Fault detection in roller bearing operating at low speed and varying loads using Bayesian Robust New Hidden Markov Model

Henry Ogbemudia Omoregbee* and P. Stephan Heyns

Centre for Asset Integrity Management
Department of Mechanical and Aeronautical Engineering
University of Pretoria, Pretoria, South Africa.

Abstract

This paper uses Bayesian robust new hidden Markov modeling (BRNHMM) for bearing fault detection and diagnosis based on its acoustic emission signal. A variational Bayesian approach is used that simultaneously approximates the distribution over the hidden states and parameters with simpler distribution hence using Bayesian inference for the estimation of the posterior HMM hyperparameters. This allows for online detection as small data sets can be used. Also, the Kullback-Leibler (KL) divergence is effectively used to access the divergence of the probability function of the BRNHMM, to find its lower bound approximation and by applying a linear transform to the maximum output probability parameter generation (MOPPG). The training set result obtained from BRNHMM is then compared to the result from artificial neural network (ANN) fault detection for same complex system of low speed and varying load conditions which are difficult from a diagnostic perspective, as found in rolling mills.

Keywords: Acoustic emission; Artificial neural networks (ANN); Bayesian robust new hidden Markov Modeling (BRNHMM); Kullback-Leibler (KL) divergence; Maximum output probability parameter generation (MOPPG); Probabilistic neural network (PNN); RBF network

1. Introduction

Monitoring the health condition of low speed and heavy-duty mechanical systems and the detection of damage progression from a very early stage in the life cycle is of great importance in many industries. To fully inspect the health conditions of these mechanical systems, condition monitoring systems are used to collect real-time data from the systems and therefore massive amount of data are acquired after a prolonged period of operation. These massive amounts of data are often collected and analyzed off-line [1, 2].

Fault diagnosis of rolling element bearings using vibration or acoustic signature analysis is the most commonly used to prevent breakdowns in machinery. 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 [3-5]. 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 [6, 7].

*Corresponding author. Tel.: +27748329194,
E-mail address: omoregbeeho@gmail.com

Traditional techniques such as the Fast Fourier Transform (FFT) which finds application in the frequency domain analysis can convert signals into the frequency domain, but they do not provide both time and frequency information simultaneously. Later studies showed that time-frequency analysis methods such as wavelet transform can be used to detect faults and can be applied to non-linear and non-stationary signal processing. However, the wavelet transform has some deficiencies which include, border distortion, energy-leakage and interference terms [8]. Cheng et al. [9] used Hilbert-Huang Transform (HHT) and Self Organizing Feature Map (SOM) to overcome these deficiencies in his gear fault identification. The HHT, which includes Empirical Mode Decomposition (EMD) and Hilbert transform, comprise of local characteristic time scale of a signal which could decompose the complicated signal into many Intrinsic Mode Functions (IMF). This could not diagnose the gear fault accurately so, the SOM was introduced for better enhanced performance. SOM is an unsupervised neural network learning technique with self-adaptive and self-learning features. It has shown that it can be used to transform arbitrarily high-dimensional input data into a low-dimensional map. However, there is an increase in computational cost [10-12].

Dual-Tree Complex Wavelet Transform (DT-CWT) threshold denoising and Laplacian Eigenmaps (LE) was used by Chen et al. [13, 14] for planetary gear fault identification. It involves the original high-dimensional feature set, comprising the time domain features, frequency domain features, fractal box dimension of the denoised signal, and permutation entropy being constructed from multi-domain. DT-CWT threshold denoising method performs better in signal denoising of planetary gears. Nevertheless, it is difficult to obtain accurate diagnostic information from the denoised signal and the computation is too cumbersome. The introduction of LE as observed in Ref. [13] reduces the dimension of the original feature set so that low-dimensional sensitive features can be obtained.

Chen et al. [15] also proposed a weak feature information extraction method of planetary gear based on Ensemble Empirical Mode Decomposition (EEMD) and Adaptive Stochastic Resonance

(ASR). In their method, the original signal was decomposed into IMF with small modal aliasing by EEMD. The aspect of the ASR is composed of Particle Swarm Optimization (PSO) and Stochastic Resonance (SR). The weak fault feature information can be extracted from the output signal of the ASR system as soon as the signal reconstructed by effective IMFs is inputted. The PSO algorithm is used to optimize the critical parameters of SR. Usually the SR method is limited by adiabatic approximation theory and linear response theory which can only deal with small parameter signals.

