifiers. However, in this work it was adopted for fault classification of non-linear classifiers as the features generated which helps to foster it were optimized using the genetic algorithm. The relevance of support vector machines for regression is that it is a Bayesian form representing a generalized non-linear model of identical functional form of the SVM. Bayesian techniques also provide a general rigorous framework for dynamic state estimation problems.

A very nice connection can be established between SVMGA and BRNHMM formalisms. This is since probability measures can be attached to the SVs, thus allowing posterior probability measure as the output of the classification task. Moreover, the classification task is done by solving a function which is regularized. This choice of the regularization parameter and the kernel type can be done via the Bayesian perspective. One important difference is that while Bayesian is using all the training data to infer the model, the SVMGA is using only the determined SVs for the same purpose. Mathematical relation expressing these two-models can be found in (Costache, Liénou and Datcu, 2006).

The Bayesian robust new hidden Markov model (BRNHMM) with support vector machine for regression and genetic algorithm (for diagnostics) are validated in chapter 3 and chapter 4 respectively, while the eXtended Takagi-Sugeno fuzzy recursive least square algorithm was used for determining the RUL of the REB in chapter 5.

Both the BRNHMM and the exTSFRLSA are basically two schools of thought to the probability theory which are the frequency and the Bayesian method are. To Bayesian method, the idea of an event having an intrinsic probability is ridiculous. Fuzzy uses much the same tools as probability theory. The two are very closely related as the basic connectives in fuzzy are defined in almost the same way as the corresponding operations in probability theory.

Fuzzy logic is a hacky but computationally efficient way to approximate probabilities reasoning. Many fuzzy algorithms are generalization of probability-based algorithms with a specific membership function, but fuzzy systems have the potential for much more flexibility than probabilities models.

Probability also provides a more analog representation of a concept as it allows for there to be a chance of a result being one way or the other without certainty of which it is, but at the end of it, it still requires that one of those state happens. Fuzzy membership allows for an object to be partially of one type (class) and practically of another type.

In a nut shell, the probability theory does not capture the essential properties of meaning (partial truth) which is the goal of fuzzy and fuzzy does not capture the essential property of meaning (partial knowledge) which is the goal of probability theory.

The objectives of this thesis align with the development of a diagnostic framework that:

- extracts diagnostic information from machine response signals at slow speed, by using AE signals to detect damage in low speed rotating bearings, thereby offering high sensitivity.
- extracts diagnostic information in such a manner that it is sensitive to the presence of machine faults, yet also robust in time-varying operating conditions,
- establish and pinpoint a proposed and robust acoustic emission index which makes clustering of faults easy and for easy decision of whether the samples obtained from faulty bearings come from the same distribution
- is not dependent on fault historic data, destructive tests, or extensive manual preparation of training data,
- allows for a simple intuitive representation of the extracted information, such that a non-expert, or a simple classification algorithm may interpret it.

Few researchers have proposed robust techniques for fault diagnosis of rolling element bearing (REB) of rotating machines operated under variable speed and load conditions. The approach in which this problem will be tackled in this work is given in a schematic diagram as depicted in Figure 2.1 in the next page.

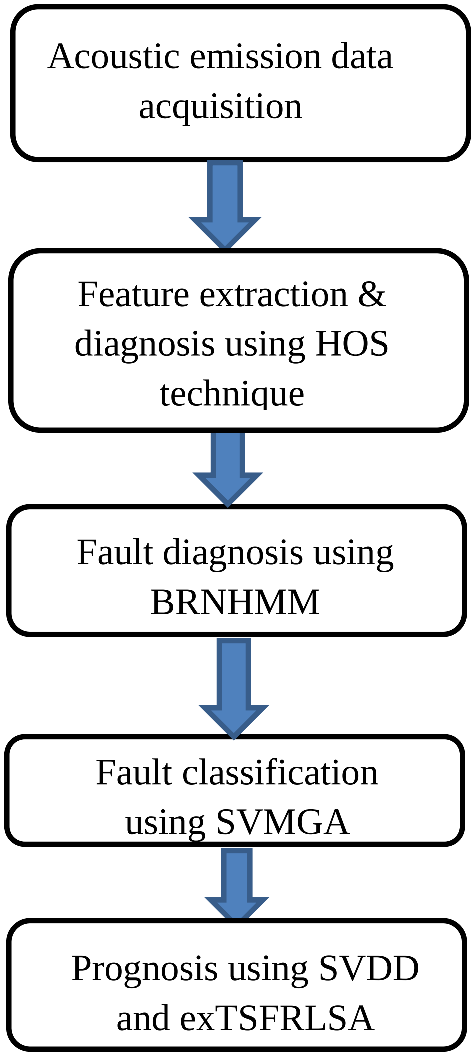

Figure 2.1, Schematic process of work done.

The exact goal that each of the algorithms meet is highlighted in the summary section in each section where they are found, but a brief prelude to each section summary is given below.

The first major decision made in this research work was to effectively decide between the use of acoustic emission or vibration signals for the diagnostics of faults under low speed variable operating conditions. Acoustic emission was chosen as it has been proven useful for fault diagnostic in low speed rolling element bearings. AE has demonstrated its viability to detect damage in low rotating bearing thereby offering high sensitivity (Mba, 2003; Moreno-Munoz et al., 2007).

The second objective of this research work is to propose robust acoustic emission indices which simplify the clustering of faults and allow for better decisions on whether the samples obtained from faulty bearings come from the same distribution. HOS like hyper-flatness and hyper-skewness are effectively combined with Kullback-Leibler (KL) divergence and Lempel-Ziv complexity to formulate an indicator (energy coherent factor) which made use also of kurtosis (in its formulation). The novelty with this work is that outliers are taking care of while faults are successfully clustered into categories. This algorithm is derived and validated in chapter 2 with a prelude of this subject claim given in paragraph 2.4 and 2.5 of this section.

The Bayesian Robust New Hidden Markov Model (BRNHMM) was successfully applied for the diagnostic of bearing operating under varying load and speed conditions. This is a semi-non parametric statistical model that effectively make use of hidden Markov model (HMM) to overcome the disadvantage stemming from the use of large data sets (Soualhi et al., 2012). Kullback-Leiber (KL) divergence was effectively used to obtaining a faster iteration process with no pre-processing, making it ideal for online fault detection and for easy fault classification during its incipient stage. More on this algorithm and its validation is given in chapter 3.

The fourth objective is the use of AE for high sensitive data analysis at low speed and varying load conditions. This objective help investigate the establishing of a practical diagnostic model, whereby, online AE signal mixed with noise as acquired in condition monitoring (CM) is transformed into a set of features which establishes the machine health condition for fault diagnosis (Niu, et al. 2005). Features of detection, selection, extraction and classification of data under varying load and speed conditions using models like the support vector machines (SVMs) for regression and genetic algorithms (GA) which belongs to the statistical learning theory was focused on here and genetic algorithm was successfully use to optimize the best feature for its classification (Smola and Schollkopf, 2003; Xiangyang and Wanqiang, 2014) more of this is given in paragraph 6 of this section.

In this thesis the use of HOS with the combination of Kullback-Leibner divergence and Lempel-Ziv complexity was effectively used to formulate the energy coherent factor in chapter 2, and then moved on to chapter 3 to use the same Kullback-Leibner divergence and Bayesian method to obtain the variational free energy lower bound of the model for REB at low speed and under varying load condition. Also, fault classification was obtained at this level, using statistical non-parametric method.

In chapter 4, another statistical non-parametric methods and algorithms (the SVMGA) was used for solving the pressing problems for diagnostics while support vector data descriptive (SVDD) was use for the prognostics of rolling bearing element operating at varying speed and load conditions.

In chapter 5 the prognostics and estimation of the remaining useful life in rolling bearing element was discussed. Here the eXtended Takagi-Sugeno with fuzzy recursive least square method was used to prove the usefulness of prognostics of REB.

This research attempts to further identify faults in the incipient stage on rolling element bearings running at low speed and under varying load conditions. The algorithms derived in this thesis can be augmented to include other aspects of artificial intelligent, other statistical learning theories etc. Some of these aspects have already being discussed with their drawbacks given which can be worked upon.

1.4 Document layout.

Chapter 1 gives the problem statement, literature survey and the research objectives. Chapter 2 describes the data driven pre-processing and diagnosis of rolling element bearing using higher order statistical method. The pre-processing considers how data obtained for fault diagnosis in slow bearing can first be normalized using the Kullback-Leibler divergence estimation which is known to be useful and good of continuous distribution and to enhance the signal to noise (S/N) ratio for statistical indicators to be able to diagnose incipient fault in rolling element bearing using acoustic emission.

Chapter 3 covers the diagnosis of REB using the Bayesian Robust New Hidden Markov Model as the BRNHMM has proven useful and reliable in detecting faults that are in their early stage where the signal strength seems to be week in amplitude (as in the case of bearing rotating at low speed) and buried in noise to which acoustic emission method proves important.

In chapter 4, fault classification and recognition in rolling bearing element using artificial intelligent such as support vector machines for regression and genetic algorithms (SVMGA) was covered as this method showed effective classification of faults under the varying load and speed condition which other classification proved difficult to classify as the fault under view is multiple in nature.

Chapter 5 involves the prognostics and estimation of the remaining useful life in rolling bearing element was discussed. Here the eXtended Takagi-Sugeno with fuzzy recursive least square method was used to prove the usefulness of prognostics of REB. In chapter 6 is the conclusion, general summary and a discussion of the contribution of the entire research work.

2.0 Chapter Two Data pre-processing and diagnosis of roller element bearing REB.

2.1 Diagnostics of slow rotating bearings using HOS and acoustic emission.

Having discussed diagnostics of slow rotating bearings already in sub-section 1.2.5 emphasis will be laid here on the related diagnostics signal best suited for its slow rotating speed. Acoustic emission (AE) is most suited for this condition and is a high frequency phenomenon whereby transient elastic waves are generated by rapid (and spontaneous) release of energy from a localized source or sources within a material (Niu, 2005; Welling, 2005). AE–based condition monitoring in rolling element bearings is not new and it has been studied widely for condition monitoring at higher speeds (> 600 rpm). However not very much has been done at low speeds between 10 – 600 rpm. Low speeds such as these are also often associated with widely varying operating conditions which makes spectral based analysis techniques less appropriate (Caesarendra et al., 2016).

The general motivation behind the use of HOS is that they extract information due to deviations from Guassianity. They also detect and characterize the non-linear properties of mechanisms which generate time series via relations of their harmonic components. Much has already been said on the first point of motivation in paragraph 1 of sub-section 1.2.5. The second point is based on the valid assumption that higher order statistics play a key role in detecting and characterizing the type of non-linearity in the system from its output data.

Statistical moments tend to describe the shape of the amplitude distribution of vibration data collected from a bearing and are sensitive to the impact impulses. From previous research on the application of statistical moments to condition monitoring in rolling element bearings, it is known that the third and fourth normalized central moments restrain the selective range of statistical parameters (Niu, 2005).

The presence of outliers which result from noise generated by moving parts within and around the machine renders the estimated moments unstable and this could become severe for higher order statistical moments like skewness, kurtosis, hyperskewness and hyperflatness as found in section 1.2.3.3. Higher orders expose one to the effects of outliers (measurement errors) thereby causing significant variance and which no longer captures the correlation with damage.

(Caesarendra et al., 2016) recently proposed a feature extraction method of the AE time domain waveform signal using the largest Lyapunov exponent (LLE) algorithm thereby demonstrating that the LLE feature can detect indications of failure from AE hit parameters such as the amplitude, RMS, counts and AE Burst, energy with the Average signal level (ASL) etc.(Zaeri et al., 2011; Ulus S. and Erkaya S, 2016). However, the process is too cumbersome and unfit for online fault detection.

The use of Kullback-Leibler (KL) divergence and Lempel-Ziv complexity may offer a solution to this issue; since by introducing KL divergence one reduces the sensitivity to outliers but essentially retain the advantage of higher emphasis on irregularity (Fernando, 2008). The Lempel-Ziv complexity is an analysis tool used for non-linear dynamic systems. It, specifically measures the generational rate of new patterns along a digital sequence and is closely related to important source properties such as entropy and compression ratio (Amigó et al., 2004; Yan and Gao, 2004; Hong and Liang, 2009). KL divergence is defined as the mean of the log-likelihood ratio which is the exponent in large deviation theory. It is also known as information divergence and relative entropy that measures the distance between two density distributions.

The performance of these techniques improves under inchoate and early-stage bearing fault detection due to the following

1) Reliability under poor signal-to-noise ratio (SNR): As an efficient system it detects and diagnose fault reliably under initial conditions.

2) Immunity to noise: In harsh industrial conditions, SNR may vary drastically over the time in industrial environment.

The main contribution of this chapter is proposing a new indicator called the energy coherent factor for detecting anomalies in faulty bearings using Kullback-Leibler (KL) divergence and HOS elements as it utilizes the AE signal in detecting the state of health of bearings. The new indicator responds to the amplitude of the energy of the emission.

2.2 Background to HOS model.

2.2.1 Skewness (3rd order statistical moment).

Skewness characterizes the degree of asymmetry of distribution around its mean and a measure of the lopsidedness of the distribution. A distribution that is skewed to the left (i.e. the tail of the distribution is longer on the left) will have a negative skewness while a distribution that is skewed to the right will have a positive skewness. Hence it is reasonable that measurements must be conducted over a sufficiently long period to encompass at least one complete cycle of the modulation to be certain to have measured the maximum skewness.

$$Skewness = \frac{\sum_{i=1}^{N}(x_i - \mu)^3}{N\delta^3} \tag{2.1}$$

To extract the feature parameter from a fault signal contaminated by noise and to accurately identify the fault type, a skewness wave ($SKW_{skewness}$) of the faulty damage bearing is taken as it is needed in the computation to arrive at the Kullback-Liebler information wave (KLW) which is presented later (Niu, Zhu, and Ding, 2005).

$$SKW_{skewness} = \frac{\sum_{i=(j-1)\times M+1}^{j\times M}\left(x_i - \mu_j\right)^3}{M\delta_j^3} \tag{2.2}$$

where M is the number of data points in a small region, and μ_j and δ_j respectively are the mean and standard deviation of the signal series x_i in the j^{th}small region. The skewness formula of (2.2) will be used to first generate a skewness wave to the fault-based signal from the bearing housing. Table 2.1 shows the statistical moments of higher order relation giving its moment number in increasing order.

Table 2.1, Statistical moments of higher order relation.

Moment number	Central Moment	Standardized Moment
1	0	0
2	Variance	1
3	-	Skewness
4	-	Kurtosis (or flatness)
5	-	Hyperskewness
6	-	Hyperflatness

2.2.2 Kurtosis (4th order statistical moment).

Kurtosis as a fitness parameter also offers the advantage of having high values in the presence of the fault signal where it is usually zero when only background noise is present. The kurtosis spectral wave ($KW_{kurtosis}$) was generated on the fault diagnostics signal (i.e the signal generated from the faulty bearing housing) (Petropulu, 2000; Niu X., 2005).

$$KW_{kurtosis}(j) = \frac{\sum_{i=(j-1)\times M+1}^{j\times M}(x_i - \mu_j)^4}{M\delta_j^4} \quad (2.3)$$

where μ_j and δ_j are, as above the mean and standard deviation of the signal series x_i in the j^{th} small region. M is the number of data points in a small region and $j = 1 - L,\ L = {}^{M}/_{N} \leq {}^{f_s}/_{f_A}$ where f_s and f_A are the sampling and the analysis frequencies respectively. The signal data $x_i(i = 1 - N)$ is divided into smaller regions. The points of the kurtosis are connected in order to derive the kurtosis wave $KW(j)$ (Wang and Chen, 2009).

2.2.3 Hyperflatness (6th order statistical moment).

Hyperflatness is a statistical power of the sixth order. It is a powerful means of estimating fault diagnosis in bearing signal since it belongs to the kurtosis family.

$$(HFIW) = \frac{\sum_{i=(j-1)\times M+1}^{j\times M}\left(x_i - \mu_j\right)^6}{N\delta^6} \tag{2.4}$$

2.2.4 Feature extraction using KL and Lempel-Ziv complexity.

Kullback-Leibler (KL) divergence is used to compare the current estimate of the feature variable of $P_1(x)$ to the reference feature variable $P_2(x)$ as expressed in (2.5). It is defined as

$$KL(P_1P_2) = \int P_1(x)log\frac{P_1(x)}{P_2(x)}dx \tag{2.5}$$

where $P_1$and P_2 are two probability distributions which have probability density $P_1(x)$ and $P_2(x)$ respectively and $KL(P_1P_2)$ can be used to compare the two PDFs.

There are two fundamental properties of KL:

- Non-negativity: $KL(P_1P_2) \geq 0$ with equality if and only if $P_1 = P_2$
- Asymmetry: $KL(P_1P_2) \neq KL(P_2P_1)$ (Dube, Dhamande and Kulkarni, 2013).

Kullback-Leibler divergence (Fernando, 2008) captures the expected log-likelihood ratio which gives a statistical interpretation of power loss, when the wrong distribution is used for one of the hypotheses. Moreover, technically it is a vital part of probability theory with a deep connection to large deviations theory and to statistical inference to ergodic theory i.e. outliers associated to data measurement. This could be important in the analysis if the damage detection is compensated so that in the use of HOS, these deviations which could have irregular results in the analysis are well taking care of. KL divergence has interesting statistical properties; finding parameters of a statistical model by maximizing the likelihood is analogous to finding the parameters minimizing the divergence (Williams, 2001; Bondu and Grossin, 2010).

The KL information quantity may be calculated from the expected value of the reference feature variables given by

$$KL = SKW_{skewness}(t)log\frac{SKW_{skewness}(t)}{HFIW(t)}dt \tag{2.6}$$

Hence to extract the Kullback feature of the fault signals, the "Kullback-Leibler Wave (KLW)" is expressed based on the KL information quantity expressed above in (2.6)

Therefore,

$$KLW(t) = SKW_{skewness}(t) log \frac{SKW_{skewness}(t)}{HFIW(t)} \quad (2.7)$$

It is on the KLW(t) signal that the spectral analysis will be performed (Fernando, 2008). Getting the envelope wave from the absolute values of the Kullback-Leibler information Wave (KLW) which is presented below as

$$|KLW(j)| = \left| SKW_{skewness}(j) log \frac{SKW_{skewness}(j)}{HFIW(j)} \right| \quad (2.8)$$

Kullback-Leibler (KL) divergence was used together with the hyperflatness wave to obtain the final reduced signal on which the Lempel-Ziv complexity was used for fault diagnosis. To be able to detect a defect (change) in the vibrating signal, there is the need to be able to interpret the contribution of each variable to the divergence and then this contribution is normalized. This contribution is based on the idea of (Moreno-Munoz et al., 2007) and can be evaluated in (2.9).

$$\text{Contrib\_L} = \frac{KW_{kurtosis} - SKW_{skewness}}{HFIW} \quad (2.9)$$

2.2.5 Complexity measurement.

The Lempel-Ziv complexity is an alternative tool for signal analysis involving nonlinear dynamics and has been used for monitoring the effect of anesthesia in patients (Yan and Gao, 2004). It is used in transforming a signal to be analyzed into a data sequence whose elements are given in symbols of which such transformation is often referred to as a “coarse-graining” operation. Complexity analysis is known to help in focusing on the intrinsic characteristics of the overall dynamics of a signal and neglecting details contained in lower hierarchical components, especially as it relates to an AE signal whose analysis cut across a narrow broad band spectrum where the impact fault occurs in the damage bearing.

The Lempel-Ziv complexity reflects on the number of all different subsequences contained in the original signal data sequence (Amigó et al., 2004; Hong and Liang, 2009). Considering a generalized normalized complexity, where A* denotes the length of all finite length sequences over the finite symbol set A, and $l(P)$ denotes the length of a sequence $P \in A^*$ with (2.10).

$$A^n = \{P \in A^* | l(P)| = n\}, \qquad n \geq 0 \tag{2.10}$$

For every $P \in A^n$, the Lempel-Ziv complexity can be expressed as

$$C(n) < \frac{n}{(1-\varepsilon_n) log_\alpha(n)} \tag{2.11}$$

where $\varepsilon_n \to 0$ if $n \to \infty$, and α is the number of different symbols in the symbol set A.