Pandya et al. [16] used a method for an automation of fault diagnosis of ball bearings having localized defects (spalls) on the various bearing components. The system he used uses the wavelet packet decomposition which made do with the 'rbio5.5' real mother wavelet function for feature extraction from the vibration signal. Although this method helped to alleviate the time-invariant characteristic of the wavelet packet coefficient that most wavelet packet exhibit, its classification output yield rate still needs improvement as it was just a-little above 90%. Also, Pandya and Wadhwani [17] used the wavelet packet decomposition which made use of the 'db8' mother wavelet function for feature extraction from the vibration signal and the best node of wavelet packet tree was performed using best tree algorithm along with minimum Shannon entropy criteria with a combination of artificial neural network for automatic fault classification of rolling element bearing and they were both able to obtain an overall fault classification rate of 97% and 98.33% respectively. However, the process of computation and mathematical expression is very cumbersome.

Li et al. [3, 18] used neural networks to diagnose motor rolling bearings by combining it with the time-frequency domain analysis for the diagnosis of a fault. Likewise, the Deep Neural Network (DNN) is a promising tool for fault diagnosis of both the planetary gear and the rolling element bearing, as ANN has deficiencies in diagnosing complex systems which are non-stationary and non-linear. With the use of ANN for diagnosing such complex system, features are manually extracted depending on much prior knowledge about signal processing. Likewise, ANN have shallow architectures thereby limiting it to learn complex non-linear relationships in fault

diagnostics issues. In Ref. [1] DNN was used to overcome these deficiencies in their work. They used DNN to implement both fault feature extraction and intelligent diagnosis. DNN are implemented by first pre-trained by an un-supervised layer-by-layer learning and then fine-tuned using a supervised algorithm. Lin et al. [19] 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 Phua et al. [20] 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 [21-23].

HMM has found application in so many fields like in electrical, accounting, mechanical etc. In electrical, it has been used for fault diagnostics in industrial machines, synchronous motors, reluctance motors etc. [3, 24, 25]. Jadhav and Bhandari [26, 27] in their work stated that the most common fault in the rotary electric motors (REM) is the bearing related faults which are responsible for about 50% of all rotary machine faults and he was able to detect and diagnose fault in synchronous motors using hidden Markov Model. Reluctance motor was successfully diagnosed for fault by classification method using hidden Markov Model by Lerner et al. [29] while industrial machines was considered in Ref. [30].
In accounting and business-related fields, HMM was used by [20, 27, 31] to show that it can be used for the detection of frauds. Their objective was to detect the anomaly during the transaction and confirm frauds by asking some security code questions.

The Hidden Markov Model (HMM) is a parametric statistical method that has the capability of pattern classification and is suitable for dynamic time series of signals that are non-stationary, has poor repeatability and reproducibility. It is a finite set of states,

where each state is linked with a probability distribution. The transitions among these states are governed by a set of probabilities called transition probability. HMMs are often referred to as the 'gold standard' for the difficult task to perform speech recognition [32, 33]. Nelwamondo et al. [6] used HMM combined with the use of Mel-frequency cepstral coefficient, fractal and Gaussian mixture models for early classifications of bearing faults. Lerner et al. [29] used a HMM based semi-nonparametric approach for fault detection and diagnosis in synchronous motors. Soualhi et al. [34] 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 Jadhav and Bhandari [27] were the first to use a Bayesian Hidden Markov Model-based approach for detecting anomalies in electronic systems. Table 1 in the appendix shows a brief state-of-the-art-approach with the merits and demerits associated with them as found in the introduction.

In this work, a Bayesian robust new HMM (BRNHMM) is used and compared to neural network pattern recognition for fault classification and detection under varying load and speed conditions. It works well for both static and dynamic signals and is thus effective in their modeling hence its suitability for online fault detection, since small data samples are used in its diagnostics. It also represents constructively the uncertainty in our model structure and parameters as it helps to over-come the limitation of miss-specification of the prior. In addition to the above, it averages overall possible values giving unbiased estimates, robust predictions and allows the comparison of assumptions underlying different models. Among the disadvantages is its intense computational burden especially for models involving many variables and the posterior distributions are also more difficult to incorporate into a meta-analysis [9, 11, 22]. 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. The paper is structured as follows. In Sec. 2 the theorem and algorithm

of BRNHMM based on fault diagnosis approach is briefly introduced. The experimental procedure is described in Sec. 3. In Sec. 4 the results from the experiment are presented and a conclusion is drawn.

2. Theoretical background of Bayesian Robust New Hidden Markov Model [28]

HMMs have proven to be very effective among the various stochastic approaches, in modeling both static and dynamic signals. HMM is a finite-state machine which changes its state during every time increment.

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 does 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 [33, 35].

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 $\lambda = (A, C_{jm}, \mu_{jm}, \sum_{jm}, \pi)$ 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 [29, 36]. These three basic algorithms represent the three basic problems to be solved respectively. The three problems to be solved by these algorithms are:

- **Evaluation**, with an observation sequence $O = o_1, o_2, ..., o_T$, and a model $\lambda = (A, B, \pi)$ how should one effectively compute $P(O|\lambda)$ giving that it is the probability of the observation sequence, when given the model. The forward-backward procedure is used for solving this problem.

- **Decoding**, with an observation sequence $O = o_1, o_2, ..., o_T$, and a model λ, how does one choose a corresponding state sequence $Q = q_1, q_2, ..., q_T$, that is optimal to the observation sequence generated. The Viterbi algorithm is used in this situation.

- **Training**, this relates with how one adjusts the parameters $\lambda = (A, B, \pi)$ to maximize $P(O|\lambda)$ which is the likelihood of all observation sequences. This is a problem of determining the reference model faults which is solved by 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 $O = o_1, o_2, ..., o_T$ 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 an 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 Lth fault state or to maximize the $P(O|\lambda_L)$ of the probability of observation sequence O given the model λ_L. Expectation maximization (EM) is another term for which Baum-Welch re-estimation algorithm is known as Ref. [27]. It is needful for one to write out the log probability of the hidden variables and observations to derive the EM algorithm for learning the parameters.

$$\log P\left(S_{1:T}, Y_{1:T}\right) = \log P(S_1) + \sum_{t=1}^{T} \log P(Y_t|S_t) + \sum_{t=2}^{T} \log P(S_t|S_{t-1}) \qquad (1)$$

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

2) In the context of the EM algorithm, the variational Bayesian is the learning algorithm for continuous HMM as the 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 2 below

$$F(\theta) = \int S(\theta) \log P(O|\theta) \partial\theta - KL[S(\theta)\|P(\theta)] \qquad (2)$$

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)

$$KL[S\|P] = \int S(\theta) \log \frac{S(\theta)}{P(\theta)} \partial\theta \qquad (3)$$

if KL is positive and greater than zero, and F provides a lower bound on the model log-likelihood [37]. 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 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 $P(S, \theta|D, M)$.

$$\log P(D|M) = \log \int \partial\theta P(D,\theta|M) \qquad (4)$$

$$\geq \int \partial\theta \, Q(\theta) \log \frac{P(D,\theta|M)}{Q(\theta)} \qquad (5)$$

$$= \int \partial\theta Q\left(\theta\left[\log P(D|\theta,M) + \log \frac{P(\theta|M)}{Q(\theta)}\right]\right) \qquad (6)$$

$$\geq \int \partial\theta Q\left(\theta\left[\Sigma_S Q(S) \log \frac{P(S,D|\theta,M)}{Q(S)} + \log \frac{P(\theta|M)}{Q(\theta)}\right]\right) \qquad (7)$$

$$\equiv F\left(Q(\theta), Q(S)\right) \qquad (8)$$

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 fig. 1 below shows the flow chart of the training to fault detection procedure.
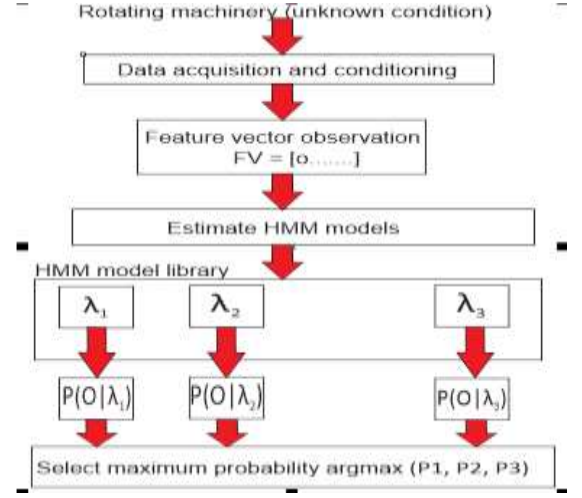

Fig. 1 The HMM pattern-based classification for BRNHMM see (Ref. [33]).

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. Experimental setup

The test rig setup comprises of brushless AC motor (Rockwell Automation MPL-3680B) mounted on a NSK 6309 single row bearing used to drive the system as shown in Fig. 2. The angular velocity of the motor was retrieved from one of the angular outputs made available in the motor drive, which is a Rockwell Automation Kinetix 6000 series BM-01 and allows a continuous speed variation from 0 to 3600 rpm. A Soundwel AE sensor with model number SR 150M with a frequency range of 25-530kHz was used for measurement.