2.2.6 Proposed indicators.

The reason for proposing a damage indicator is based on the contribution in the formulation from the equation (2.9) which is based on the formulation by (Moreno-Munoz, et al., 2007). Here it is further expatiated upon and a complexity measurement is introduced. The indicator is computed from the sixth statistical power, the Kullback-Leibler Wave (KLW), kurtosis wave ($KW_{kurtosis}$), skewness wave ($SKW_{skewness}$) and the skewness series. The motivation for proposing this indicator as a damage analyzing tool is that it tends to describe the shape of the amplitude distribution of the acoustic signal by trying to explain its sensitivity at low speeds, as indicators from the past is needed to determine the damage in a bearing. Which from the literature has some drawback such as being poor indicators when dealing with very low speeds in bearings and hence need to be improved upon. The following describes the calculation procedure:

1. Compute the Lempel-Ziv complexity C(n) of the original Kullback-Leibler wave C(KLW), of the sixth statistical moment (hyperflatness C(HFIW)), (and complexity of the series skewness, i.e. C(skewness)).
2. Compute the original information wave of the kurtosis ($KW_{kurtosis}$)and of the skewness ($SKW_{skewness}$).
3. Finally, in (2.12) we have the indicator formulation as;

$$Energy\ coherent\ factor = \frac{\left[\frac{|KW_{kurtosis} - SKW_{skewness}|}{C(HFIW)}\right]}{\left[\frac{|C(HFIW) - SKW_{skewness}|}{C(KLW)}\right]} \tag{2.12}$$

This formulation is linked to equation 2.9 where we see the relation between kurtosis wave, skewness wave and the hyper-flatness information wave being laid.

2.3 Experimental setup.

The test rig shown in Figure 2.1 was used to simulate the progression of damage in a bearing. Data acquired from this set-up were subsequently used to extract the features required for the computation of damage indicators based on HOS. The progression of the damage was traced with an acoustic emission signal from an acoustic emission transducer. Only the radial response was monitored in this case.

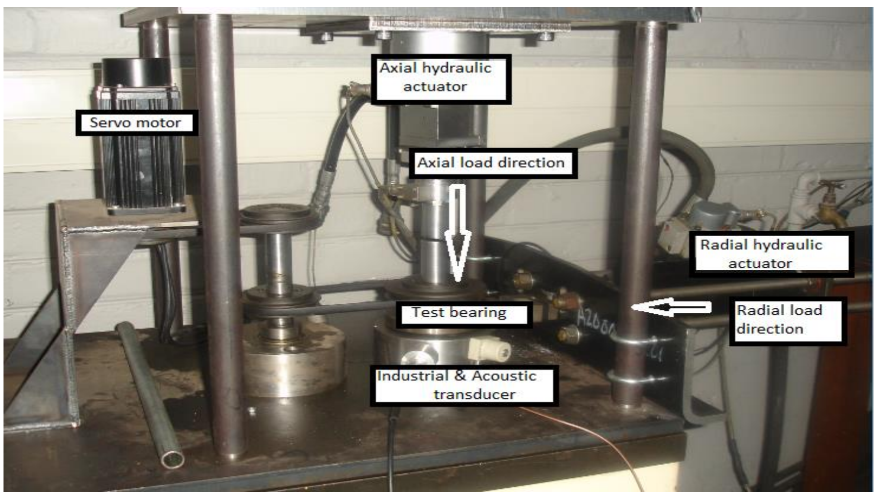

Figure 2.1, Experimental test rig.

The test rig depicted in Figure 2.1 (above) features the use of two servo-hydraulic actuators to introduce constant amplitude axial and radial loads on a test bearing. The purpose of introducing the two actuators is to allow simulating a scenario where coupling between axial and radial loads could be considered. The rotating speed for the slow rotating bearing ranges from 70 to 100 rpm. The actuator forces applied is however effectively sinusoidal. Figure 2.2 (below) shows the schematic diagram of the built test rig.

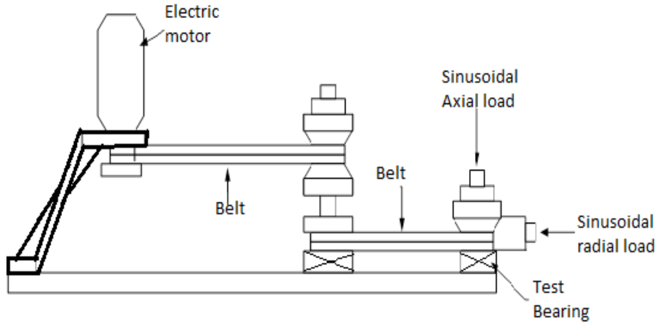

Figure 2.2, Schematic diagram of test rig setup.

Three test bearing scenarios were considered. Firstly, an undamaged bearing was considered. In a second bearing grounded metallic debris was mixed with grease and introduced into the bearing. In the last bearing a simulated crack was introduced (see fig. 2.3)

A taper roller bearing (Timken HR 30307 J) was used to be able to artificially introduce the localized-defect, since it can be dismantled from the outer raceway. Surface damage was seeded on the outer raceway of the bearing (as shown in Figure 2.3) with the use of a small hand drilling machine to which a small disk was mounted, which was then used to introduce a groove on the outer raceway of the taper bearing.

Figure 2.3, Seeded damage on outer race of a bearing.

Details of the selected bearing are reported in Table 2.2 (below).

Table 2.2, Bearing specification.

Contents	Parameters
Bearing specification	Timken taper roller bearing HR 30307 J
Bearing outer diameter	80 mm
Bearing inner diameter	35 mm
Bearing width	22.75 mm
Bearing roller diameter	12 mm
The number of rollers	14
Rated speed with grease	4800 rpm

A brushless AC motor (Rockwell Automation MPL-B680B), mounted on a NSK 6309 single row bearing was used to drive the system. The angular velocity of the motor was retrieved from one of the analogue outputs available on the motor drive, (Rockwell Automation Kinetix 6000 series BM-01). This system allows continuous speed variation from 0 to 3600 rpm. A Soundwel AE sensor with model number SR 150M and a frequency range of 25-530 kHz was used in the experiment.

The first bearing (undamaged) was loaded sinusoidally with forces of amplitude of 300N at a frequency of 2 Hz on the axial load and an amplitude of 700 N at a frequency of 1Hz in the radial, while bearing two which was debris induced was loaded sinusoidally with forces of amplitude of 400 N at a frequency of 2 Hz in the axial direction and an amplitude of 800 N at a frequency of 1 Hz in the radial direction. Bearing three with crack at the outer race was loaded sinusoidally with forces of amplitude of 500 N on the axial at a frequency of 2 Hz and 900 N at a frequency of 1 Hz in the radial direction. The reason for applying different loads axially and radially is to simulate real life scenarios of how the application of load under varying conditions affects the bearing in its life cycle.

The speed of the servo motor was set at 70 rpm, 80 rpm, 90 rpm and 100 rpm and ran on bearings 1, 2, and 3 respectively. The vibration signatures for the three test bearings were collected for the four speeds, using an FFT analyzer, a National Instruments data acquisition card (BNC-2110) with a shielded BNC connector block.

Figure 2.4 (below) shows the flow chart for arriving at the indicator proposed for use in this work. First the loads are applied in both the axial and radial direction and the acoustic transducer is used to obtain the signatures from which the skewness, kurtosis and hyperflatness values are obtained, which are then used to find the KLW before being applied to the indicator function.

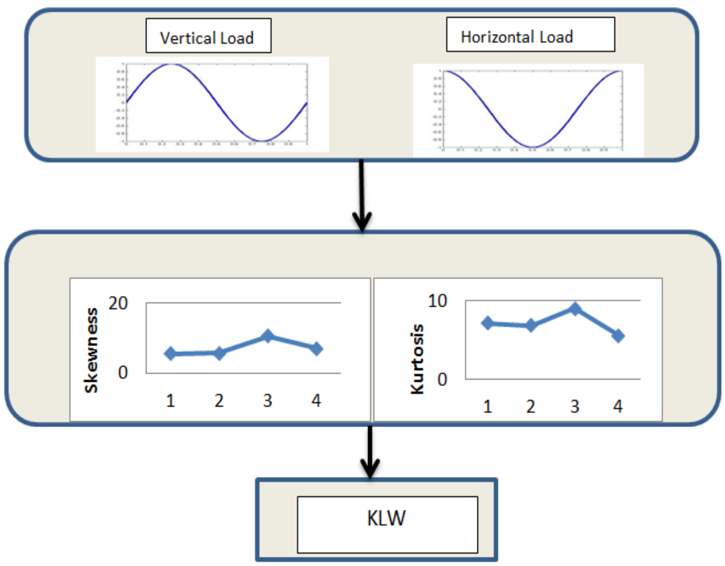

Figure 2.4, Diagram for arriving at the indicator values.

Sinusoidal axial and radial loads were applied to the test bearings and acoustic emission signals were collected from the bearing. These signals were grouped into smaller groups and analyzed further obtaining skewness and the kurtosis wave to which the Kullback-Leibler principle was applied to obtain the Kullback-Leibler wave (KLW) as is explained in section 2.2.4.

2.4 Parameter extraction based on statistical higher moments for KL formulation.

A bearing in a good condition has a Gaussian acceleration probability density distribution, whereas a damaged bearing results in a non-Gaussian distribution with dominant tails, because of the relative increase in the number of high acceleration level events. High-order statistics are used to infer new properties about the data of non-Gaussian processes to which the AE belongs (Moreno-Munoz *et al.*, 2007). An acoustic emission signal was considered in this work to test the computational robustness of the proposed indicators and to validate the indicator which help in the analysis to capture the impact within a narrow broad band spectrum, given that the impact response in the damage bearing is very short.

2.5 Experimental results.

Figure 2.5 (below) shows typical acoustic emission signals obtained from the bearing test rig. The length of the data record taken was 60000 data samples. It cannot easily be observed that the amplitude of the wave formed as shown in the figure varies within some given time interval.

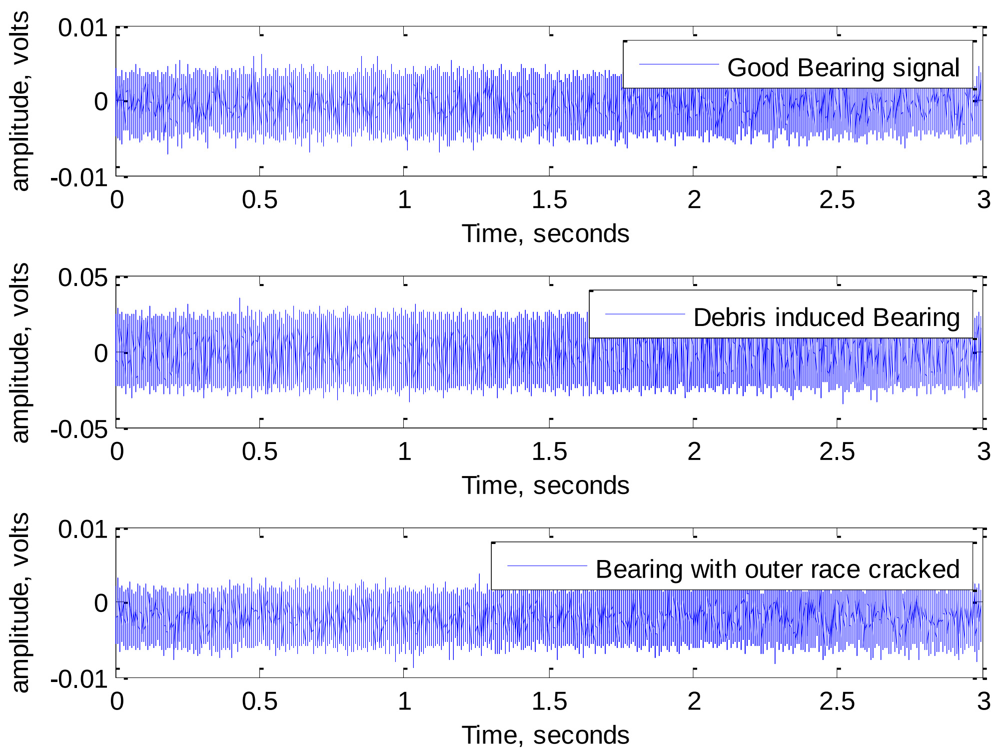

Figure 2.5, Acoustic emission obtained from test rig at 100rpm.

Frequencies of the order of 100 kHz are involved with AE and the fault characteristic frequencies caused by the defective bearing and its harmonics are difficult to detect in the corresponding spectrum by conventional FFT-based envelope analysis especially at low speeds and very low speeds (< 10 rpm), as it occurs within a narrow band spectrum their harmonics as depicted in fig. 2.6 below are also difficult to obtain especially for the outer race cracked bearings.

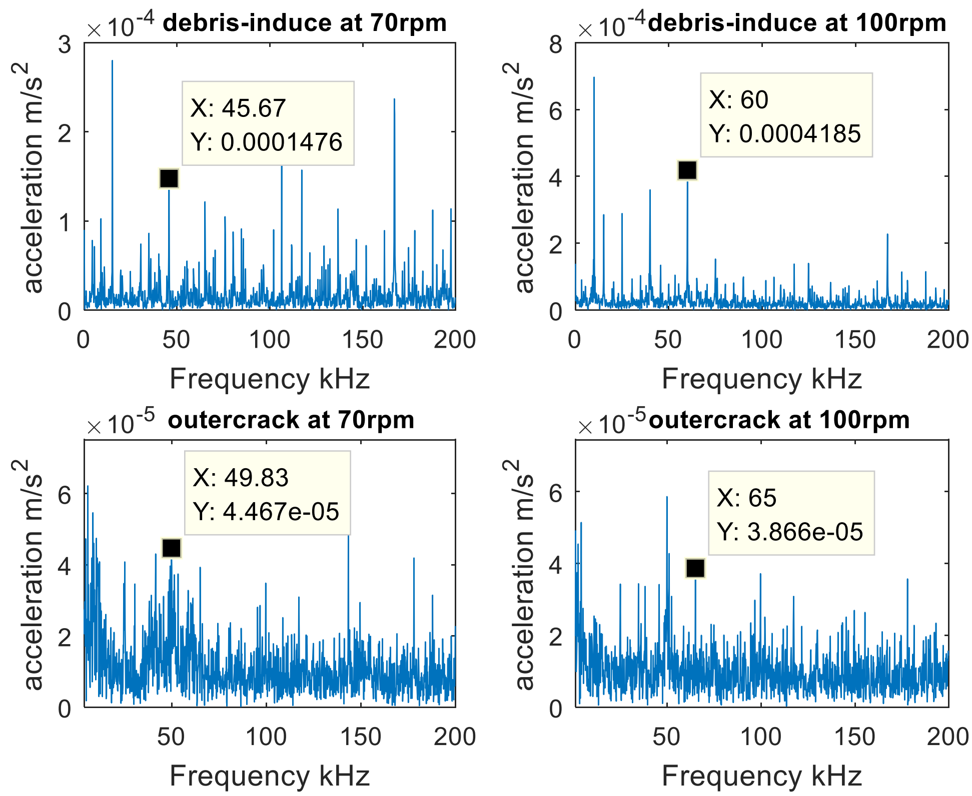

Figure 2.6, The FFT of the signal at 70 and 100 rpm.

Figure 2.7 below shows the evolution of the descriptor established for the acoustic signal of the specific defect for the indicator. Referring back to equation 2.12, the value obtained does consider the energy and the amplitude of the signals for the indicator. The rotational speeds used for the test are 70 rpm, 80 rpm, 90 rpm and 100 rpm. To be able to cluster at the different speeds, it is necessary to obtain an absolute value of the difference in the subgroup division of the indicator equation formulated, as sometimes negative values are obtained which could cause the signal at the various speed levels to cross each other when plotted causing clustering/separation of the signal to be difficult. Skewness could sometimes give a negative result hence the denominator should always be positive and smaller than the numerator to obtain favorable answers from the indicators. The high energy and amplitude exhibited by the ringing pulses generated as the roller passes over the groove induced on the outer ring cannot be observed on the time signal given in Figure 2.5. That is why the "energy coherent factor" is formulated to see if it does reflect this energy exhibited. This is shown in Figure 2.7.

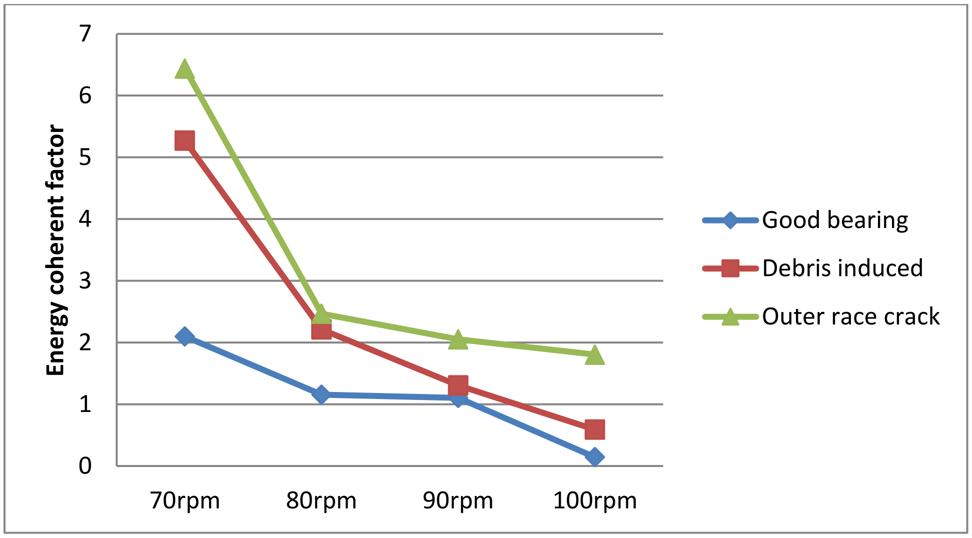

Figure 2.7, Energy coherent factor plot.

By observing figure 2.7, we find that at a low speed of 70 rpm the energy levels of the three scenarios are clearly displayed and as the speed increases (after a longer life cycle of use) there is a reduction in the energy level which further reduces as the speed and life cycle increases which explains the description provided in sub-section 1.2.3.3 on how the kurtosis value relating to that of a good bearing stays below 3.

The good bearing indicator line as shown in the plot above did not show much deviation from the zero line, while the debris induced, and the outer race crack bearing indicator lines show significant deflection away and from the zero line respectively.

2.6 Summary.

HOS seem like an intuitive thing to do. Examples are the use of skewness and kurtosis. However, the direct outliers (which originate from non-uniform load) that make it worse for signal analysis as a peakedness fault in the measured data become difficult to measure. These outliers which are being created in measured data being analyzed are common with higher order moment.

Lower order statistical moments such as skewness and kurtosis which are commonly used could not be easily used to identify the faults at incipient stage especially as it relates to very low speeds, but in this chapter the condition of bearings at very low speeds are successfully classified into clusters of good bearings, debris induced, and outer race cracked bearings. This is because HOS parameters are much more sensitive than lower order statistical moments. This however has the problem of outliers affecting the analysis. This is taken care of here by KL divergence and Lempel-Ziv complexity.

HOS is effectively combined with KL divergence and Lempel-Ziv complexity to formulate an indicator that makes use of kurtosis which is known to be a relatively poor insensitive of faults especially on the outer race of a bearing. With HOS, information is successfully extracted which deviate from Gaussianity, making it easier to detect and quantify non-linearities in time series especially at very low speed.

For this, it becomes necessary to introduce statistical techniques to “smooth” the signal, by introducing even higher order parameters like (HFIW in energy coherent factor), and this leads to results where the energy coherent factor is less sensitive to speed but separate well based on damage.

3.0 Chapter Three Diagnosis of REB under slow rotating condition.

3.1 Basic concept and algorithm of BRNHMM for REB diagnosis.

As continuation with the work done in chapter 2 where the energy coherent factor of the varying speed and load signal was formulated, there was a need to find the bounds of the energy level determined and as reported in section 1.2.5.2, HMM was combined with Bayesian method with the aid of the Kullback-Leibler divergence (which we saw in chapter 2) to achieve this goal. HMMs have proven to be very effective among the various stochastic approaches, in modeling both static and dynamic signals. It is a finite-state machine which changes its state during every time increment.

This chapter focuses on the introduction of the application of HMM in combination with Bayesian theory in a new robust method through the application of the Kullback-Leibler divergence which is a non-parametric statistical method in continuation to chapter 2 which also is a non-parametric statistical method. It also shows the intermediate results to provide a thorough step-by-step guideline of how the HMM is used for process monitoring.