Fig. 2. Test rig setup

Two servo-hydraulic actuators were used for loading sinusoidally, axial and radial loads on test bearings whose purpose is to allow simulating a real-life scenario of the bearings of a functional machine experiencing varying cyclic loads at different rotational speeds and frequencies. The test bearings used were three taper roller bearings (Timken HR 30307 J) which were inserted into the test rig one after the other with two having been introduced with defects and the third bearing been left undamaged. The test rotating speeds for the slow rotating bearings ranged from 70 to 100 rpm. Fig. 3 shows a schematic view of the test rig setup.
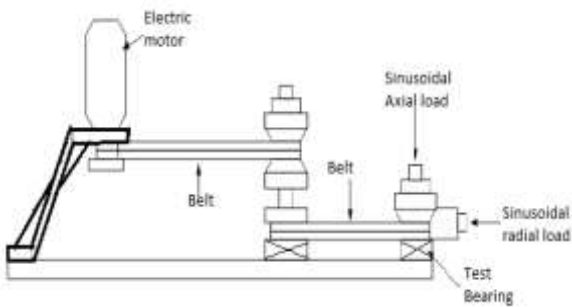

Fig. 3. Schematic diagram of test rig setup

The first bearing (good) was loaded sinusoidally with forces of amplitude 500N at a frequency of 2Hz on the axial load and amplitude of 900N at a frequency of 1Hz on the radial load. Bearing two, in which debris was introduced, was also loaded sinusoidally with forces of amplitude of 500N at a frequency of 2Hz in the axial direction and amplitude of 900N at a frequency of 1Hz in the radial direction. Bearing three, with a crack on the outer race was loaded

sinusoidally with forces of amplitude of 500N on the axial at a frequency of 2Hz and 900N at a frequency of 1Hz in the radial direction as the other two bearings. The three roller bearing AE signatures were collected for four speeds, set at 70 rpm, 80 rpm, 90 rpm and 100 rpm using an FFT analyzer, a National Instrument data acquisition card (BNC-2110) with a shielded BNC connector block.

The induced crack was seeded on the outer raceway of the bearing (as shown in Fig. 4) 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.

Fig. 4. Seeded damage on outer race of a bearing

4. Results

Fig. 5 shows a typical acoustic signal obtained from the bearing test rig, the length of the data taken was 60000 data samples. It cannot be observed that the amplitude of the wave formed as shown in the figure varies within some given time interval.
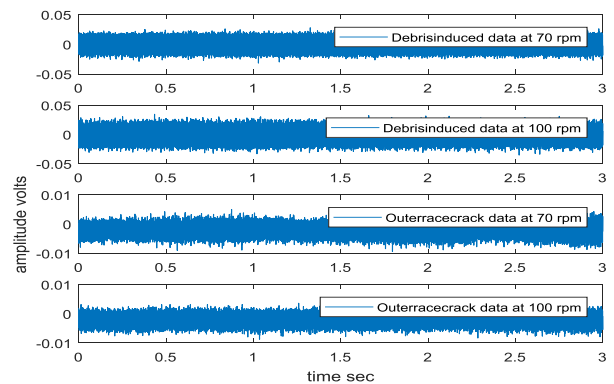

Fig. 5 Acoustic signal from the test experiment

As it was stated before, taper roller bearings (Timken HR 30307 J) were used to be able to artificially introduce the localized-defect,

since it can be dismantled from the outer raceway. The surface damage was seeded on the outer raceway of the bearing and the bearing characteristic is given in the table below. The ball-pass frequency of the outer ring was found to be 25.103 KHz and 35.862 KHz for the two speeds respectively.

Table 2: Bearing characteristics

Contents	Parameters
Bearing specification	Timken taper roller bearing HR 3030 7 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

It should however be noted that with AE, frequencies of the order of 100 kHz are involved 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. 6 below are also difficult to obtain especially for the outer race cracked bearings.
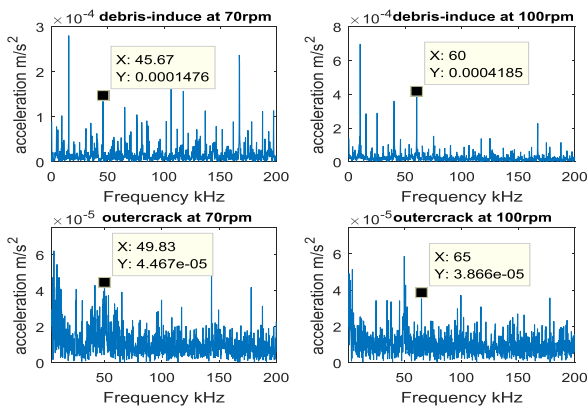

Fig. 6 FFT plot of the acoustic signal

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.

Fig. 7 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).
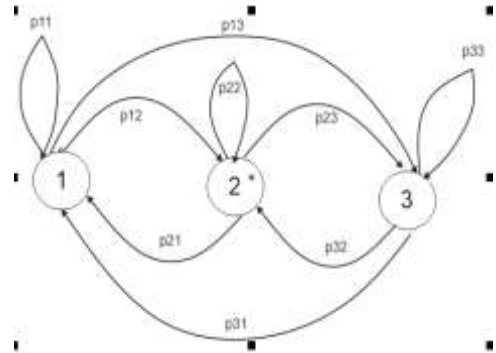

Fig. 7   BRNHMM state model for the bearing fault classification

The transition between each state is presented in Table 3 below.

Table 3: 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 speed 70, 80, 90 and 100 rpm.

Figs. 8 a, b, and c and 9 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.
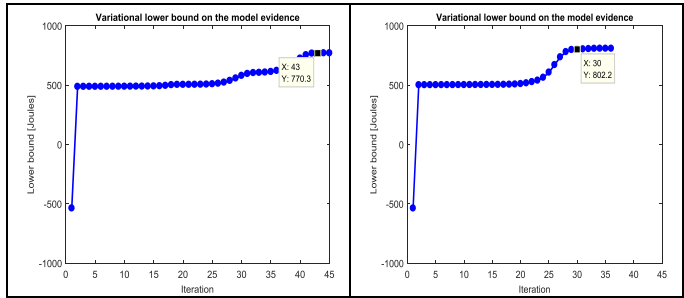
Fig. 8a Variational lower bound for good bearing with training at 70 rpm

Fig. 8b Variational lower bound for debris induced bearing with training at 70 rpm
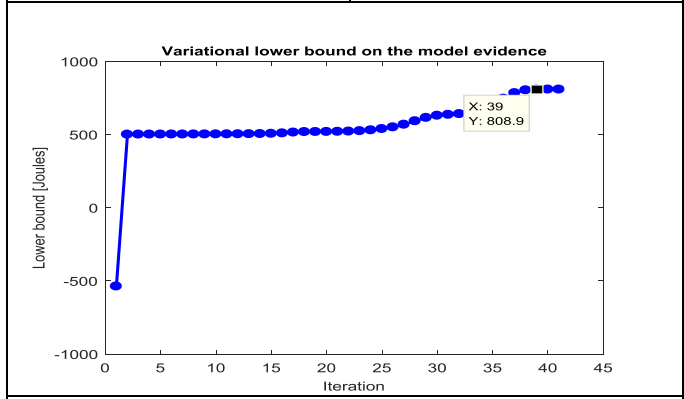
Fig. 8c Variational lower bound for outer race cracked bearing with training at 70 rpm

It is evident from figs. 8 a, b & C that for the good bearing model the free energy of the lower bound which is at 770.3 joules begin to stabilize after around 43 iterations at 70 rpm, whereas for the anomalous models the energy of the lower bound was at 802.2 joules and stabilized a bit faster with that of the debris stabilizing the lowest after 30 iterations. While for the outer crack race it stabilizes at 39

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.
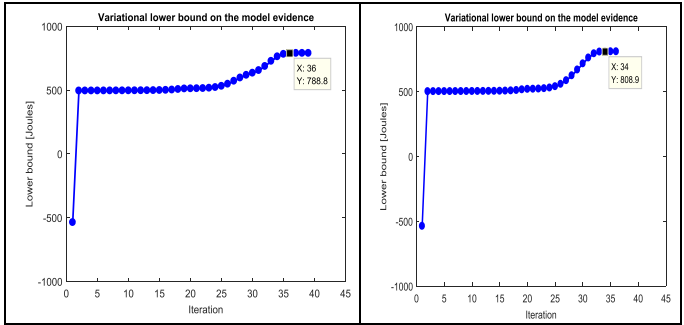
Fig. 9a Variational lower bound for good bearing with training at 100 rpm

Fig. 9b Variational lower bound for debris induced bearing with training at 100 rpm
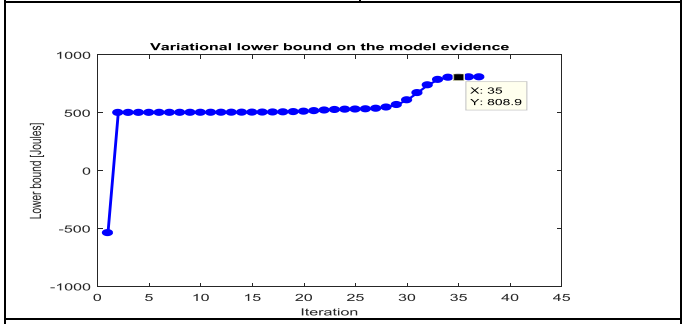
Fig. 9c Variational lower bound for outer race cracked bearing with training at 100 rpm

From figs. 9 a, b and c for the good bearing model the free energy of the lower bound is at 788.8 joules and begins to stabilize after around 36 iterations at 100 rpm, whereas for the anomalous models the energy of the lower bound is at 808.9 joules and stabilized after 34 iterations. The outer crack race model stabilized after 35 iterations at 100 rpm with the energy of the lower bound being at 808.9 joules. These experiments indicate that energy of the lower bound is usually in the highest when cracks are formed in 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 good 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 both the healthy and anomalous models is shown in figs. 10 a-c. As stated in paragraph 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.