HMM is usually approached in probabilistic manner. Probability theory is a branch of mathematics concerned with the analysis of random phenomena which gives us a consistent frame work for the manipulation of the uncertainty. Probability can be measured using a Bayesian approach which makes use of probabilities to quantify the degree of belief in different models.

Bayesian probability is measured as a probability distribution over a given parameter. It treats the entity or parameter of interest as a random variable making it possible to estimate the uncertainty associated with the estimation process using a single observed data set. This makes this approach much more flexible for most practical analyses where enough real-world data is not available (Li et al., 2005).

HMM models can be grouped into three categories namely discrete, continuous and semi-continuous models with the difference lying in their use of different output production probabilities. HMM is an extension of Markov chains. With HMMs every state do not match with an observable event, but is often connected to a group of probability distributions of state and the actual problem is often more complex than that described by Markov chains (Purushotham, Narayanan and Prasad, 2005).

The compact notation for convenience is given as $\lambda = (A, B, \pi)$. More generally it is given as $\lambda = (N, M, A, B, \pi)$ which denotes a discrete HMM i.e. discrete probability distributions. For a continuous HMM $\lambda = \left(A, C_{jm}, \mu_{jm}, \sum_{jm}, \pi\right)$ is used here for continuous density functions or distributions. Essentially there are three algorithms in HMM, namely the forward-backward procedure, the Viterbi algorithm and the Baum-Welch algorithm (Rabiner, 2009); (Ilhem, Amar and Lebaroud, 2014). These three basic algorithms represent the three basic problems to be solved respectively. The three problems are:

- **Evaluation**, given the observation sequence $O = o_1, o_2, \dots, o_T$, and a model $\lambda = (A, B, \pi)$ how should one effectively compute $P(O|\lambda)$ which is the probability of the observation sequence, given the model. The forward-backward procedure is used for solving this problem.
- **Decoding**, when given the observation sequence $O = o_1, o_2, \dots, o_T$, and the model λ, how does one choose a corresponding state sequence $Q = q_1, q_2, \dots, q_T$, which is optimal to generate the observation sequence. The Viterbi algorithm is used in this situation.
- **Training**, this relates with how one adjusts the parameters $\lambda = (A, B, \pi)$ to maximize $P(O|\lambda)$ the likelihood of all observation sequences. This last step is a problem of determining the reference model faults and is solved using Baum-Welch algorithm.

The difference between the continuous HMM and the Bayesian approach is that the CHMM model is estimated using the Baum-Welch algorithm whereas our Bayesian approach is treated as a variational approximation method.

The method proposed here for process monitoring of the rolling bearing element fault under varying condition using the BRNHMM involves **training** and **detection**. The following assumptions are made: Each training set constitutes an observation sequence $O = o_1, o_2, \dots, o_T$ and each fault state is modeled by using an HMM. The fault state also has finite training sets and lastly, that number of faults to be monitored is L. The two steps involved here in this work are designed to:

1) Build a HMM model λ_L for each fault state L. Hence there is the need to estimate the model parameters for the compact notation (A, B, π) which optimizes the likelihood of the training set of the observation sequence for the L^{th} fault state or to maximize the $P(O|\lambda_L)$ of the probability of observation sequence O given the model λ_L. The Baum-Welch re-estimation algorithm which is also known as the expectation maximization (EM) is used in this step (Ghahramanj Zoubin, 2001). One needs to write out the log probability of the hidden variables and observations to derive the EM algorithm for learning the parameters.

$$logP(S_{1:T}, Y_{1:T}) = logP(S_1) + \sum_{t=1}^{T} logP(Y_t|S_t) + \sum_{t=2}^{T} logP(S_t|S_{t-1}) \quad (3.1)$$

where S is a continuous state e.g. the state at time t taking on the value of "2" is represented as $S_t = [0\ 1\ 0 \ldots\ 0]^T$.

2) In the context of the EM algorithm, the variational Bayesian is the learning algorithm for continuous HMM as the dimensionality, cardinality and number of variables can be achieved. The negative free energy, F, is important in maximizing the marginal likelihood and can be defined by equation (3.2) below

$$F(\theta) = \int S(\theta)\, logP(O|\theta)\partial\theta - KL[S(\theta)||P(\theta)] \quad (3.2)$$

the first term is the average likelihood of the data while the second term is the Kullback-Leibler (KL) divergence in approximating the posterior S and the prior P which is given by equation (3.3)

$$KL[S||P] = \int S(\theta) log \frac{S(\theta)}{P(\theta)} \partial\theta \quad (3.3)$$

if KL is positive and greater than zero, and F provides a lower bound on the model log-likelihood (Fernando, 2008). When KL is zero, F becomes equal to the model log-likelihood and $S(\theta)$ becomes equal to the posterior $P(\theta)$ making the model to converge.

3) Applying the Jensen inequality twice, the model can be lower bounded of which the basic idea is to simultaneously approximate the distribution over both the hidden states and parameters with a simpler distribution. This iteratively maximizes F as a function of two

free distributions $Q(S)$ and $Q(\theta)$. Maximizing F is equivalent to minimizing the KL divergence between $Q(S)\,Q(\theta)$ and the joint posterior over the hidden states and parameters $P(S,\theta|D,M)$.

$$logP(D|M) = log\int \partial\theta\ P(D,\theta|M) \tag{3.4}$$

$$\geq \int \partial\theta\ Q(\theta)log\frac{P(D,\theta|M)}{Q(\theta)} \tag{3.5}$$

$$= \int \partial\theta\ Q(\theta)\left[logP(D|\theta,M) + log\frac{P(\theta|M)}{Q(\theta)}\right] \tag{3.6}$$

$$\geq \int \partial\theta\ Q(\theta)\left[\sum_s Q(S)log\frac{P(S,D|\theta,M)}{Q(S)} + log\frac{P(\theta|M)}{Q(\theta)}\right] \tag{3.7}$$

$$\equiv F\big(Q(\theta),Q(S)\big) \tag{3.8}$$

where M stands for model, and D the data of the system. Detecting the unknown fault type, features are extracted from the acoustic emission signal and vectors are formed for vector quantization which is followed by the estimation of the model likelihood for all possible models, $P(O/\lambda_L)$, $1 \leq l \leq L$. The model with the highest likelihood is the best score for representing the fault condition i.e. $L^* = \underbrace{argmax}_{1\leq l\leq L}[P(O|\lambda_L)]$ the figure 3.1 below shows the flow chart of the training to fault detection procedure.

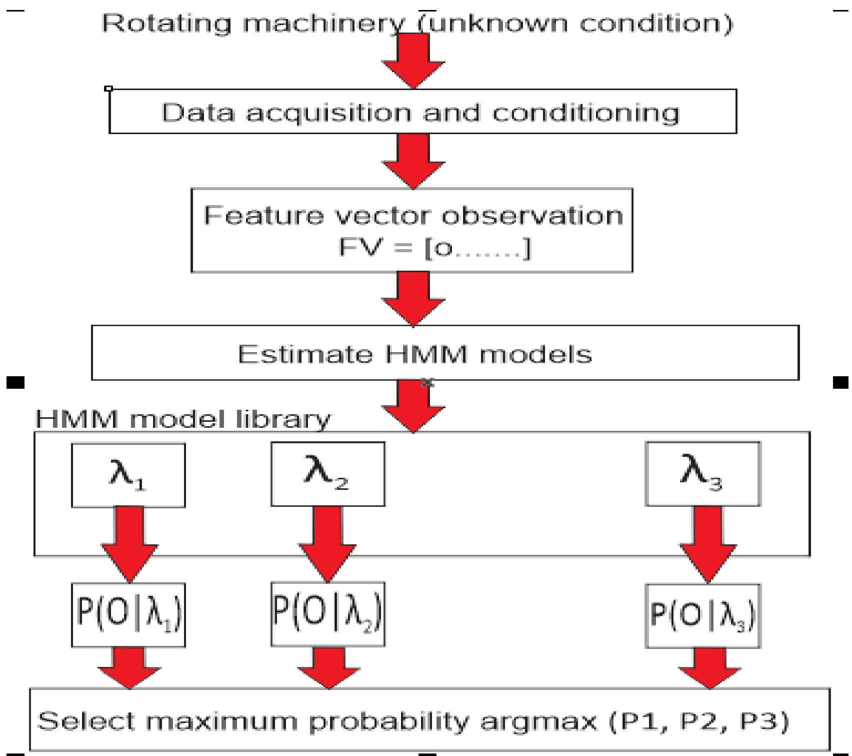

Figure 3.1, The HMM pattern-based classification for BRNHMM see Li et al. 2005.

It is assumed that the output probability density function can be written as $P_t(x) = \sum_{m=1}^{M} C_{jm} N(x|\theta_{jm})$ with $\sum_{m=1}^{M} C_{jm} = 1$, where C_{jm} is the mixture coefficient and $N(x|\theta_{jm})$ is the Gaussian density. Some Matlab functions were written to help facilitate the program among which are the bsxfun, erfc (complementing error function), genparam (which generates initial input parameters from a M-by-N training data) and the BRNHMM (that constructs a Bayesian robust new hidden Markov model with many hidden states and real-value features).

3.2 Discussion of BRNHMM result.

Fault classification was performed for three classes, which are the good bearing, debris induced bearing and an outer race cracked bearing. Features were extracted from the AE signals obtained from these bearings and used to train both BRNHMM and artificial neural networks (ANN) for fault classification of low speed bearings which were loaded sinusoidally along the axial and radial directions. The essence of this work is to show which model or network between BRNHMM and ANN better classify faults, regardless of the load and speed applied simultaneously, especially at varying conditions mimicking rolling mill plants that operate at low speed and varying load condition, for achieving fault classification.

Figure 3.2 below represents the BRNHMM state model diagram of the fault classification with respect to the experiment performed. The BRNHMM has three states with state 1 representing the good bearing (G), 2 stands for the debris induced bearing (D) while 3 represents the bearing with the crack defect in the outer race (O).

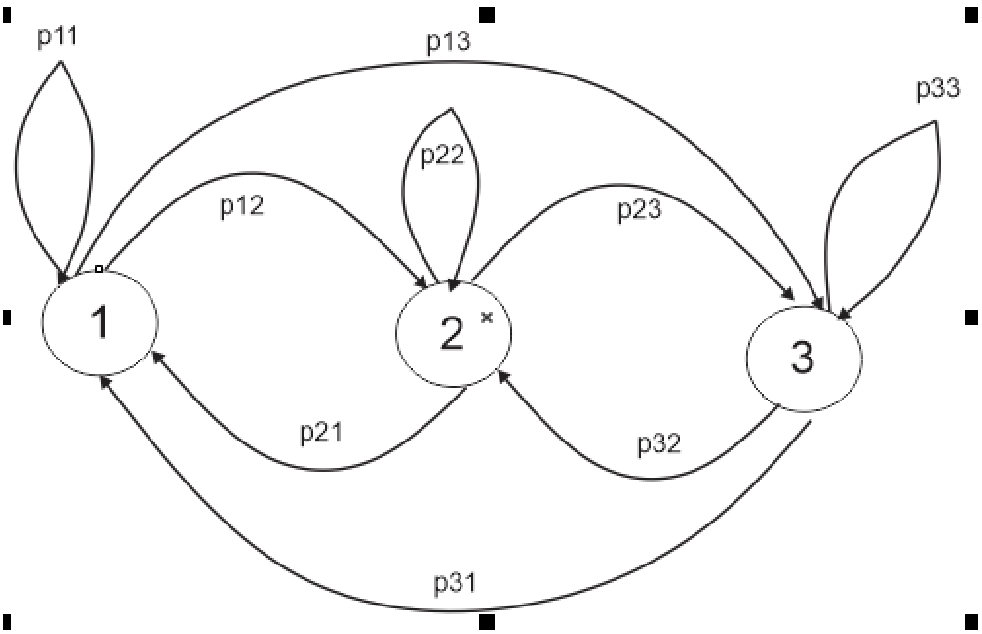

Figure 3.2, BRNHMM state model for the bearing fault classification.

The transition between each state is presented in Table 3.1 below.

Table 3.1, Transition between each state in the BRNHMM model.

G	P11	P12	P13
D	P21	P22	P23
O	P31	P32	P33
	G	**D**	**O**

From the experimental run, training was done with good bearing data, debris induced bearing data and the outer race crack bearing data each for speeds 70, 80, 90 and 100 rpm.

Figures 3.3 a, b, and c and 3.4 a, b, and c present the variational lower bound on the model that took place in the training process for optimizing the posterior parameters at 70 and 100 rpm respectively. We limit ourselves to showing just the plots of the variational lower bounds at 70 and 100 rpm because of space limitation.

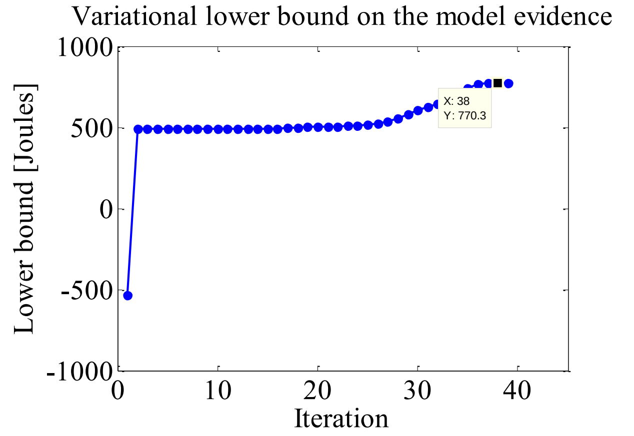

Figure 3.3, a) Variational lower bound for good bearing with training at 70 rpm.

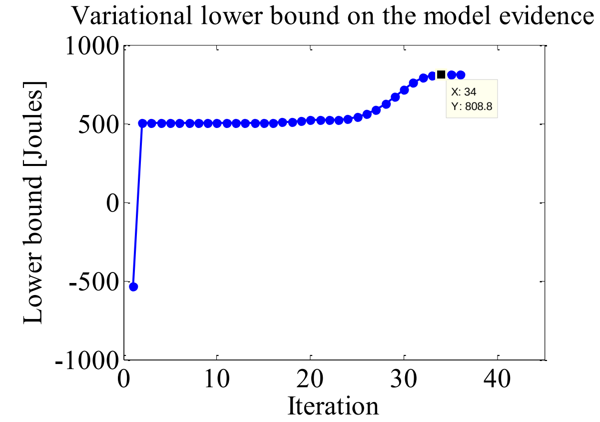

Figure 3.3, b) Variational lower bound for debris induced bearing with training at 70 rpm.

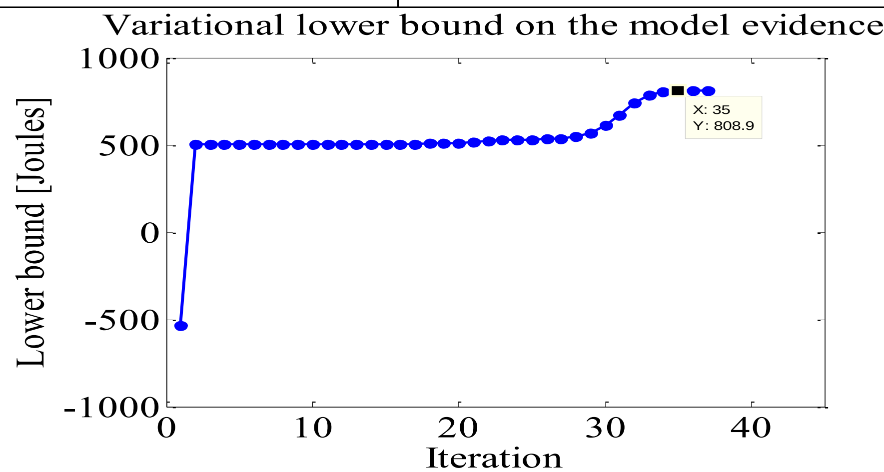

Figure 3.3, c) Variational lower bound for outer race cracked bearing with training at 70 rpm.

In figure 3.3 good bearing model, the free energy of the lower bound which is at 770.3 joules begin to stabilize after around 38 iterations at 70 rpm, whereas for the anomalous models the energy of the lower bound was at 808.8 joules and stabilized a bit faster with that of the debris stabilizing the lowest after 34 iterations. While for the outer crack race it stabilizes at 35 iterations with the energy of the lower bound been at 808.9 joules. As found in these experiments conducted, it indicates that when a crack is found in a working bearing it tends to exhibit higher energy on the working bearing, followed by reduced energy when debris are found in the bearing. Although the difference in energy in the third case compared to case two is less than 3%, this is due to several reasons of which a few could be that the speed variation is not much between both bearings and for the fact that data are captured at varying frequency intervals of applications.

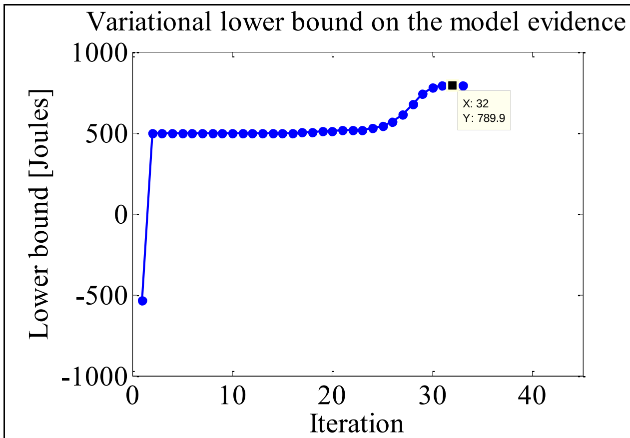

Figure 3.4, a) Variational lower bound for good bearing with training at 100 rpm.

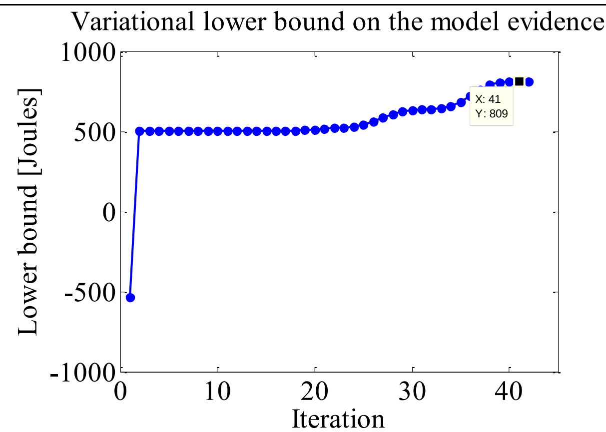

Figure 3.4, b) Variational lower bound for debris induced bearing with training at 100 rpm.

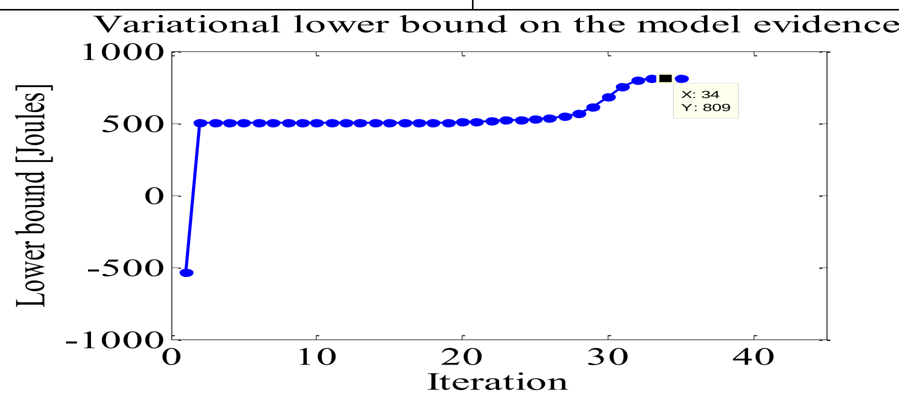

Figure 3.4, c) Variational lower bound for outer race cracked bearing with training at 100 rpm.

From figures 3.4 a, b and c for the good bearing model the free energy of the lower bound is at 789.9 joules and begins to stabilize after around 32 iterations at 100 rpm, whereas for the anomalous models the energy of the lower bound is at 809 joules and stabilized a-bit longer with that of the debris stabilizing the highest after 41 iterations. The outer crack race model stabilized after 34 iterations at 100 rpm with the energy of the lower bound being at 809 joules. These experiments indicate that energy of the lower bound is usually highest when cracks are formed in the bearing than when debris is formed when at a low speed (with conclusion drawn from the four speeds considered) and is at its lowest in a well working bearing.