Fig. 10 a) Bayesian robust new HMM for normal bearing model at 100 rpm

Fig. 10 b) Bayesian robust new HMM for debris induce bearing model at 100 rpm

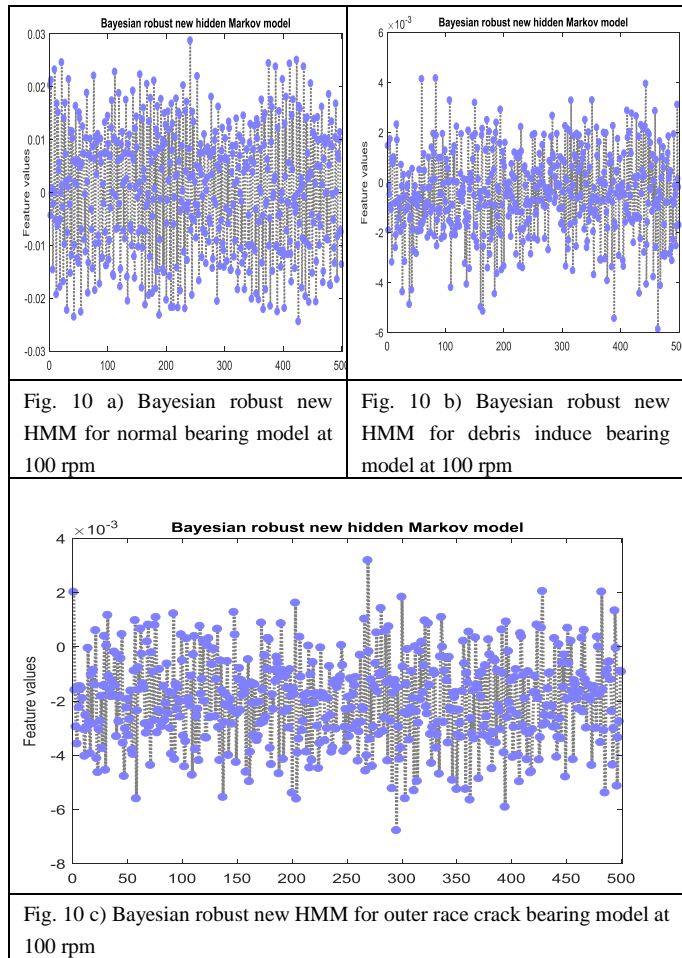Fig. 10 c) Bayesian robust new HMM for outer race crack bearing model at 100 rpm

Table 4 shows the features equations for the training done in ANN.

Table 4: Features extracted from the AE time histories data and their equations [7, 37]

Feature	AE time histories abbreviation	Equation
Kurtosis	$X_{kurtosis}$	$\dfrac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^4}{(N-1)\sigma^4}$
Skewness	$X_{skewness}$	$\dfrac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^3}{(N-1)\sigma^3}$
Crest factor	X_{CF}	$\dfrac{\max\left(|x_i|\right)}{\sqrt{\dfrac{1}{N}\sum_{i=1}^{N}x_i^2}}$
Root-mean-square	X_{rms}	$\sqrt{\dfrac{1}{N}\sum_{i=1}^{N}x_i^2}$
Standard deviation	$X_{std} = \sigma_x$	$\sqrt{\dfrac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}$
Variance	$X_{var} = \sigma_x^2$	$\dfrac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2$
Norm	$X_{norm} = \|X\|_2$	$\sqrt{\sum_{i=1}^{N}|x_i|}$
Mean	$X_{mean} = \bar{X} = \mu_x$	$\dfrac{1}{N}\sum_{i=1}^{N}x_i$
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 network has three layers (the input layer, the hidden layer and the output layer) with the

hidden layer having 10 states. Table 5 shows the result obtained after the training was done for the ANN.

Table 5: 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

Table 6 shows the result of the log-likelihood for the different model run for BRNHMM.

Training Mode		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

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 ANN was not able to classify correctly at a low speed of 70 rpm but shows correct classification at higher speed see table 5. Unlike the ANN, the BRNHMM correctly classify for the difficult condition of low speed and varying load.

The total percent of correctly classified cases with the ANN as specified in fig. 11 is 60.9 % while the misclassified cases is 39.1 %. This result from the confusion matrix show that ANN is not a very good classifier for the varying load and speed condition of rolling element bearing. However, the confusion matrix for the Bayesian Regularization method as shown in fig. 12 proves that the Bayesian method gives better result than ANN.