The ANN was trained with thirteen well-established statistical features found in the literature. These features were extracted from the acoustic emission (AE) time history data and have been used for fault classifications in other literature. They are the mean, standard deviation, crest factor, root-mean-square (RMS), variance, norm, sum, minimum, maximum, median, range, skewness and kurtosis. The Bayesian robust new hidden Markov model for the models of the three categories of state used in the training process for optimizing the posterior parameters for both healthy and anomalous models is shown in figure 3.5. As stated in paragraph 4.2 “the model with the highest likelihood is usually the best score for representing the fault condition”, it was found after computation that for each model under the different speed conditions, the model with the highest likelihood was that of the condition represented.

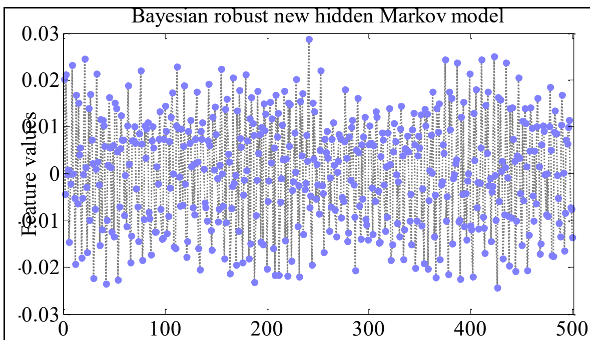

Figure 3.5, a) Bayesian robust new HMM for normal bearing model at 100 rpm.

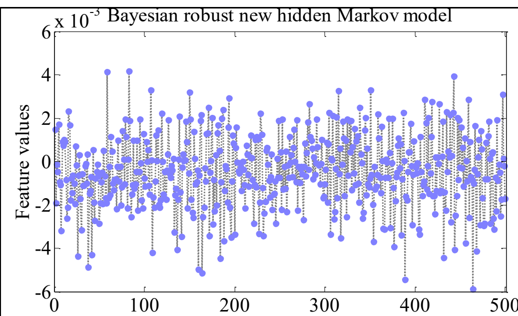

Figure 3.5, b) Bayesian robust new HMM for debris induce bearing model at 100 rpm.

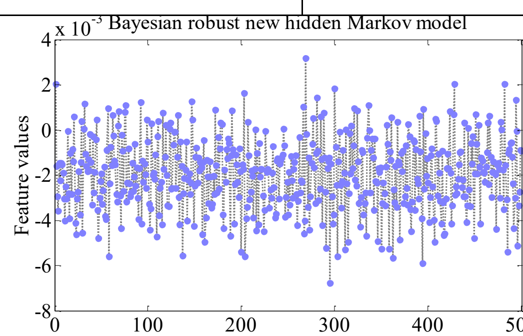

Figure 3.5, c) Bayesian robust new HMM for outer race crack bearing model at 100 rpm.

Table 3.2 shows the feature equations for the training done in the ANN.

Table 3.2, Features extracted from the AE time histories data and their equations (Rao and Horton, 2011; Zaeri et al., 2011).

Feature	Symbol	Equation
Kurtosis	$X_{kurtosis}$	$\frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^4}{(N-1)\sigma^4}$
Skewness	$X_{skewness}$	$\frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^3}{(N-1)\sigma^3}$
Crest factor	X_{CF}	$\frac{\max(\lvert x_i \rvert)}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}}$
Root-mean-square	X_{rms}	$\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$
Standard deviation	$X_{std} = \sigma_x$	$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}$
Variance	$X_{\mathrm{var}} = \sigma_x^2$	$\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2$
Norm	$X_{norm} = \lVert X \rVert_2$	$\sqrt{\sum_{i=1}^{N}\lvert x_i \rvert}$
Mean	$X_{mean} = \bar{X} = \mu_x$	$\frac{1}{N}\sum_{i=1}^{N} x_i$
Sum	X_{sum}	$\sum_{i=1}^{N} x_i$
Median	X_{median}	$median(x)$
Minimum	$X_{\min}$	$\min(x)$
Maximum	$X_{\max}$	$\max(x)$
Range	X_{range}	$x_{\max} - x_{\min}$

With respect to the training just as it was for the BRNHMM in the ANN training, the good bearing represent index 1, the debris induced bearing represent index 2 while the outer race crack bearing index 3. Targets were set for each class index while the simulated result was computed using Matlab. The ANN network has three layers (the input layer, the hidden layer and the output layer) with the hidden layer having 10 states. Table 3.3 shows the result obtained after the training was done for the ANN.

Table 3.3, ANN output classification result.

	Result at speed 70 rpm, took 27 iterations		
Target class	**2**	1	3
Simulated class	**1**	**1**	**3**
	Result at speed 80 rpm, took 21 iterations		
Target class	2	1	3
Simulated class	2	1	3
	Result at speed 90 rpm, took 42 iterations		
Target class	2	1	3
Simulated class	2	1	3
	Result at speed 100 rpm, took 26 iterations		
Target class	2	1	3
Simulated class	2	1	3

It was observed from the training run that at low speed under the varying load and low speed conditions which mimicked the rolling mill plants, the ANN was not able to classify correctly at a low speed of 70 rpm but shows correct classification at higher speed see table 3.3. Unlike the ANN, the BRNHMM correctly classify for the difficult condition of low speed and varying load.

Table 3.4 shows the result of the log-likelihood for the different model run for BRNHMM.

Table 3.4, Training model log-likelihood result.

Training Model	Log-likelihood				
	Fault class	**Speed**			
		70 rpm	**80 rpm**	**90 rpm**	**100 rpm**
Good bearing data	**G**	**131.4538**	**184.3939**	**96.9743**	**45.3567**
	D	121.1390	119.3415	91.3265	42.0180
	O	129.5494	179.6677	96.5922	45.2185
Debris induced data	**G**	73.3347	177.5201	125.1642	246.7132
	D	**86.0431**	**247.1741**	**131.2389**	**250.1108**
	O	84.4187	246.8940	131.1483	250.0533
Outer race crack data	**G**	139.1618	97.9514	339.7586	60.5178
	D	148.1032	167.1507	346.0991	63.9386
	O	**149.8086**	**172.9499**	**346.2743**	**64.0194**

3.3 Summary.

Presented here in this work is a system satisfactory for on-line evaluation of low speed and varying load application in roller bearing by use of BRNHMM classifiers. Training the model on data obtained directly from the experiment was performed successfully and this helps to ensure the validity of the model. It is a robust method in that the BRNHMM works satisfactorily with raw signals with little or no pre-processing and the effective use of Kullback-Leibler divergence method for obtaining faster iteration in the various models made this work unique to the other methods in existence.

Through the experimental study of roller bearing rotating at low speed and varying load condition, we have shown how BRNHMM is able to represent the most relevant aspects of the sensory signals. Hence it is evident that from the diagnostics accuracy shown by BRNHMM that its representation when compared to that of ANN is quite satisfactory as it classifies accurately even at a low speed of 70 rpm which is not achievable with the use of ANN.

Section 1.2.5.2 records the disadvantage of the use of HMM which makes it unfit for on-line use (due to the involvement of large data sets to train the model for fault classification). This was overcome here by using the BRNHMM combined with the selection of the output symbol vector by the Gaussian density of continuous HMM and this has become a successful tool for process monitoring and fault detection in roller bearing operating in a complex scenario like as found in rolling mills.

4.0 Chapter Four Fault classification, diagnosis and prognostics in REB.

4.1 Fault classification.

Support vector regression machine (SVRM) theory is formed based on the principle of support vector machine (SVM) which is used for time series prediction. The main difference between ANN and SVM or SVRMs is in the principle of risk minimization (RM). In SVM, the structural risk minimization (SRM) principle is used, thereby minimizing an upper bound on the expected risk while in ANN, traditional empirical risk minimization (ERM) is used minimizing the error on the training data and the difference in RM leads to better generalization performance in SVM than ANN (Shen et al. 2014; Hariharan & Srinivasan, 2009; Rao & Horton 2011; Kim et al. 2006). The aim of SVRMs is to obtain a function $F(x)$ which can predict the output y_i within the error limit of ε with the estimation function $f(x)$ being as flat as possible to ensure good generalization and variance. The function is given as follows

$$f(s) = w \cdot x + b \tag{4.1}$$

where w is the weight vector and b is a constant. This function is obtained by solving the following optimization problem

SVM for regression uses the same principles as SVM, for classification. With regression, a margin of tolerance ϵ is set in approximation to the SVM. Once trained, the SVMGA will generate predictions using the formula:

$$f_{(x)} \equiv \sum_{i=1}^{m} \vartheta_i \emptyset(x, x_i) + b \tag{4.2}$$

For us to minimize the error, we will be individualizing the hyper-plane that maximizes the margin knowing fully well that of the error is been tolerated.

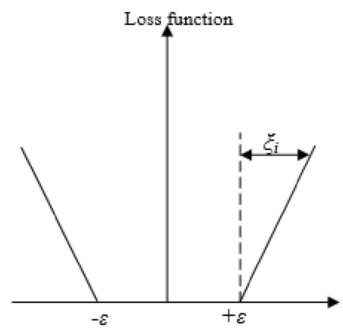

Figure 4.1, The loss function of SVM.

Figure 4.1 shows the loss function for SVR. The quadratic optimization problem becomes

$$\min_{w} \frac{1}{2} w^T \cdot w \tag{4.3}$$

$$s.t. \begin{cases} y_i - (w^T \cdot \emptyset(x) + b) \leq \epsilon \\ (w^T \cdot \emptyset(x) + b) - y_i \leq \epsilon \end{cases}$$

where $\emptyset(x)$ is the kernel function and w is the margin.

If the bound is added to the set, the tolerance on error that can be committed will be given as;

$$\min_{w,b} \frac{1}{2} w^T \cdot w + C \sum_{i=1}^{m} (\xi_i + \xi_i^*) \tag{4.4}$$

$$s.t. \begin{cases} y_i - (w^T \cdot \emptyset(x) + b) \leq \epsilon + \xi_i \\ (w^T \cdot \emptyset(x) + b) - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \quad i = 1 \dots \dots \dots \dots \dots m \end{cases}$$

where ξ_i and ξ_i^* are the slack variables and C is a positive constant which penalizes the errors larger than $\pm\varepsilon$ using ε- insensitive loss function given. Now being a minimization problem, we can set all constraints ≥ 0 by multiplying through by a negative sign:

$$\mathrm{R} = \min_{w,b} \frac{1}{2} w^T \cdot w + C \sum_{i=1}^{m} (\xi_i + \xi_i^*) \tag{4.5}$$

$$s.t. \begin{cases} -y_i + (w^T \cdot \emptyset(x_i) + b) + \epsilon + \xi_i \geq 0 \\ y_i - (w^T \cdot \emptyset(x_i) + b) + \epsilon + \xi_i^* \geq 0 \\ \xi_i, \xi_i^* \geq 0, \quad i = 1 \dots \dots \dots \dots \dots m \end{cases}$$

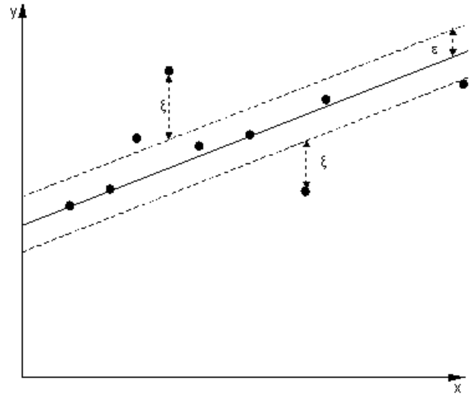

Figure 4.2, The regression line of SVM.

Figure 4.2 shows the regression line of SVM, the upper and lower boundary lines. For solving the optimization problem given in eq. (4.3) the following Lagrangian is needed:

$$L = \frac{1}{2} w^T \cdot w + \sum_{i=1}^{m} C(\xi_i + \xi_i^*) - \sum_{i=1}^{m} (\eta_i \xi_i + \eta_i^* \xi_i^*) - \sum_{i=1}^{m} \alpha_i (\varepsilon + \xi_i - y_i + w \cdot x_i + b) - \sum_{i=1}^{m} \alpha_i^* (\varepsilon + \xi_i^* + y_i - w \cdot x_i - b) \quad 4.6)$$

Subject to $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0,\ i \dots \dots \dots m$ are Lagrange multipliers that must be satisfied with the partial derivatives of the Lagrange equation L w.r.t. the primal variables w, b, ξ_i, ξ_i^* having to vanish for optimality

$$\begin{cases} \dfrac{\partial L}{\partial b} = \sum_{i=1}^{m}(\alpha_i^* - \alpha_i) = 0 \\ \dfrac{\partial L}{\partial w} = w - \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)\emptyset(x_i) = 0 \qquad \rightarrow w = \sum_{i=1}^{m}\emptyset(x_i)(\alpha_i - \alpha_i^*) \\ \dfrac{\partial L}{\partial \xi_i} = C - \alpha_i - \eta_i = 0 \qquad \rightarrow \eta_i = C - \alpha_i, \quad \alpha_i \in [0, C] \\ \dfrac{\partial L}{\partial \xi_i^*} = C - \alpha_i^* - \eta_i^* = 0 \qquad \rightarrow \eta_i^* = C - \alpha_i^*, \quad \alpha_i^* \in [0, C] \\ \dfrac{\partial L}{\partial \eta_i} = \sum_{i=1}^{m}\xi_i = 0 \\ \dfrac{\partial L}{\partial \eta_i^*} = \sum_{i=1}^{m}\xi_i^* = 0 \end{cases} \tag{4.7}$$

Substituting eq. (4.7) into eq. (4.6), the dual optimization problem is given, for the sake of space a lot of substitution has been omitted to arrive at the following

$$max\frac{-1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\emptyset(x_i)(x_j)\,(\alpha_i - \alpha_i^*)\left(\alpha_j - \alpha_j^*\right) - \varepsilon\sum_{i=1}^{m}(\alpha_i + \alpha_i^*) + \sum y_i(\alpha_i - \alpha_i^*) \tag{4.8}$$

$s.t. \sum_{i=1}^{m}(\alpha_i - \alpha_i^*) = 0$ and $\alpha_i, \alpha_i^* \in [0, C]$

By solving the optimization problem, a linear regression function is presented as follows

$$f(x) = \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)(x_i \cdot x) + b \tag{4.9}$$

To compute for b we exploit the Karush-Kuhn-Tucker (KKT) conditions which state that at the point of the solution the product between dual variables and constraints must vanish,

$$\alpha_i(\varepsilon + \xi_i - y_i + \langle w, x_i\rangle + b) = 0 \tag{4.10}$$

$$\alpha_i^*(\varepsilon + \xi_i^* + y_i - \langle w, x_i\rangle - b) = 0$$

And

$$(C - \alpha_i)\xi_i = 0$$

$$(C - \alpha_i^*)\xi_i^* = 0 \tag{4.11}$$

which allows us to make relevant conclusions that are useful. Firstly, only samples (x_i, y_i) with corresponding $\alpha_i^{\langle * \rangle} = C$ lie outside the ε-insensitive tube and secondly $\alpha_i, \alpha_i^* = 0$, meaning that there can be never a set of dual variables α_i, α_i^* which are both simultaneously nonzero (Hamadache and Lee, 2014). Thus, allowing us to conclude that

$$\varepsilon - y_i + \langle w, x_i \rangle + b \geq 0 \text{ and } \xi_i = 0 \text{ if } \alpha_i < C \tag{4.12}$$

$$\varepsilon - y_i + \langle w, x_i \rangle + b \leq 0 \text{ if } \alpha_i > 0 \tag{4.13}$$

The kernel function is applied here to map the input vector into a high dimensional space because the linear regression function is not enough to process the non-linear problem. Hence the regression function is derived as follows

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) K(x_i, x) + b \tag{4.14}$$

where $K(x_i, x) = \varphi(x_i) \cdot \varphi(x)$ is a symmetric positive defined kernel function given by the Mercer's theorem (Xie, 2011; Sloukia, Bouarfa and Medromi, 2013). In this work, the Gaussian and the exponential kernel functions are adopted which are respectively examples of the radial function kernel and the cubic polynomial function

$$K(x, y) = exp\left[\frac{-\|x - y\|^2}{2\sigma^2}\right] \tag{4.15}$$

where σ is a positive real number, alternatively it could be implemented using

$$K(x, y) = exp(-\gamma \|x - y\|^2) \tag{4.16}$$

The adjustable parameter σ plays an important role in the performance of the kernel and hence should be carefully tuned (by either using the classical technique which employs some method of determining a subset of centres or by first clustering to select a subset of centres) to the problem being solved. If overestimated, the exponential will behave almost linearly, and the higher dimensional projection will start to lose its non-linear power, and if underestimated the function will lack regularization making the decision boundary highly sensitive to noise in the training data.

4.1.1 Support Vector Machines.

A genetic algorithm is employed here to configure the SVR for regression. In these genetic algorithms consecutive populations of feasible solutions are created. It then evolves a population of chromosomes as potential solutions to an optimization problem for which the optimal solution is often obtained after a series of iterative operations.

To evaluate the fitness function of the chromosomes and genetic operators, selection and reproduction are employed to create new populations. A typical genetic algorithm requires: (1) A genetic representation of the solution domain and (2) A fitness function to evaluate the solution domain.

Once the genetic representation and the fitness function are both defined, the GA proceeds to initialize a population of solutions and then to improve it through repetitive application of the mutation, crossover, inversion and selection operators.

The three most important aspects of using GA are: (1) definition of the objective function, (2) definition and implementation of the genetic representation, and (3) definition and implementation of the genetic operators. Once these three have been defined, the generic GA should work fairly well.

Since a solution must represent a SVM for regression, the corresponding chromosomes are composed by two genes i.e. one for each SVM for regression parameter and the values of the genes which are obtained in ranges [0.01, 60000] and [1.0E-6, 8] for the genes representing C and γ respectively. The values of C and γ are limited into certain ranges to assure the generalization capability of the SVM for regression. These chromosomes are constructed by using a binary coding system (see table 4.2). To determine the optimal values of the regularization parameter C and γ which determines the tradeoff between the fitting error of the SVR for regression model and the model complexity and which assures the optimal accuracy and generalization stability simultaneously, the GA is used. It should be noted that the values of parameters C and γ are limited to certain ranges so as to assure the generalization capability of the SVR for regression. The values of these parameters in this work were set for [0.01, 60000] and [1.0E-6, 8]. The crossover and mutation rates were set to 0.5 and 0.1 respectively. The evolutionary process was terminated using two stopping criteria which are after 600 generations or if the fitness value of the best solution does not change after 60 generations. The formula for the fitness function used is $f = sum(x.^{*}z)/sum(z)$ and the constraints is $0 < x < 1, \quad sum(z.^{*}x) = class$

4.1.2 Genetic tuning to configure SVM.

The procedure used to develop the classifier in order to achieve the fault classification for the three conditions of consideration starts from the acoustic emission acquisition from two bearings which have faults induced and a good bearing. GA is used to determine the optimal values of the regularization parameter C and γ which determines the tradeoff between the fitting error of the SVR for regression model and the model complexity and assures the optimal accuracy and generalization stability simultaneously. It evolves a population of chromosomes as potential solutions to an optimization problem which is obtained after a series of iterative operations. The fitness function in GA is used to evaluate the goodness (i.e. the fitness) of the chromosomes and the genetic operators based on selection to create new populations (i.e. generations). The individual which gives the best solution in the final population was taken to define the best approximation to the optimum problem of investigation. After the signals created by the machine with normal and faulty conditions were measured, the optimized developed classifier was then used to classify the operational conditions of the machines. The GA is used to find the optimal values of C and γ that assures the optimal predication accuracy and generalization ability of the SVM for regression simultaneously. To determine the individuals that are included in the next generation (i.e. survivals) we employed tournament selection where only the best n solutions are copied straight into the next generation.

The three groups of bearing problems and their classification is presented in Table 4.1, while

Table 4.1, Bearing grouping and their classification.

Fault Grouping	Classification	Bearing speed (rpm)	Class index	Binary code used
Good bearing	Good	70	1	0 0 1
Good bearing	Good	80	1	0 0 1
Good bearing	Good	90	1	0 0 1
Good bearing	Good	100	1	0 0 1
Debris induced bearing	Debris induced	70	2	0 1 0
Debris induced bearing	Debris induced	80	2	0 1 0
Debris induced bearing	Debris induced	90	2	0 1 0
Debris induced bearing	Debris induced	100	2	0 1 0
Outer race crack bearing	Outer race	70	3	1 0 0
Outer race crack bearing	Outer race	80	3	1 0 0
Outer race crack bearing	Outer race	90	3	1 0 0
Outer race crack bearing	Outer race	100	3	1 0 0

4.2 Results and discussion.

The skewness and kurtosis indicator used as input to the system often provide good detection at high speeds and decreases sharply in their detection abilities as the speed decrease as is reported in chapter 2 of this work.

A test was done on neural network pattern recognition and table 4.2 shows the result obtained after the training was done for the neural network (NN).

Table 4.2, NN output classification result.

	Result at speed 70 rpm, took 27 iterations		
Target class	**2**	**1**	**3**
Simulated class	1	1	3
	Result at speed 80 rpm, took 21 iterations		
Target class	2	1	3
Simulated class	2	1	3
	Result at speed 90 rpm, took 42 iterations		
Target class	2	1	3
Simulated class	2	1	3
	Result at speed 100 rpm, took 26 iterations		
Target class	2	1	3
Simulated class	2	1	3

The network has three layers (the input layer, the hidden layer and the output layer) with the hidden layer having 10 states. It was observed from the training run that at low speed under the varying load and low speed conditions which mimicked the rolling mills plants, the NN was not able to classify correctly at a low speed of 70 rpm but shows correct classification at higher speed. Figure 4.4 shows the confusion matrix for the neural network classification method.

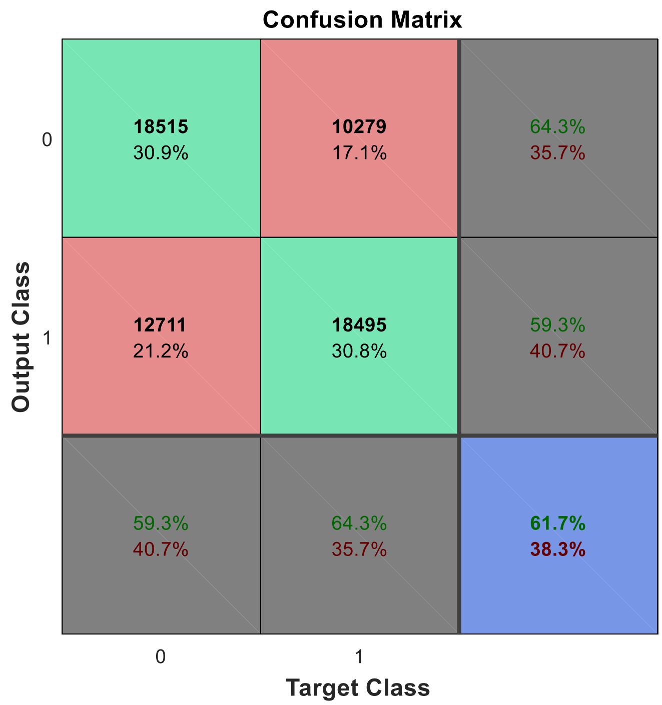

Fig. 4.4, **The confusion matrix for the NN classification method.**

The total percentage of correctly classified cases with the NN as specified in Fig. 7 is 61.7 % while the misclassified cases is 38.3 %. This result from the confusion matrix shows that NN is not a very good classifier for the varying load and low speed condition of rolling element bearing.

As indicated in section 4.3, the number of inputs to the network for the SVM for regression in this work is thirteen of which GA was used to optimize and select the best few features and also to reduce the dimensionality that describes the problem. The SVMGA can be used for several applications in the field of engineering of which it is an innovation of SVM. Here it is used to develop the condition monitoring of the bearing i.e. to identify whether the bearing is defective or normal and to achieve this the Gaussian kernel, exponential kernel, spline, radial basis function (rbf), polynomial, periodic and sigmoid function were used to perform various test of classification, so as to classify the simulated output to the target class that was set for the various conditions of the bearings as indicated in the experimental setup to check for which that can be used under this scenario. Only the Gaussian kernel functions proved effective for this purpose under this test at varying load and speed and at low speed condition.

The most suitable method was chosen according to the application constraints and the number of training samples, for easy recognition of classes. Target data for recognition was set to consist of vectors of all zero values except for a 1 in element I, where I is the class they represent (see table 4.1). Table 4.2 presents the values of the best possible classification values generated by GA by looking for weight values that produces the best fitness result based on the set of constraints provided in section 4.3. The debris induced bearing was classed 2, while the good bearing was classed 1 and the outer race cracked bearing was classed 3.

Table 4.2, Genetic algorithm best fitness values for the bearings.

Bearing types	Applying classifier class	Generated GA possible class value matching to a measurement			Speed (rpm)
		Good bearing	Debris induced bearing	Outer race cracked bearing	
Good bearing	1.000000e+00	1.0000	-0.3968	-2.9452	70
Debris induced bearing	2.000000e+00	-0.6540	2.0000	1.2329	
Outer race cracked bearing	3.000001e+00	-1.0547	0.7102	3.0000	
Good bearing	9.999999e-01	1.0000	0.1936	-2.9858	80
Debris induced bearing	2.000000e+00	-0.7467	2.0000	0.3255	
Outer race cracked bearing	3.000000e+00	-1.1085	0.1383	3.0000	
Good bearing	1.898637e+00	1.0000	0.8860	1.8986	90
Debris induced bearing	1.142809e+00	1.1428	-0.5207	-0.7566	
Outer race cracked bearing	3.000000e+00	0.0005	1.4212	3.0000	
Good bearing	1.000000e+00	1.0000	0.4193	-0.9147	100
Debris induced bearing	2.000000e+00	-1.2400	2.0000	-0.2920	
Outer race cracked bearing	1.274203e+00	1.2742	0.0401	-0.8171	

The results obtained for the classification of the bearings at different low speeds using the kernel functions of Gaussian and exponential is presented in table 4.3 and 4.4 below.

Table 4.3, Result of bearing classification using Gaussian kernel function.

Class	Kernel function	Debris bearing	Good bearing	Outer race crack bearing	Speed (rpm)	Classifying result
Target	Gaussian	2	1	3	70	Correctly classified
Simulated		2	1	3		
Target	Gaussian	2	1	3	80	Correctly classified
Simulated		2	1	3		
Target	Gaussian	2	1	3	90	Correctly classified
Simulated		2	1	3		
Target	Gaussian	2	1	3	100	Correctly classified
Simulated		2	1	3		

Table 4.4, Result of bearing classification using exponential kernel function.

Class	Kernel function	Debris bearing	Good bearing	Outer race crack bearing	Speed (rpm)	Classifying result
Target	Exponential	2	1	3	70	Incorrectly classified
Simulated		0	2	-2		
Target	Exponential	2	1	3	80	Correctly classified
Simulated		2	1	3		
Target	Exponential	2	1	3	90	Correctly classified
Simulated		2	1	3		
Target	Exponential	2	1	3	100	Correctly classified
Simulated		2	1	3		

In table 4.4 it is observed that at the lowest bearing speed of 70 rpm the exponential kernel function could not classify the three bearing conditions according to the index assigned to them. There was wrong classification of the good bearing, outer-race defect and the exponential function could not classify the debris bearing as shown in figure 4.4, Index 2 was repeated twice but in opposite direction. At every other speed, it rightly classified the bearing condition. The precision obtained from the classification made by the Gaussian kernel function was perfect at the various speed category of the bearing condition. The simulated classes rightfully met their respective target class.

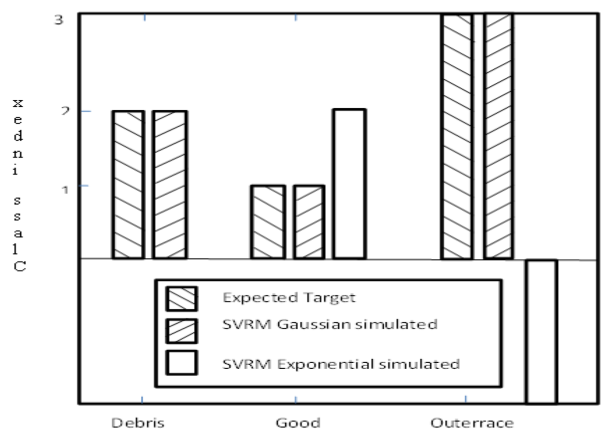

Figure 4.4, Bearing classification as simulated at 70 rpm.

The change in speed as indicated in the simulation done in table 4 did not affect the expected result for the Gaussian kernel. We were restricted to just three classes here in this work; the class of the good bearing indicated from table 1 as index 1, the debris induced bearing which is indicated as index 2 and the outer race cracked bearing indicated as index 3. To reduce space consumption bearing classification simulated at 90 rpm is given in figure 4.5.

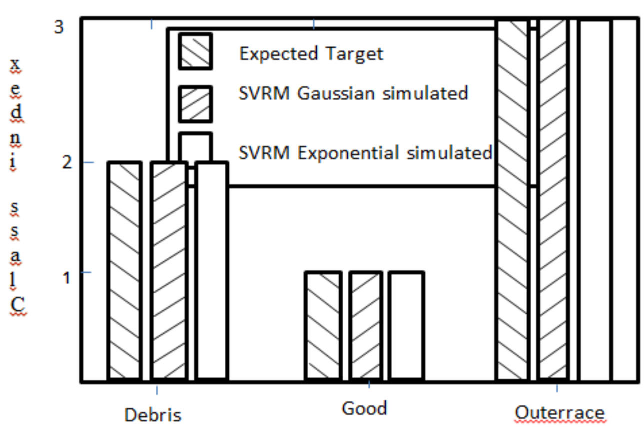

Figure 4.5, Bearing classification as simulated at 90 rpm.

4.3 Prognostics with SVDD.

Modern processes do not satisfy classical method assumptions, such as normality or linearity. To overcome this issue, introduction of new techniques from statistical machine learning theory has been applied. Control charts based on Support Vector Data Description (SVDD), a popular data classifier method inspired by Support Vector Machines, benefit from a wide variety of choices of kernels, which determine the effectiveness of the whole model. Among the most popular choices of kernels is the kernel principal component analysis, which enables SVDD to obtain a flexible data description, thus enhances its overall predictive capability. This work explores an even more robust approach by incorporating the Mahalanobis distance-based kernel (hereinafter referred to as Mahalanobis kernel) to SVDD for a varying load problem of a REB at low speed.

4.3.1 Theoretical background to support vector data descriptive algorithm for prognostics.

Support Vector Data Descriptive (SVDD) model is like the support vector machine, it maps its data into a high dimensional space. The computing tasks of SVDD model are concerned with the calculation of the radius (R) and centre (α) of the hypersphere by using the given data samples. R^2 is the distance from the centre of the hypersphere (α) to any of the support vectors on the boundary, where the radius of the hypersphere generated by the SVDD is used as a health indicator.

Using a hypersphere with minimized radius in the feature space to describe the dataset is the main work of SVDD i.e. all the samples should be in the hypersphere. Let {xi, i=1,2-----N} be the given training dataset with the dataspace, where N is the number of samples (α) and R denote the centre of the hypersphere respectively. The R of the hypersphere can be obtained by calculating the distance from its centre of any support vector with $0 \leq X_i \leq C$.

To determine whether a test data z is within the hypersphere its distance to the centre of the hypersphere must be calculated. Minimizing the radius of the hypersphere is the problem that can be described by the following quadratic equation with the inequality constrains as stated below (Benkedjouh et al., 2012).

$$minF(R, a, \xi_i) = R^2 + C \sum_{i=1}^{n} \xi_i \quad (4.17)$$

$$s.t. \begin{cases} x_i - a^2 \leq R^2 + \xi_i, & i = 1,2, \ldots \ldots, n. \\ \xi_i \geq 0, & i = 1,2, \ldots \ldots, n. \end{cases} \quad (4.18)$$

ξ_i is the slack variables of the $i - th$ training sample and C is a constant which determines the trade-off between the hypersphere volume and the training errors. The Lagrangian dual form of equation (12) can be restated as follow:

$$max_\alpha \sum_{i=1}^{n} \alpha_i k(x_i, x_i) - \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j k(x_i, x_j) \quad (4.19)$$

$$s.t.\begin{cases}\sum_{i=1}^{n}\alpha_i = 1\\ 0 \le \alpha_i \le C, \quad i = 1,2,\ldots,n\end{cases} \tag{4.20}$$

where $k(x_i, x_j) = \xi(x_i)^T \xi(x_j)$

from optimality condition we obtain,

$$R^2 = k(x_i, x_i) - 2\sum_{j=1}^{n}\alpha_j k(x_i, x_j) + \sum_{k=1}^{n}\sum_{j=1}^{n}\alpha_k \alpha_j k(x_k, x_j) \tag{2.21}$$

If $x_k \in SV < C$ for any set of support vectors which have $\alpha_k < C$, then

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i \alpha_j (x_i, x_j) = const. \tag{4.22}$$

Hence,

$$R = \sqrt{f(x_s)} \tag{4.23}$$

4.3.2 Discussion on the use of support vector data descriptive algorithm for prognostics.

The actual work in SVDD is to describe the dataset by using a hypersphere with minimized radius in the feature, meaning the samples should be in the hypersphere thus if outliers appear in the dataset, their corresponding distances to the center of the hypersphere will not be strictly smaller than R because of the formulated constrained convex optimization problem used which minimizes the function with the constraint condition. The outliers which have large distances from the center are penalized. Figures 4.3a and b below show the result obtained from SVDD plot at 70 and 90 rpm speeds.

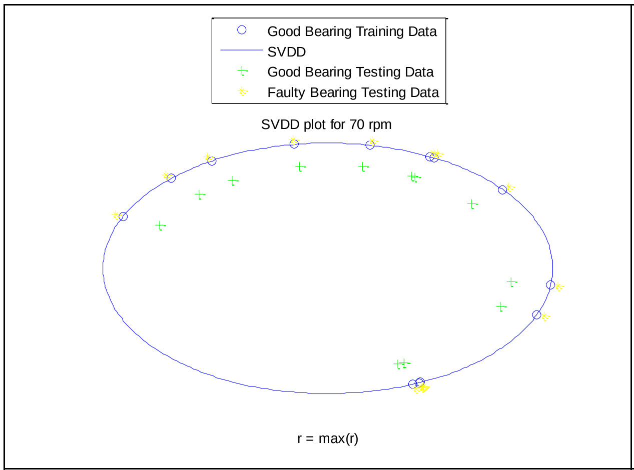

Figure 4.6, a) SVDD plot of the bearing life at 70 rpm.

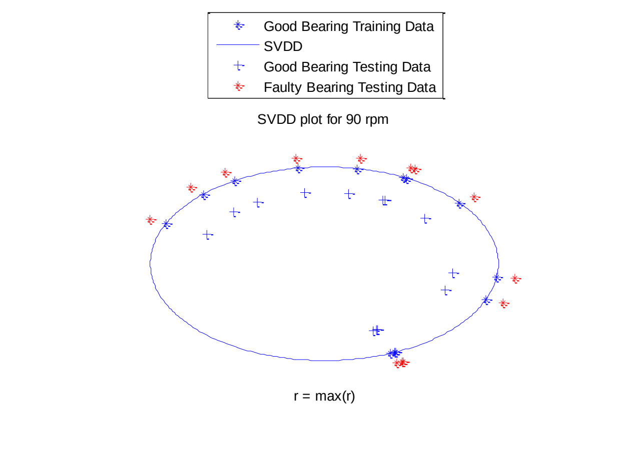

Figure 4.6, b) SVDD plot of the bearing life at 90 rpm.

As it can be seen, when the bearing was still in a good shape the datasets were mapped within the support vector data descriptive line (thus having its radius R less than the radius of the SVDD). The samples used for the training lied on the support vector line while the faulty bearing dataset was located outside the hypersphere. The Mahanalobis kernel was used to map the target data onto the bounded spherically shaped area in the feature space and the outlier objects were mapped outside this area it was also used for feature reduction and mapping from high dimensional space to fewer principal components there by taking care of the outliers.

4.4 Summary.

With respect to section 1.2.5.1, the introduction of SVMGA help to solve the problem associated to free parameter selection thereby providing solution which regularizes the parameter that controls the generalization performance of SVM and the ε-insensitive zone which determines the number of support vectors. With respect to the illustration, this work also involves extracting representative statistical parameters by using GA-based feature extractor from raw acoustic emission dataset and using it to classify the inputs for the SVM for regression.

The GA was not only able to enhance the classification performance, but also reduced the dimensionality that describes the problem. These features extracted cleared the doubts in obtaining intensity from the time spectrum. Thus, the features extracted proved to be good indicators of defect intensity and it also showed that with the Gaussian kernel function, effective classification can be achieved over exponential kernel function with SVMGA as it classified correctly in all conditional cases of the bearing faults with all the data set generated in this experimental work for which just a few sample of data results was showed due to space limitation.

The SVMGA algorithm has high accuracy in classification performance and wider generalization ability in small group of samples using the learning and testing pattern as revealed by the simulation results. Moreover, a combination of GA and SVM for regression has been used for intelligent fault diagnostics, showing that the proposed method can reduce the dimensionality in dataset, solve the problem of outliers and be used to optimize features parameter selection for training purpose. SVMGA proves to be a better classifier that maximizes the fault classification accuracy as it could identify faults and classify better with the Gaussian kernel function than any other kernel function like the polynomial, the rbf, and the spline with the sigmoid kernel function.

There are several challenges associated with the use of only SVM which is known for its time series prediction, among which is the free parameter selection. To date there is no universal method for hyper-parameter selection which contains some parameters such as the kernel parameter, the regularization parameter that control the generalization performance of SVM and the ε-insensitive zone which determines the number of support vectors. Hence if one is using arbitrary SVM parameters, the performance of SVM could differ over a wide range. This makes SVMGA (an innovation to SVM) a better option for solving classification problem especially at low speed conditions as it gives a decision that is much unique than SVM while retaining its classification accuracy.

Hence the SVMGA with the use of Gaussian kernel function proof to be more superior to the exponential kernel function in the classification of faults. Compared with the traditional classification methods reported in many other studies, this study aims at the effective classifying of different bearing fault pattern especially at low speed and at varying load condition for which the results showed whether the bearing is normal or defective.

The support vector data descriptive (SVDD) method based on Mahanalobis distance is considered within the frame-work of condition-based maintenance and predictive maintenance. The simulation work showed that Mahanalobis distance when used for feature reduction gives better output and that its use with SVDD can be used to predict bearing fault prognosis.

This new approach aimed to make SVDD boundary closely fit the contour of the target data. The proposed method can reduce the effect of outliers and yield higher classification rate and hence it is possible to fully exploit the potentiality of SVDD, whose performance depend on the configuration of the parameters.

The approach presents a flexible tool to support the strategies of project managers that might prefer to maximize a specific performance criterion, it also effectively set its parameters in order to improve fault predictions and can be deployed in critical real-time applications where the bearing failure affects the performance of the machine.

5.0 Chapter Five Prognosis and estimation of remaining useful life.

5.1 Introduction.

Faults in bearings often result in severe vibration of rotating machinery and timely detection of faults goes a long way to reduce downtime for the system and financial loss for owners. The ability to detect and sometimes isolate faulted components and its failure condition is known as diagnosis while prognostics is defined as the capability to predict the progression of a fault condition of a component failure and to estimate the remaining useful life (RUL) (Baruah and Chinnam, 2005).

Rather than just understanding the phenomenon which has just appeared like a failure, a prognostic process is all about the anticipation of its manifestation so as to take adequate action as soon as possible. The efficiency of a prognostic system is highly dependent on its ability to perform "good" predictions as reliability indicators follow from it. To further expatiate and give more meaning to the RUL obtained in chapter 4 the eXtended Takagi-Sugeno in combination with the well- known Paris law is used here.

The eXtended Takagi-Sugeno assumes the fuzzy rule-based system structure to be fixed. With there being some divergences in literature on the definition of prognostic, the International Organization for Standardization defines it that 'prognostic is the estimation of time to failure and risk for one or more existing and failure modes' (Tobon-Mejia *et al.*, 2012). It is well known that the fuzzy rule-based (FRB) systems are universal function approximators, they are suitable for extracting interpretable knowledge and provide a promising framework for designing effective and powerful prognostic, classification and control systems.

An incremental clustering procedure that takes account of the non-stationary nature of the data pattern and generates clusters that are used to form fuzzy rule-based systems antecedent part in on-line mode, is often used as a first stage of the non-iterative learning process. The decoupling of the learning task into a non-iterative, recursive clustering with a modified version of the well-known recursive parameter estimation technique leads to a powerful construct known as the eXtended Takagi-Sugeno (exTS). From control point of view, this equates to the system structure identification, a process that is usually ignored with attention being more often paid in most system to their systems parameter identification (tuning, adjustment, and adaptation) (Angelov and Zhou, 2006; El-Koujok, Gouriveau and Zerhouni, 2009).