The total percent of correctly classified cases with the Bayesian method as specified in fig. 12 is 99.1 % while the misclassified cases is 0.9 %.
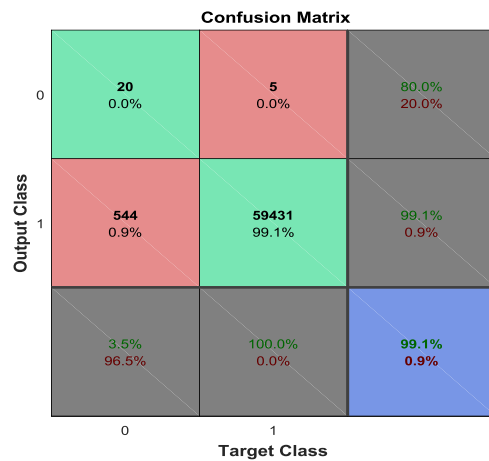

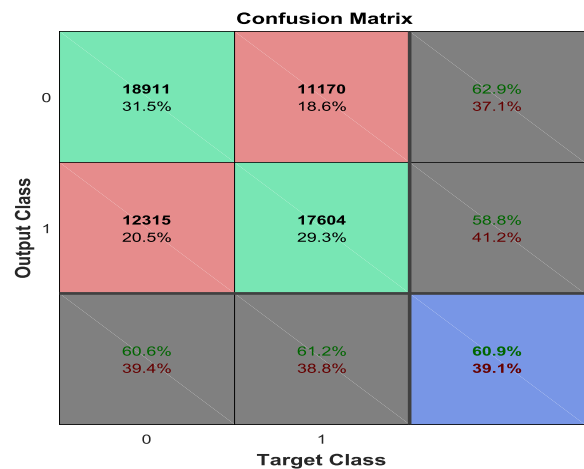Fig. 11 Confusion matrix for the Bayesian method

Fig. 12 The confusion matrix for the ANN classification method

5. Conclusion

Presented here in this paper 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. BRNHMM works satisfactorily with raw signals with little or no pre-processing and uses Kullback-Leibler divergence method for obtaining faster iteration in the various models which thus makes this work unique to the other methods in existence. Through the experimental study of roller bearing rotating at low speed and varying load condition, BRNHMM has been shown to be 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. The 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.

6. *Reference*

[1] F. Jia, Y. Lei, J. Lin, X. Zhou and N. Lu, 'Deep neural networks : A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data', *Mechanical Systems and Signal Processing*, 72–73, (2016) pp. 303–315.

[2] S. John Sakellariou, K A. Petsounis and D. F. Spilios, Vibration based fault diagnosis for railway vehicle suspensions via a functional model based method: A feasibility study, Journal of Mechanical Science and Technology, 29 (2) (2015) pp. 471-484

[3] B. Li,. M-Y. Chow, Y. Tipsuwan and J. C. Hung, 'Neural-network-based motor rolling fault diagnosis', *IEEE Transactions on Industrial Electronics*, 47(5), (2000) pp. 1060–1069.

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

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

[6] F. V. Nelwamondo, T. Marwala, and U. Mahola, '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), (2005) pp. 1–19.

[7] H. Ocak, and K. A. Loparo, '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*, (2001) pp. 1–4.

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

[9] G. Cheng, Y-L. Cheng, L-h. Shen, J-B. Qui and S. Zhang, 'Gear fault identification based on Hilbert – Huang transform and SOM neural network', *Measurement*, 46(3), (2013) pp. 1137–1146.

[10] Z. Derouiche, M. Boukhobza, B. Belmekki and J. M. Rouvaen, 'Application of neural networks for monitoring mechanical defects of rotating machines', *Journal of Energy and Power Engineering*, 6, (2012) pp. 276–282.

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

[12] R. Zaeri, A. Ghanbarzadeh, B. Attaran and S. Moradi 'Artificial neural network based fault diagnostics of rolling element bearings using continuous wavelet transform', *IEEE*, 11(3), (2011) pp. 753–758.

[13] X. Chen, G. Cheng, H. Li and Y. Li 'Fault identification method for planetary gear based on DT-CWT threshold denoising and LE †', *Journal of Mechanical Science and Technology*, 31(3), (2017) pp. 1035–1047.

[14] Z. Shuai, Z. Yongxiang and Z. Jieping 'Rolling element-bearing feature extraction based on combined wavelets and quantum-behaved particle swarm optimization', Journal of Mechanical Science and Technology, 29 (2) (2015) pp. 605-610.

[15] X-H. Chen, G. Cheng, X-L. Shan, X. Hu, Q. Guo and H-G. Lin 'Research of weak fault feature information extraction of planetary gear based on ensemble empirical mode decomposition and adaptive stochastic resonance', *Measurement*. Elsevier Ltd, 73, (2015) pp. 55–67.

[16] D. H. Pandya, S. H. Upadhyay and S. P. Harsha, 'Ann based fault diagnosis of rolling element bearing using time-frequency domain feature', *International Journal of Engineering Science and Technology*, 4(6), (2012) pp. 2878–2886.