The basic principle is of a gradual evolution of the fuzzy rule-based model structure in terms of fuzzy rules and their components- fuzzy sets, including variables of the antecedent part. The emerging evolving fuzzy systems paradigm mimics the evolution of individuals in nature during their life-cycle, specifically the autonomous mental development typical of humans: learning from experience, knowledge generation from routine operations, inheritance, gradual change, and rules extraction from data (Angelov and Filev, 2005; Angelov and Zhou, 2006).

With the eXtended Takagi-Sugeno (exTS) being a well-established tool for dealing with complex systems experiencing multiple operating modes, it has recently begun to find application in on-line identification methods. The problem associated with its on-line use has also been addressed in conjunction with model identification, control, fault detection, and signal processing. The exTS model is a fuzzy model whose rule-base and parameters continually evolve by adding new rules with more summarization power, and thereby modifying existing rules and parameters. The exTS learning algorithm is based on a recursive evaluation of the information potential of the new data points and the focal points of the rules hence adopting its name eXtended Takagi-Sugeno fuzzy recursive least square algorithm (exTSFRLSA) (Angelov, and Victor, 2004; Angelov and Zhou, 2008).

This work provides an approach for modeling the degradation feature of bearings thereby giving useful information on the RUL of the bearings by using the eXtended Takagi-Sugeno fuzzy recursive least square algorithm (exTSFRLSA) for prognostics purpose. The exTSFRLSA as used in this work is used for tuning, adjusting, and adapting the parameters involved in the propagation model by comparing predicted and measured defect sizes as in (Li et al., 1999; Angelov and Zhou, 2006), hence the instantaneous rate of defect propagation can be captured despite defect growth behavior variation and also to increase processing time of computation. The exTSFRLSA is a method that applies the principles of eXtended Takagi-Sugeno (which is a fuzzy rule-based phenomenon), the recursive least squares method and combining it with the Paris law, hence capturing crack growth which is applicable to online data capture.

In mathematical modeling of real world problems that cannot be specified, which has much complexity in the controlling of the system especially under uncertainty and vagueness, fuzzy theory is considered most suitable for solving its problem. The exTSFRLA is found to have clearly defined value of outputs of all fuzzy rules and is found prevalent in practical applications due to its speed hence, is very useful in system where the RUL of REB running under varying speed and load conditions is needed to be determined.

The solution to the generalized problem of inducing expert knowledge from large amounts of data as found with most work that uses fuzzy algorithm was solved here as good results were drawn from its application here. Another contribution this work brings is making the process less restricted to knowing a generic model for the fault development and its capabilities to work with any other area of non-linear modeling.

5.2 Prognostics with eXtended Takagi-Sugeno.

It is obvious to say that real systems are complex systems to deal with and their behavior are often non-linear and non-stationary, especially rolling element bearings operating under varying speed and load conditions. However, monitoring systems have evolved, and it is now easy to carry out on-line analysis. Thus, data-driven approaches are increasingly applied to machine prognostics. The eXtended Takagi-Sugeno fuzzy models have shown improved performance over conventional approaches hence, they can perform the degradation modeling step of prognostics. (Angelov, Filev and Member, 2004; Angelov and Zhou, 2006)

The eXtended case of Takagi-Sugeno models can be described as a set of fuzzy rules of the following form (El-Koujok, Gouriveau and Zerhouni, 2009):

$$R^i: IF\left(x_i.is.close.to.ב_1^{i^*}\right) AND \dots AND\left(x_n.is.close.to.ב_n^{i^*}\right)$$

$$THEN\left(y^i = x_e^T \pi^i\right) \quad i = 1,2,\dots.,R \qquad (5.1)$$

Where R^i denotes the i^{th} fuzzy rule; R is the number of fuzzy rules; x_e is the extended input vector; $x_{e=}[1, x^T]^T$ which is formed by appending the input vector $x = [x_1, x_2, \dots, x_n]^T$; $x_j.is.close.to.x_j^{i^*}$ denotes the i^{th} fuzzy sets of the j^{th} fuzzy rule; $j = [1, n]$; $ב^{i^*}$ is the focal point of the i^{th} rule antecedent; $y^i = [y_1^i, y_2^i, \ldots, y_m^i]$ is the m-dimensional output of the i^{th}sub-system..

The exTS model output is calculated by weighted averaging of individual rules' contributions

$$y = \sum_{i=1}^{R} \tau^i\,(x) y^i = \sum_{i=1}^{R} \tau^i\,(x) x_e^T \pi^i \quad (5.2)$$

Where $\tau^i(x) = \frac{\mu^i(x)}{\sum_{j=1}^{R} \mu^j(x)}$ is the firing level of the i^{th} rule;

Although the recursive least squares algorithm is a powerful tool for adaptive filtering and predicting, when implemented in a finite precision environment it can suddenly become unstable due to quantization effects and divergence becomes a problem. This gives rise to separate effects like with the use of an exponentially forgetting factor which makes errors grow exponentially and the propagation of the error due to recursive use of the algorithm (Paleologu, Ciochin and Enescu, 2009).

An important characteristic of recursive least square (RLS) is the computation of the estimate of the inverse correlation matrix from the input data which helps the minimization process. It has some drawbacks related to the high computational complexity and large dynamic range of the algorithm's variables.

These issues are overcome by introducing fuzzy to the RLS algorithm to make it computationally fast and to fine tune the dynamic range of the variables in the algorithm. Compared to the least mean square (LMS) algorithm, the relevance of the least square algorithm is to minimize the sum of the square of the difference between the desired signal and the filter output (Paleologu, Ciochin and Enescu, 2009; Sankararaman et al., 2009). The RLS algorithm has a faster convergence speed (but not as compared to when fuzzy is introduced) and does not exhibit the eigenvalue spread problem.

A deterministic eXtended Takagi-Sugeno fuzzy recursive least square defect-propagation model is developed to estimate the size of defect and the rate of defect growth. The defect growth model is assumed to be similar to the Paris law (Bechhoefer, 2008; Sankararaman et al., 2009; Abdessalem et al., 2016).

$$\frac{dD}{dt} = C_0(\Delta D)^n \quad (5.3)$$

where ΔD is the instantaneous defect area and parameters C_0 and n are material constants which often vary with factors other than the instantaneous defect size (Li et al., 1999; Li, Kurfess and Liang, 2000). Figure 5.1 shows the exTSFRLSA system which estimates the size of the defect, and the rate of the defect growth.

The prognostic model compares the future bearing defect size to the measurement inferred bearing condition at time $t+\Delta t$. The adaptive fuzzy algorithm is used here to take advantage of the predictive error, thereby fine-tuning the model parameters together with the recursive least square algorithm hence improving the accuracy continuously in following the time varying defect growth behavior. The initial parameters α, β and t_0 which must be determined for the model can be obtained with the following equations:

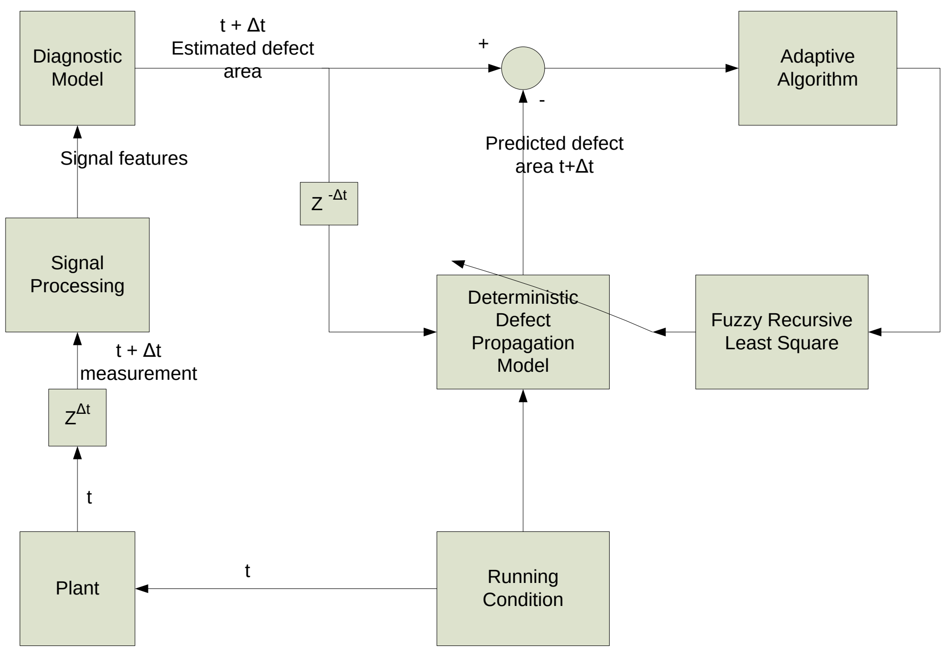

Figure 5.1, exTSFRLSA prognostic system (Li et al., 1999).

$$\alpha = \frac{1}{n-1}\log\left(\frac{C_0}{1-n}\right) \tag{5.4}$$

$$\beta = \frac{1}{1-n} \tag{5.5}$$

$$t_0 = \left(\frac{C_0}{1-n}\right) D_0^{n+1} \tag{5.6}$$

where $C_0 = 0.0702$ and $n = 0.6875$ (Li et al., 1999) are material dependent constants, D_0 is the smallest defect area that can be detected by a given diagnostic system and t_0 is the time for the smallest defect area to occur.

A recursive least squares (RLS) algorithm with a forgetting factor is often used to adaptively update the values of α, β and t_0 that need to be estimated in the model.

The output of the fuzzy network is,

$$Y(t) = \log(\Delta \mathrm{D}) \tag{5.7}$$

where $\Delta \mathrm{D}$ is the range of strain during a fatigue cycle.

$$Y(t, \vartheta(t-1)) \qquad \rightarrow estimate\, of\ Y(t)$$

The RLS algorithm is given as follows:

$$Y(t) = [\Psi^T(t-1)\vartheta(t-1) + \varepsilon(t)] \tag{5.8}$$

where

$$\Psi^T(t-1) = [y(t-1), y(t-2), ---y(t-n),$$
$$u(t-1-N), ---u(t-r-N), 1]$$
$$\vartheta^T(t-1) = [a_1, a_2, ---a_n, b_1, b_2, ---b_r, d]$$
$$\min_{\hat{\vartheta}} \sum [\underbrace{\Psi^T(i-1)\vartheta(i) - y(i)}_{\substack{\text{"least square"} \\ \text{prediction value of } y}}]^2 \tag{5.9}$$

where y is the output value, u is the input, d is the disturbance or noise and N is the time delay. The covariance matrix $p(t)$ which is a (3X3), is given by the following equation

$$P(t) = \lambda^{-1}\left(P(t-1) - \frac{P(t-1)\Psi(t)\Psi^T(t)P(t-1)}{\lambda + \Psi(t)P(t-1)\Psi(t)}\right) \tag{5.10}$$

The forgetting factor λ falls within the range of $0 < \lambda \leq 1$. The initial covariance matrix must be chosen as a unit matrix scaled by a positive scalar which is typically in the region of 1-1000 (Li et al., 1999).

$$\vartheta = [\alpha \;\; \beta \;\; t_0]^T \tag{5.11}$$

5.3 Experimental tests.

The same test rig set up which is found in figure 2.1 was used the second time. Three bearings were used in the experiment with two of the new bearings (of the same type) (Timken HR 30307 J) being subjected to artificially localized-defects since it can be dismantled from the outer raceway. The simulated crack was seeded on the outer raceway of the bearings with the use of a spark erosion machine which seeded a groove of width 1.5mm width by 1mm depth on the outer raceway of the taper bearings respectively. Figure 5.2 show a photograph of the seeded crack on the outer race of the test bearing.

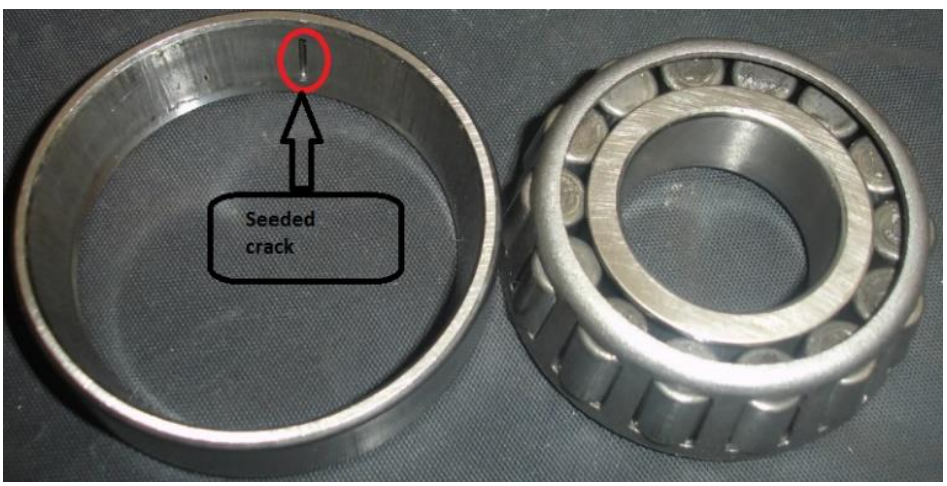

Figure 5.2, Crack on outer race introduced by spark erosion machine.

A brushless AC motor (Rockwell Automation MPL-B680B), mounted on a NSK 6309 single row bearing was used to drive the system. The angular velocity of the motor was retrieved from one of the analogue outputs available on the motor drive, (Rockwell Automation Kinetix 6000 series BM-01). This system allows continuous speed variation from 0 to 3600 rpm. A Soundwel AE sensor with model number SR 150M with a frequency range of 25-530 kHz was used.

All three bearings were loaded sinusoidally with 500N in the axial direction at a frequency of 2Hz and 900N in the radial direction at a frequency of 1Hz. The reason for applying the loads is that it helps us to simulate real life scenario where subsequent rotation of the bearing is slow in their application as it is here.

The speed of the servo motor was set at 70 and 90 rpm for the bearings. The vibration signatures for the test bearing was collected for all these speeds, using an FFT analyzer, a National Instruments data acquisition card (BNC-2110) with a shielded BNC connector block.

5.4 Discussion of result.

As discussed under the experimental procedure in section 5.3, to accelerate the defect propagation process, an initial crack was seeded on the outer race way of two bearings. An initial crack of 1.5 mm width by 1mm depth was initiated on the bearings and samples were collected from the bearings with the AE sensor at 30-minute intervals while the running was interrupted about every 15 hours to see if the bearings were damaged. It was observed by measurement with a venier caliper and a microscope that the crack has propagated to about a width of 1.6mm. After several random pits sprawled out after running for about 16 days each on both bearings the tests were stopped. Each of the damaged bearing data was used for the eXTSFRLSA analysis respectively.

Data 1 group was obtained from the test ran on the damaged bearing which was operated on a speed of 70 rpm while data 2 group set came from the test ran on the damaged bearing operated on a speed of 90 rpm.

Training using data 1 group took place at a learning rate of 0.5 and after the model was trained it was then used to validate the results obtained for data 2. Data 1 and 2 groups comprise of 52 data sets each of 60,000 samples. Data 1 group was then partitioned into 2 groups, one of 25 data sets which was used for training the algorithm, while the other group of 27 data sets was used to test and validate the learned algorithm. The incremental training changes the weights and biases of the fuzzy network after presentation of each individual input vector. This incremental training is often referred to as adaptive training or on-line training.

The process involves un-supervised learning where the weights and the biases are modified in response to the network input only. Data 2 set was also validated with respect to that obtained from the 27 data sets of data 1 and the result obtained was found to be valid as compared to that of data 1 set. A plot of the estimated defect propagation model parameters in the simulation of continuous growth of the defect area is given below in figure 5. Four simulations were performed starting with an initial estimate of $\alpha = -4.5$, $\beta = 3.2$ and $t_0 = 1.5$.

Table 5.1, initial estimation of defect propagation model parameters in the simulation of continuous growth of defect area.

	Simulation of continuous growth of defect area on data group 1 at 70 rpm		
Prediction no.	α_1	β_1	t_{01}
1	-4.5	3.2	3.6
2	-4.2	3.4	2.4
3	-5.0	2.5	2.0
4	-4.0	2.5	2.4
	Simulation of continuous growth of defect area on data group 2 at 90 rpm		
Prediction no.	α_2	β_2	t_{02}
1	-3.5	2.2	1.4
2	-3.5	2.2	1.5
3	-3.5	2.5	1.3
4	-2.0	2.2	1.3

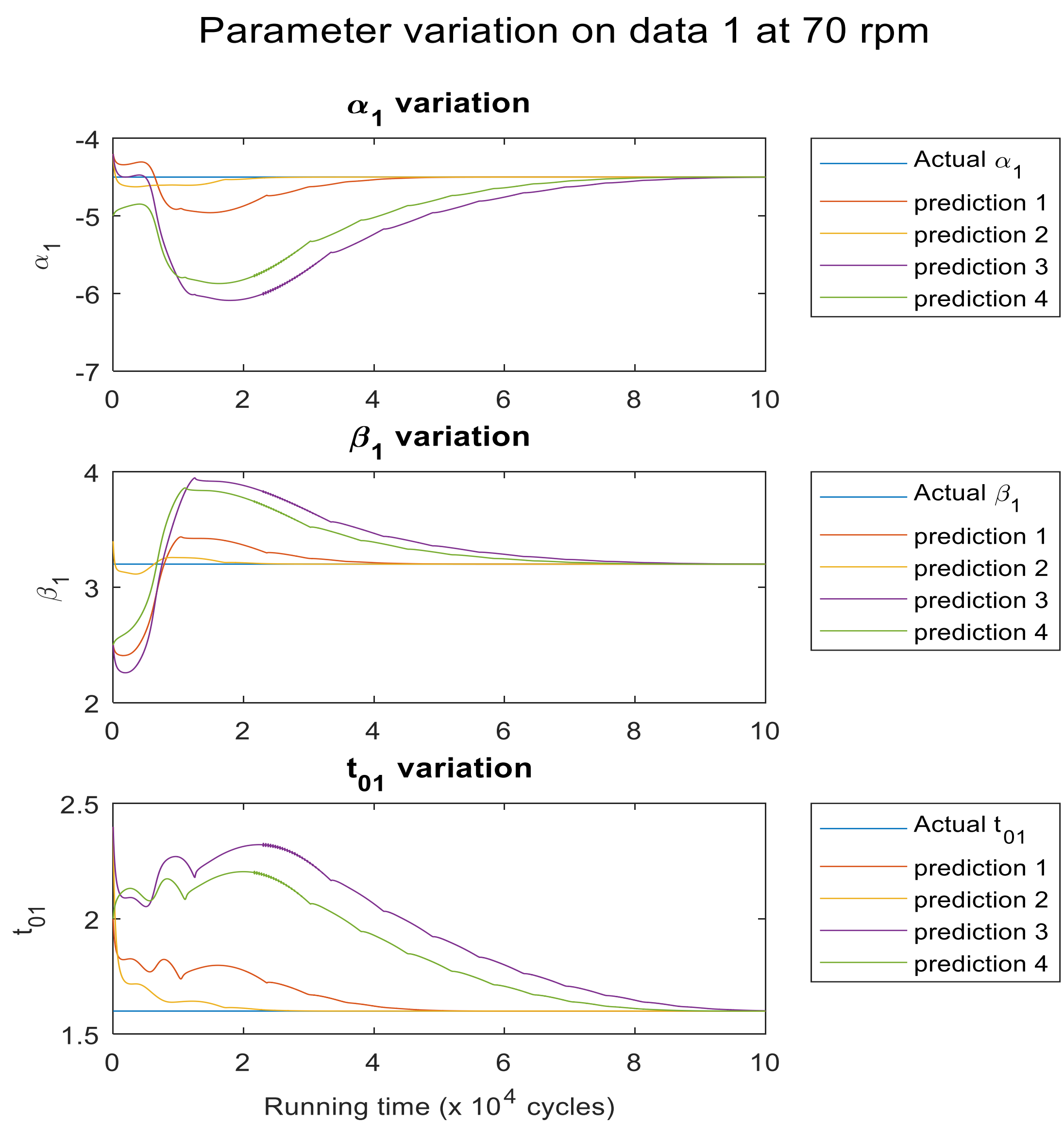

Figure 5.3, Estimated defect propagation model parameters for data 1 at 70 rpm using exTSFRLSA.

Figure 5.3 and 5.4 show the estimated defect propagation model parameters for both bearing 1 and bearing 2 running at speeds 70 and 90 rpm respectively. It was found that from the first simulation on data set 1, as the value of α_1 tends to be farther away from zero value and t_{01} becomes large there is a convergence experience, at around 10×10^4 cycles. But as α_2 tends to be closer to the zero line with t_{02} been smaller, as shown in figure 6 there is divergence of the parameter simulation from the actual value. A total time of 154.22 secs was taken to run each simulation to completion which was 81 % faster compared to the RLS simulation, because of the fuzzy network that was introduced to the recursive least square algorithms.

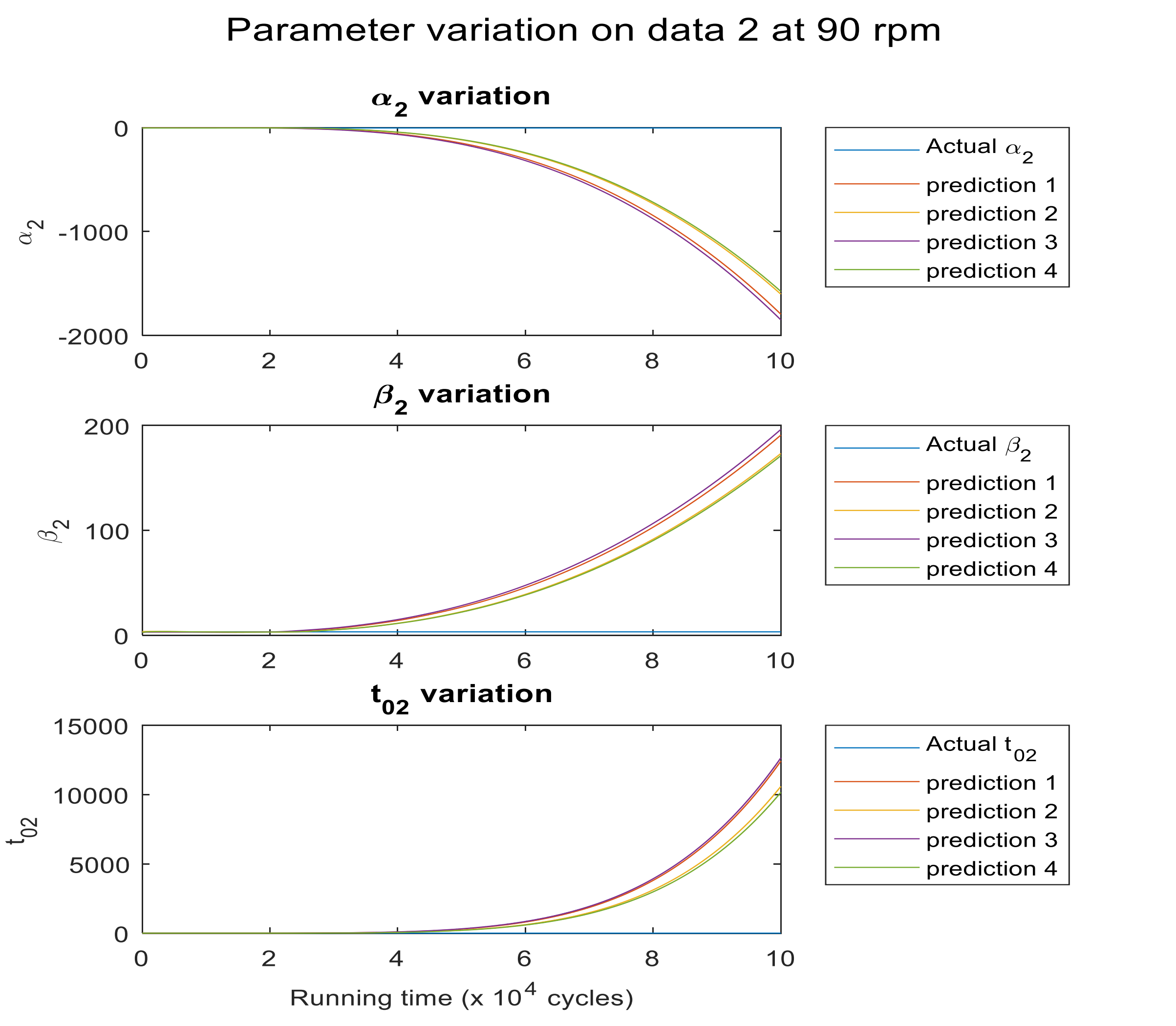

Figure 5.4, Estimated defect propagation model parameters for data 2 at 90 rpm using exTSFRLSA.

Figure 5.5 shows the plot of the prediction with adaptation at the crack growth. In accounting for the effect of disturbance which are inherent to the signal processing of in-direct measurements, a zero mean and 0.3 standard deviation was added to the normally distributed noise component. A forgetting factor λ of 0.99 was used in the simulation.

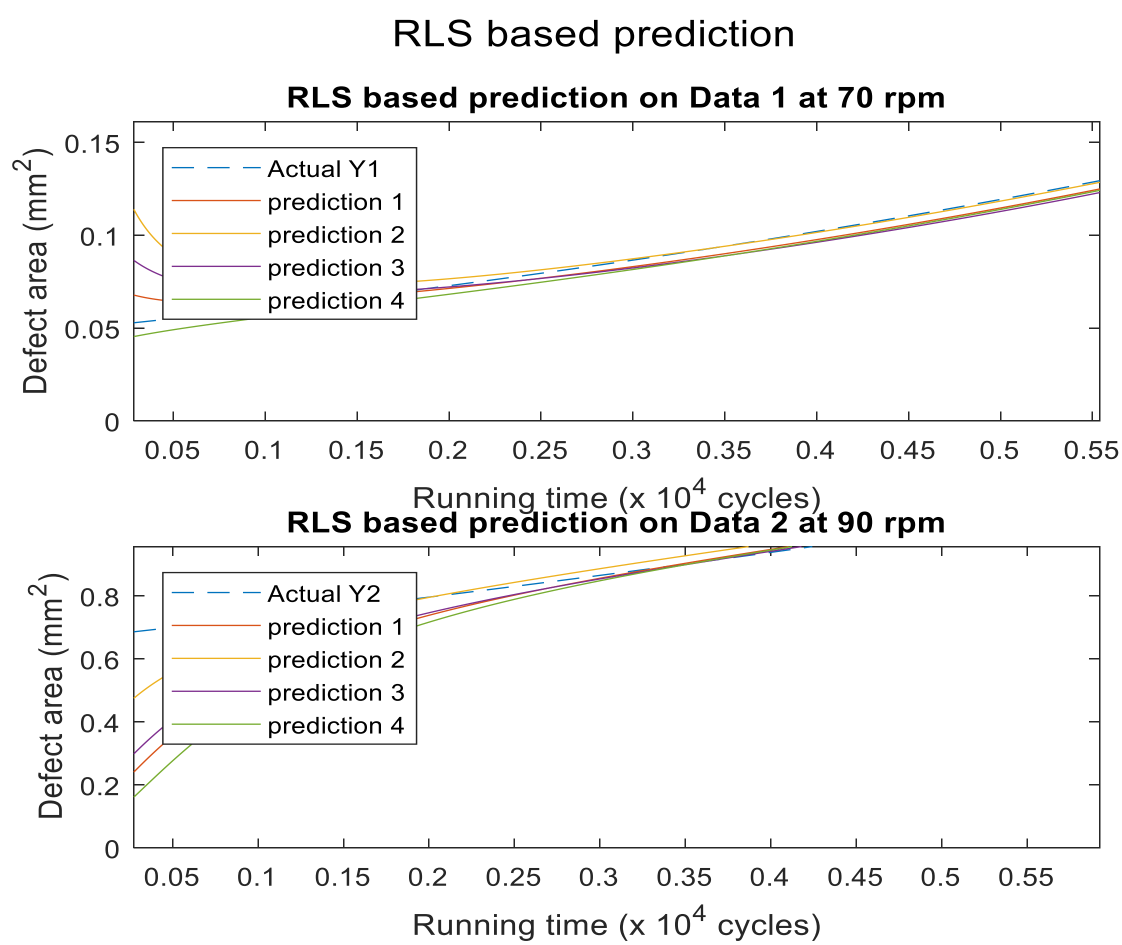

Figure 5.5, Plot of the prediction with adaptation at the crack growth using RLS.

After performing the same simulations to examine the feasibility of the eXTSFRLSA prognostic system, it was observed that the predictions plots did not deviate much from the actual plot of the two main bearings used as shown from figures 5.4 and 5.5. In the first set of simulations, a defect is assumed to grow continuously according to the deterministic propagation model of both equations 5.2 and 5.10. The equation 5.10 is however very important for use with equation 5.2 as it makes it possible for estimating many data without knowing a good equation representation. The eXTSFRLSA uses the unsupervised learning and emphasizes on the development of an evolving mechanism for rule-based evolution. We gradually update the centroid of the time series input based on the data. From Goguen's argument, any system satisfying certain axioms is equivalent to a system of fuzzy sets. Since axioms are intuitively plausible for system of all inexact concepts, the theorem allows us to conclude that inexact concepts can be represented by fuzzy sets.

EXTSFRLSA suitability for on-line identification means that for one to get a particular value at a particular time, it will depend on the previous value that has been obtained for example, if the system starts from time t0 one can use the integrator block such that the new value will depend on the previous, the time t2 will depend on the value of t1. Hence, fine tuning of the parameters makes the equation gradually becomes the right value expected.

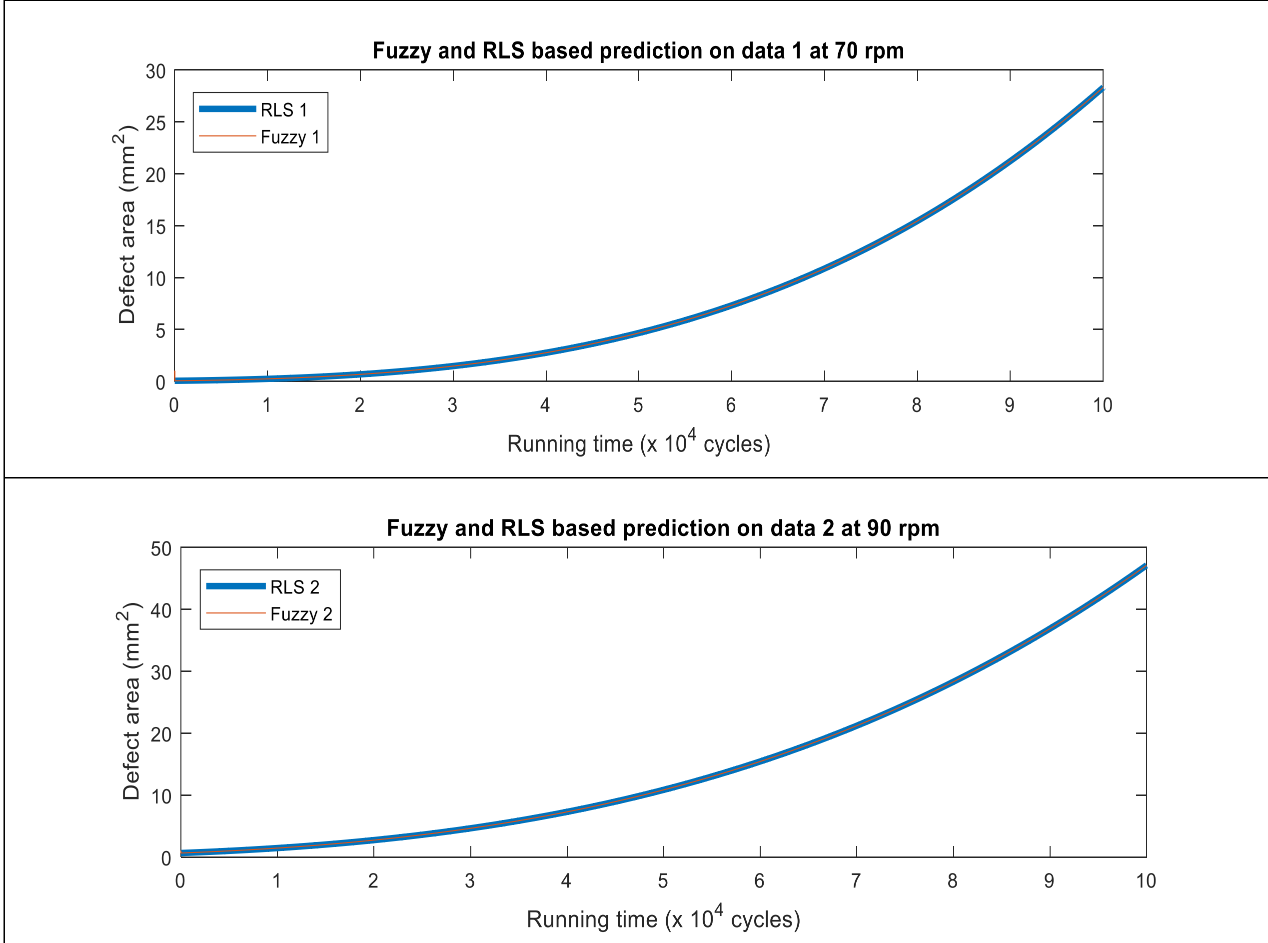

Figure 5.6, Plot comparison between RLS and eXTSFRLSA.

The growth defect area is compared for both the use of eXtended Takagi-Sugeno fuzzy and the RLS of the generic diagnostic model. The plots for the eXTSFRLSA shows just a single line being superimposed on that obtained from the data after using RLS thereby showing that eXTSFRLSA is well suitable for the prognostic of REB operated at varying speed and load conditions.

5.5 Summary.

This work presents an approach to online identification of the exTSFRLSA model for the health assessment of bearing. The eXtended Takagi-Sugeno fuzzy recursive least square algorithms (exTSFRLSA) is considered within the frame-work of condition-based maintenance and predictive maintenance. It is based on the recursive, noniterative building of the rule base by unsupervised learning. The concept introduced in this work made do with eXtended Takagi-Sugeno fuzzy and the recursive least square in combination with the famous Paris law which simulation work showed that it can effectively be used for RUL prediction for bearing fault prognosis.

exTSFRLSA on the other hand proves to be a good defect predictive system analyzer as it can respond before-hand to the bearing defect propagation process without a priori knowledge of the bearing under use before catastrophic failure occurs. It is a method that provide knowledge of the time-variant nature of defect growth in bearings by providing the best prediction possible.

The application of eXtended Takagi-Sugeno fuzzy to the algorithm made RLS algorithm which involves uncomplicated mathematical operations with more computational resources have more faster convergence speed and provides better fine-tuning of the algorithm parameters. It is computationally effective, as it does not require re-training of the whole model.

6.0 Chapter Five Conclusion, summary and contributions.

6.1 Conclusion and summary of contributions.

This research presented a data driven approach to condition monitoring based on statistical non-parametric feature extraction. Simulation and laboratory results showed the techniques are capable of classifying condition of non-steady loads and speeds of which such conditions are found in many industrial applications and include draglines in the mining industry and large rolling mills in many materials processing environments.

The objectives of the proposed research were presented in chapter 1 as the development of a diagnostic framework that:

- extracts diagnostic information from machine response signals at slow speed, by using AE signals to detect damage in low speed rotating bearings, thereby offering high sensitivity.
- extracts diagnostic information in such a manner that it is sensitive to the presence of machine faults, yet also robust in time-varying operating conditions,
- is not dependent on fault historic data, destructive tests, or extensive manual preparation of training data,
- allows for a simple intuitive representation of the extracted information, such that a non-expert, or a simple classification algorithm may interpret it.

My goals are to effectively diagnose faults in rolling element bearings at incipient stage which are operated under varying loads and speeds condition using higher order statistics and to find the best algorithm that could be used to train data obtained under this condition at short training time thereby making it suitable for online application. Also, to establish a convenient prognostic algorithm which can use less data compared to fuzzy (known for its large data application).

In the second chapter, HOS was combined with KL divergence and Lempel-Ziv complexity to formulate an indicator (energy coherent factor) that made use of kurtosis and hyper-flatness as these were used to successfully extract information from signals which deviated from Gaussianity making it easier to detect and quantify non-linearities in time series especially at low speed.

In chapter 3, two training methods were also explored in the laboratory demonstration. In the first method, training was done separately for each condition so that Bayesian decision boundaries were tailored for each condition. Training the model on data obtained directly from the experiment was performed successfully and this helps to ensure the validity of the model. Though HMM is a very robust method for fault classification, literature records the disadvantage of the use of HMM which makes it unfit for on-line use (due to the involvement of large data sets to train the model for fault classification). This was overcome here by using the BRNHMM combined with the selection of the output symbol vector by the Gaussian density of continuous HMM and this has become a successful tool for process monitoring and fault detection in roller bearing operating in a complex scenario like as found in rolling mills.

By using a GA-based feature extractor, statistical parameters were extracted from raw acoustic signal and used to classify the inputs for the SVM for regression in chapter 3. These features extracted cleared the doubts in obtaining intensity from the time spectrum. Thus, the features extracted proved to be good indicators of defect intensity and it also showed that with the Gaussian kernel function, effective classification can be achieved over exponential kernel function with SVMGA as it classified correctly in all conditional cases of the bearing faults with all the data set generated in this experimental work for which just a few sample of data result was showed due to space limitation.

SVMGA proves to be a better classifier in chapter 4 which maximizes the fault classification accuracy as it could identify faults and classify better with the Gaussian kernel function than any other kernel function like the polynomial, the rbf, and the spline with the sigmoid kernel function.

The eXtended Takagi-Sugeno fuzzy recursive least square algorithms (exTSFRLSA) is considered within this frame-work with the concept which combined eXtended Takagi-Sugeno fuzzy and the recursive least square in combination with the famous Paris law for RUL prediction for bearing fault prognosis.

All this has been a great significant improvement over past research which cited the non-steady operation as a reason for missed detections.

6.2 Suggestions for further research.

While this method is a viable option for machine fault detection in industrious environments, some improvements are needed. The main advantage in adapting other anomaly discovery environments is the ability to model frequency location dependency and discern fault types. This analysis may include the use of other intelligent anomaly discovery techniques, such as hierarchal statistics.

The study in this work was basically on the Timken tapered roller bearing HR 30307 J. It is expected that further studies be carried out on other bearing types. It is also expected that much larger bearing datasets be involved for further research, likewise much broader range of speed and load conditions should be used especially lower speeds in the range of 10 rpm to 60 rpm. There should be better distinction between training and other data.

References.

Abdessalem, A. B. *et al.* (2016) 'Stochastic modelling and prediction of fatigue crack propagation using piecewise-deterministic Markov processes', *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, Part O.

Aherwar A. and Khalid S. (2012) 'Vibration analysis techniques for gearbox diagnostic : a review', *International Journal of Advanced Engineering Technology*, III(II), pp. 04–12.

Amigó, J. M. *et al.* (2004) 'Estimating the entropy rate of spike trains via Lempel-Ziv complexity', *Neural Computation*, 16(4), pp. 717–736.

Andrade F.A., E. I. and B. M. N. M. (2001) 'A new approach to time-domain vibration condition monitoring: gear tooth fatigue crack detection and identification by the Kolmogorov-Smirnov test', *Journal of Sound and Vibration*, 240(5), pp. 909–919.

Angelov, P. and Filev, D. (2005) 'Simpl _ e TS : a simplified method for learning evolving Takagi-Sugeno fuzzy models', pp. 1068–1073.

Angelov, P. P., Filev, D. P. and Member, S. (2004) 'An approach to online identification of Takagi-Sugeno fuzzy models', 34(1), pp. 484–498.

Angelov, P. and Zhou, X. (2006) 'Evolving fuzzy systems from data streams in real-time', in *2006 International Symposium on Evolving Fuzzy Systems, September, 2006*, pp. 29–35.

Angelov, P. and Zhou, X. (2008) 'On line learning fuzzy rule-based system structure from data streams', *IEEE*, 44(1524), pp. 915–922.

Antoni, Ã. (2006) 'The spectral kurtosis : a useful tool for characterising non-stationary signals', *Mechanical Systems and Signal Processing*, 20(2), pp. 282–307.

Antoni, J. and Randall, R. B. (2002) 'Differential diagnosis of gear and bearing faults', *Journal of Vibration and Acoustics*, 124, pp. 165–171.

Aye S.A (2014) *Acoustic emission-based diagnostics and prognostics of slow rotating bearings using Bayesian techniques*. University of Pretoria, South Africa.

El Badaoui, M. *et al.* (2001) 'Use of the moving cepstrum integral to detect and localize tooth spalls in gears', *Mechanical Systems and Signal Processing*, 15(5), pp. 873–885.

Barszcz, T. and Randall, R. B. (2009) 'Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine', *Mechanical Systems and Signal Processing*, 23(4), pp. 1352–1365.

Bartelmus, W. (2003) 'Diagnostic information on gearbox condition for mechatronic systems', *Transactions of the Institute of Measurement and Control*, 25(5), pp. 451–465.

Baruah, P. and Chinnam, R. B. (2005) 'HMMs for diagnostics and prognostics in machining processes', *International journal of Production Research*, 43(6), pp. 1275–1293.

Baydar, N. and Ball, A. (2000) 'Detection of gear deterioration under varying load conditions by using the instantaneous power spectrum', *Mechanical Systems and Signal Processing*, 14(6), pp. 907–921.

Bechhoefer, E. (2008) 'A method for generalized prognostics of a component using Paris law', in *American Helicopter Society 64th Annual Forum, Montreal, CA*, pp. 1–11.

Bechhoefer E. and M., K. (2009) 'A review of time synchronous average algorithms', in *Annual Conference of the Prognostics and Health Management Society, 2009*, pp. 1–10.

Behzad, M. and B. A. R. (2011) 'A new method for detection of rolling bearing faults based on the Local Curve Roughness approach', *Polish Maritime Research*, 18(2(69)), pp. 44–50.

Benkedjouh, T. *et al.* (2012) 'Fault prognostic of bearings by using support vector data description', in *IEEE Conference on Prognostics and health Management*, pp. 1–7.

Bolander, N. *et al.* (2009) 'Physics-based remaining useful life prediction for aircraft engine bearing prognosis', in *Annual Conference of the Prognostics and Health Management Society, 2009*, pp. 1–12.

Bondu, A. and Grossin, B. (2010) 'Density estimation on data streams : an application to change detection summarization of the input data stream', *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on 5:3141 - 3144*, pp. 1–12.

Caesarendra, W. *et al.* (2016) 'Acoustic emission-based condition monitoring methods : review and application for low speed slew bearing', *Mechanical Systems and Signal Processing*. Elsevier, 72–73, pp. 134–159.

Costache, M., Liénou, M. and Datcu, M. (2006) 'On Bayesian inference , maximum entropy and support vector machines methods', in *On Bayesian Inference,MaximumEntropy and SupportVectorMachinesMethods19th September*, pp. 20061–20069.

Derouiche, Z. *et al.* (2012) 'Application of neural networks for monitoring mechanical defects of rotating machines', *Journal of Energy and Power Engineering*, 6, pp. 276–282.

Dorj, E. and Chen, C. (2013) 'A Bayesian hidden Markov model-based approach for anomaly detection in electronic systems', *IEEE*, pp. 1–10.

Dube, A. V., Dhamande, L. S. and Kulkarni, P. G. (2013) 'Vibration based condition assessment of rolling element bearings with localized defects', *International Journal of Science & Technology Research*, 2(4), pp. 149–155.

El-Koujok, M., Gouriveau, R. and Zerhouni, N. (2009) *Error estimation of a neuro-fuzzy predictor for prognostic purpose*, *IFAC Proceedings Volumes*. IFAC.

Fatima, S., Mohanty, A. R. and Kazmi, H. F. (2016) 'Fault classification and detection in a rotor bearing rig', *Journal of Vibration Engineering and Technology*, 4(6), p. 2016.

Fernando, P.-C. (2008) 'Kullback-leibler divergence estimation of continuous distributions', *IEEE International Symposium on Information Theory - Proceedings*, pp. 1666–1670.

Gebraeel, N. *et al.* (2004) 'Residual life predictions from vibration-based degradation signals : a neural network approach', *IEEE Transactions on Industrial Electronics*, 51(3), pp. 694–700.

Gebraeel, N., Elwany, A. and Pan, J. (2009) 'Residual life predictions in the absence of prior degradation knowledge', *IEEE Transactions on Reliability*, 58(1), pp. 106–117.

Gebraeel, N., Lawley, M. and Ryan, J. (2005) 'Residual life distributions from component degradation signals : a Bayesian approach residual-life distributions from component degradation', *IIE Transactions*, 37, pp. 543–557.

Gebraeel, N. Z. *et al.* (2005) 'Residual life distributions from component degradation signals : a Bayesian approach residual-life distributions from component degradation', *IIE Transactions*, 37, pp. 543–557.

Gelle G. and Delaunay, C. M. and (2000) 'Blind source separation applied to rotating machines monitoring by acoustical and vibration analysis', *Mechanical Systems and Signal Processing*, 14(3), pp. 427–442.

Geramifard, O., Xu, J. and Kumar, S. (2013) 'Fault detection and diagnosis in synchronous motors using hidden Markov model-based semi-nonparametric approach', *Engineering Applications of Artificial Intelligence*. Elsevier, 26(8), pp. 1919–1929.

Ghahramanj Zoubin (2001) 'An introduction to hidden Markov models and Bayesian networks', *International Journal of Pattern Recognition and Artificial Intelligence*, 1(15), pp. 9–42.

Goodlin, B. E. *et al.* (2002) 'Simultanous fault detection and classification for semiconductor manufacturing tools', in *201st Meeting of the Electrochemical Society, International Symposium on Plasma Processing XIV*. Abs. 413 Philadelphia, PA, pp. 1–10.

Graney, B. P. and Starry, K. (2012) *Rolling element bearing analysis*.

Guerlain, S., Brown, D. . and Mastrangelo, C. (2000) 'Intelligent decision support systems', *IEEE*, pp. 1934–1938.

Gunn, S. R. (1998) *Support vector machines for classification and regression*.

Guo, H., Jack, L. B. and Nandi, A. K. (2005) 'Feature generation using genetic programming with application to fault classification', *IEEE Transactions on Systems, Man, And Cybernetics-Part B: Cybernetics*, 35(1), pp. 89–99.

Hagan, Beale M. T, H. M. (1997) *Neural network design*. Thomson.

Hamadache, M. and Lee, D. (2014) 'Improving signal-to-noise ratio (SNR) for inchoate fault detection based on principal component analysis (PCA)', in *14th International Conference on Control, Automation and Systems (ICCAS 2014)*. Kintex, Gyeonggi-do, Korea, pp. 561–566.

Hariharan, V. and Srinivasan, P. S. S. (2009) 'New approach of classification of rolling element bearing fault using artificial neural network', *Journal of Mechanical Engineering*, ME 40(2), pp. 119–130.

Heng A. *et al.* (2009) 'Rotating machinery prognostics : state of the art , challenges and opportunities', *Mechanical Systems and Signal Processing*, 23(3), pp. 724–739.

Heyns, T. (2013) *Low cost condition monitoring under time-varying operating conditions*. University of Pretoria, South Africa.

Hong, H. and Liang, M. (2009) 'Fault severity assessment for rolling element bearings using the Lempel-Ziv complexity and continuous wavelet transform', *Journal of Sound and Vibration*, 320(1–2), pp. 452–468.

Huang, R. *et al.* (2007) 'Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods', *IEEE Transactions on Reliability*, 21(1), pp. 193–207.

Ilhem, B., Amar, B. and Lebaroud, A. (2014) 'Classification method for faults diagnosis in reluctance motors using hidden Markov models', in *2014 IEEE 23rd International Symposium on Industrial Electronics (ISIE)*, pp. 984–991.

Jamaludin, N., Mba, D. and Bannister, R. H. (2001) 'Condition monitoring of slow-speed rolling element bearing using stress waves', *Journal of Process Mechanical Engineering*, 215(4), pp. 245–271.

Jammu, N. S. and Kankar, P. K. (2011) 'A review on prognosis of rolling element bearings', *International Journal of Engineering Science and Technology (IJEST)*, 3(10), pp. 7497–7503.

Jardine, K. S. A., Lin, D. and Banjevic, D. (2006) 'A review on machinery diagnostics and prognostics implementing condition-based maintenance', *Mechanical Systems and Signal Processing*, 20(7), pp. 1483–1510.

Karacay T. and Akturk, N. (2009) 'Experimental diagnostics of ball bearings using statistical and spectral methods', *Tribology International*, 42, pp. 836–843.

Kim, Y., Tan, A. C. C. and Yang, B. (2006) 'Condition monitoring of low speed bearings : a comparative study of the ultrasound technique versus vibration measurements', in *WCEAM*, pp. 22–25.

Li, B. *et al.* (2000) 'Neural-network-based motor rolling fault diagnosis', *IEEE Transactions on Industrial Electronics*, 47(5), pp. 1060–1069.

Li, P., Jiang, Y. and Xiang, J. (2014) 'Experimental investigation for fault diagnosis based on a hybrid approach using wavelet packet and support vector classification', *The Scientific World Journal*. Hindawi Publishing Corporation, 2014. doi: ://dx.doi.org/10.1155/2014/145807.

Li, X. and Chen, W. (2014) 'Rolling bearing fault diagnosis based on physical model and one-class support vector machine', *ISRN Mechanical Engineering*, 2014, pp. 1–4.

Li, Y. *et al.* (1999) 'Adaptive prognostics for rolling element bearing condition', *Mechanical Systems and Signal Processing*, 13(1), pp. 103–113.

Li, Y., Kurfess, T. R. and Liang, S. Y. (2000) 'Stochastic prognostics for rolling element bearing', *Mechanical Systems and Signal Processing*, 14(5), pp. 747–762.

Li, Z. *et al.* (2005) 'Hidden Markov model-based fault diagnostics method in speed-up and speed-down process for rotating machinery', *Mechanical Systems and Signal Processing*, 19, pp. 329–339.

Lin, C. *et al.* (2011) 'Neural-network-based robust adaptive control for a class of nonlinear systems', *Neural Comput & Applic*, 20, pp. 557–563.

Lu, S. (2007) 'Bearing fault diagnosis based on K-L transform and support vector machine', in *Third International Conference on Natural Computation (ICNC 2007).*

Luo, J., M., N. and Pattipati K., Kawamoto M., and C. S. (2003) 'Model-based prognostic techniques', in *Proceedings Autotestcon IEEE System Readiness Technology Conference*, pp. 330–340.

Mahamad, A. K. Bin (2010) *Diagnosis , Classification and Prognosis of Rotating Machine using Artificial Intelligence*. Kumamoto University Japan.

Malhi, A. and Gao, R. X. (2004) 'PCA-based feature selection scheme for machine defect classification', *IEEE Transactions on Instrumentation and Measurement*, 53(6), pp. 1517–1525.

Marton, I. *et al.* (2013) 'Application of data driven methods for condition monitoring maintenance', *Chemical Engineering Transactions*, 33, pp. 301–306.

Mba, D. (2003) 'Acoustic emissions and monitoring bearing health', *Tribology Transactions*, 46(3), pp. 447–451.

Mcinerny, S. A. and Dai, Y. (2003) 'Basic vibration signal processing for bearing fault detection', *IEEE Transactions on Education*, 46(1), pp. 149–156.

Medjaher, K., Tobon-mejia, D. A. and Zerhouni, N. (2012) 'Remaining useful life estimation of critical components with application to bearings', *IEEE Transactions on Reliability*, 61(2), pp. 292–302.

Moreno-Munoz, A. *et al.* (2007) 'Higher-order spectral characterization of termite Higher-order spectral characterization of termite emissions using acoustic emission probes', in *SAS 2007 - IEEE Sensors Applications Symposium San Diego, California USA, 6-8 February 2007*, pp. 1–6.

Moura, J. M. F. (2009) 'What is signal processing ?', *president's Message*, 6(November), p. 2009.

Nayak, S. C., Misra, B. B. and Behera, H. S. (2014) 'Impact of data normalization on stock index forecasting', *International Journal of Computer Information Systems and Industrial Management Applications*, 6, pp. 257–269.

Nelwamondo, F. V., Marwala, T. and Mahola, U. (2005) 'Early classifications of bearing faults using hidden Markov models, Gaussian mixture models, mel-frequency ceptral coefficients and fractals', *International Journal of Innovative Computing, Information and Control*, x(0x,x 2005), pp. 1–19.

Niu X., Z. L. and D. H. (2005) 'New statistical moments for the detection of defects in rolling element bearings', *Int J Adv Manuf Technol*, 26, pp. 1268–1274.

Ocak, H. and Loparo, K. A. (2001) 'A new bearing fault detection and diagnosis scheme based on hidden markov modeling of vibration signals', in *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on 5:3141 - 3144*, pp. 1–4.

Paleologu, C., Ciochin, S. and Enescu, A. A. (2009) 'A family of recursive least-square adaptive algorithms suitable for fixed-point implementation', *International Journal on Advanves in Telecommunication*, 2(2), pp. 88–97.

Peng, Z. K. and Chu, F. L. (2004) 'Application of the wavelet transform in machine condition monitoring and fault diagnostics : a review with bibliography', *Mechanical Systems and Signal Processing*, 18(2), pp. 199–221.

Petropulu, A. P. (2000) 'Higher-order spectral analysis', in. CRC Press LLC.

Plamen Angelov, José Victor, A. D. and D. F. (2004) 'On-line evolution of Takagi-Gugeno fuzzy models', *IEEE*, pp. 1–8.

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

Rabiner, L. R. (2009) 'Tutorial on hidden Markov models and selected applications in speech recognition', *Proceedings of the IEEE*, 77(2), pp. 257–286.

Rafiee, J. *et al.* (2009) 'A novel technique for selecting mother wavelet function using an intelligent fault diagnosis system', *Expert Systems With Applications*. Elsevier Ltd, 36(3), pp. 4862–4875.

Randall, R. B. and Antoni, J. (2011) 'Rolling element bearing diagnostics — A tutorial', *Mechanical Systems and Signal Processing*, 25, pp. 485–520.

Rao, S. S. and Horton, M. J. (2011) *Mechanical vibrations Fifth Edition*.

Sankararaman, S. *et al.* (2009) 'Stochastic modelling and prediction of fatigue crack propagation using piecewise-deterministic Markov processes', in *Annual Conference of the Prognostics and Health Management Society*, pp. 1–13.

Shen, C. *et al.* (2014) 'Recognition of rolling bearing fault patterns and sizes based on two-layer support vector regression machines', *Smart Structures and Systems*, 13(3). doi: http://dx.doi.org/10.12989/sss.2014.13.3.000 000 Recognition.

Shen, C. *et al.* (2014) 'Recognition of rolling bearing fault patterns and sizes based on two-layer support vector regression machines', *Smart Structures and Systems*, 13(3), pp. 191–200.

Sikorska, J. Z., Hodkiewicz, M. and Ma, L. (2011) 'Prognostic modelling options for remaining useful life estimation by industry', *Mechanical Systems and Signal Processing*, 25, pp. 1803–1836.

Sloukia, F. E., Bouarfa, R. and Medromi, H. (2013) 'Bearings pronostic using mixture of Gaussians Hidden Markov Model and support vector machine', *International Journal of Network Security & Its Applications (IJNSA)*, 5(3), pp. 85–97.

Smola, A. J. and Schollkopf, B. (2003) *A tutorial on support vector regression*.

Soualhi, A. *et al.* (2012) 'Fault detection and diagnosis of induction motors based on hidden Markov model', in *Electrical Machines (ICEM), 2012 XXth International Conference on IEEE*, pp. 1693–1699.

Stander, C. J. (2005) *Condition monitoring of gearboxes operating under fluctuating load conditions*. University of Pretoria.

Stander, C. J. Ã. and Heyns, P. S. (2005) 'Instantaneous angular speed monitoring of gearboxes under non-cyclic stationary load conditions', *Mechanical Systems and Signal Processing*, 19(4), pp. 817–835.

Stander, C. J. Ã. and Heyns, P. S. (2006) 'Transmission path phase compensation for gear monitoring under fluctuating load conditions', *Mechanical Systems and Signal Processing*, 20(7), pp. 1511–1522.

Tandon, N. and Choudhury, A. (1999) 'A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings', *Tribology International*, 32(1999), pp. 469–480.

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.

Tong, Q. *et al.* (2017) 'A fault diagnosis approach for rolling element bearingsbased on RSGWPT-LCD bilayer screening and extreme learning machine', *IEEE ACCESS*, 5, pp. 5515–5530.

Ulus S. and Erkaya S (2016) 'An experimental study on gear diagnosis by using acoustic emission technique', *International Journal of Acoustics and Vibration*, 21(1), pp. 103–111.

Unal, M. *et al.* (2013) 'Fault diagnosis of rolling bearing based on feature extraction and neural network algorithm 2 experimental setup and test', *Recent Advances in Telecommunications, Signals and Systems*, pp. 179–185.

Unal M. *et al.* (2013) 'Fault diagnosis of rolling bearing based on feature extraction and neural network algorithm', *Recent Advances in Telecommunications, Signals and Systems Fault*, pp. 179–185.

Vinson, R. G. (2014) *Rotating machine diagnosis using smart feature selection under non-stationary operating conditions*. University of Pretoria, South Africa.

Wang, H. and Chen, P. (2009) 'Fault diagnosis method based on kurtosis wave and information divergence for rolling element bearings 2 definition of kurtosis wave', *WSEAS Transactions on systems*, 8(10), pp. 1155–1165.

Wang, W. (2007) 'A two-stage prognosis model in condition based maintenance', *European Journal of Operational Research*, 182, pp. 1177–1187.

Welling, M. (2005) *Robust higher order statistics*.

Widodo, A. *et al.* (2009) 'Expert systems with applications fault diagnosis of low speed bearing based on relevance vector machine and support vector machine', *Expert Systems with Applications*. Elsevier Ltd, 36(3), pp. 7252–7261.

Williams T. (2001) 'Rolling element bearing diagnostics in run-to-failure life time testing', *Mechanical Systems and Signal Processing*, 15(5), pp. 979–993.

Xi F., S. Q. and K. G. (2000) 'Bearing diagnostics based on pattern recognition of statistical parameters', *Journal of Vibration and Control*, 6, pp. 375–392.

Xiang, J. and Zhong, Y. (2016) 'A novel personalized diagnosis methodology using numerical simulation and an intelligent method to detect faults in a shaft', *Applied Sciences*, 6(12), p. 414.

Xiangyang, L. and Wanqiang, C. (2014) 'Rolling bearing fault diagnosis based on physical model and one-class support vector machine', *ISRN Mechanical Engineering*, 2014, pp. 1–4.

Xie, K. (2011) *Support vector machine, concept and matlab build*.

Xinwen Niu · Limin Zhu · Han Ding (2005) 'New statistical moments for the detection of defects in rolling element bearings', *International Journal of Advance Manufacturing Technology*, 26, pp. 1268–1274.

Y., J. (2011) 'Bearing performance degradation assessment using locality preserving projections and Gaussian mixture models', *Mechanical Systems and Signal Processing*. Elsevier, 25(7), pp. 2573–2588.

Yam, R. C. M. *et al.* (2001) 'Intelligent predictive decision support system for condition-based maintenance', *Int J Adv Manuf Technol*, 17, pp. 383–391.

Yan, R. and Gao, R. X. (2004) 'Complexity as a measure for machine health evaluation', *IEEE Transactions on Instrumentation and Measurement*, 53(4), pp. 1327–1334.

Yang, B. S., Han, T. and Hwang, W. W. (2005) 'Fault diagnosis of rotating machinery based on multi-class support vector machines', *846 Journal of Mechanical Science and Technology (KSME Int. J.)*, 19(3), pp. 846–859.

Yang, H., Mathew, J. and Ma, L. (2003) 'Vibration feature extraction techniques for fault diagnosis of rotating machinery - a literature survey', in *Asia- Pacific Vibration Conference*. Gold Coast, Australia, pp. 12–14.

Yang, Y., Yu, D. and Cheng, J. (2007) 'A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM', *Measurement*, 40, pp. 943–950.

Yen, G. Y. and Lin, Y. C. (1999) 'Wavelet packet feature extraction for vibration monitoring', in *Proceedings of the IEEE international Conference an Control Applications*. Kohala Coast-Island of Hawai'i, Hawai'i, USA, pp. 1573–1578.

Zaeri, R. *et al.* (2011) 'Artificial neural network based fault diagnostics of rolling element bearings using continuous wavelet transform', *IEEE*, 11(3), pp. 753–758.

Zhang, B. *et al.* (2010) 'Fault progression modeling: an application to bearing diagnosis and prognosis', in *American Control Conference Marriott Waterfront, Baltimore*, pp. 6993–6998.