[17] M. Yadav, and S. Wadhwani, 'Automatic Fault Classification of Rolling Element Bearing using Wavelet Packet Decomposition and Artificial Neural Network', *International Journal of Engineering and Technology*, 3(4), (2011) pp. 270–276.

[18] W-Y. Choi, D-H Choi and K-J Cha, 'Robust estimation of support vector regression via residual bootstrap adoption', Journal of Mechanical Science and Technology, 29 (1) (2015) pp. 279-289

[19] C-M. Lin, A-B. Ting and M-C. Li, 'Neural-network-based robust adaptive control for a class of nonlinear systems', *Neural Comput & Applic*, 20, (2011) pp. 557–563.

[20] C. Phua, D. Alahakoon and V. C. S. Lee, 'Minority Report in Fraud Detection : Classification of Skewed Data Minority Report in Fraud Detection : Classification of Skewed Data', *Sigkdd Explorations*, 6 (1) (2004) 50-59

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

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

[23] B. Yang, T. Han and W. Hwang, *Fault diagnosis of rotating machinery based on multi-class support vector machine*, Journal of Mechanical Science and Technology (KSME Int. J.), Vol. 19, No. 3, (2005); pp 846 - 859.

[24] G. Zoubin 'An Introduction to Hidden Markov Models and Bayesian Networks', *International Journal of Pattern Recognition and Artificial Intelligence*, 1(15), (2001) pp. 9–42

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

[26] O. Geramifard, J-X. Xu, S. K. Panda. 'Fault detection and diagnosis in Synchronous motoers using Hidden Markov model-based Semi-nonparametric approach', (2013) pp. 1–17.

[27] S. N. Jadhav, and K. Bhandari, 'Anomaly detection using hidden Markov Model', *International journal of computational Engineering Research*, 3(7), (2013) pp. 28–35.

[28] E. Dorj, and C. Chen, 'A Bayesian Hidden Markov Model-Based Approach for Anomaly Detection in Electronic Systems', *IEEE*, (2013) pp. 1–10.

[29] U. Lerner, R. Parr, and D. Koller, 'Bayesian Fault Detection and Diagnosis in Dynamic Systems', in *Proceedings of the seventeenth National Conference on Artificial Intelligence (AAAI-00)*, (2000) pp. 531–537.

[30] P. Sirima, and P. Pokorny, 'Hidden Markov models with covariates for analysis of defective industrial machine parts', *journal of Mathematics and Statistics*, 10(3), (2014) pp. 322–330.

[31] S. J. Stolfo, W. Fan, W. Lee, A. Prodromidis and P.K. Chan 'Cost-based Modeling for Fraud and Intrusion Detection : Results from the JAM Project', *Proceedings of DARPA Information Survivability Conference and Exposition*, (2000) pp. 130–144.

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

.

[33] Z. Li, Z. Wu, Y. He and C. Fulei 'Hidden Markov model-based fault diagnostics method in speed-up and speed-down process for rotating machinery', *Mechanical Systems and Signal Processing*, 19, (2005) pp. 329–339.

[34] A. Soualhi, G. Clerc, H. Razik and A. Lebaroud 'Fault detection and diagnosis of induction motors based on hidden *Markov model'*, *in Electrical Machines (ICEM), 2012 XXth International Conference on IEEE*, (2012) pp. 1693–1699.

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

[36] L. R. Rabiner, 'tutorial on Hidden Markov Models and selected applications in speech recognition.pdf', *Proceedings of the IEEE*, 77(2), (2009) pp. 257–286.

[37] F. Perez-Cruz, 'Kullback-leibler divergence estimation of continuous distributions', *IEEE International Symposium on Information Theory - Proceedings*, (2008) pp. 1666–1670.

Henry O. Omoregbee is a research student in the Department of Mechanical and Aeronautical Engineering at the University of Pretoria in South Africa and a member in the research group of the Centre for Asset Integrity Management (C-AIM) at the same university.

His research area is on Diagnosis and Prognosis of rolling element bearing using acoustic and vibration measurement and analysis techniques. He has a Master of Science degree in Industrial and Production Engineering from the University of Ibadan, Nigeria and a Bachelor of Engineering in Production Engineering from the University of Benin, Nigeria.

Stephan Heyns is professor in the Department of Mechanical and Aeronautical Engineering at the University of Pretoria in South Africa, as well as director of the Centre for Asset Integrity Management (C-AIM) at this university.

His personal research focuses are on machine and structural health monitoring using vibration measurement and analysis techniques including optical measurement techniques, and physical asset integrity management, and the use of machine learning and statistical analysis techniques in these applications. His current work includes structural integrity monitoring in turbomachinery using non-contact measurement, gearbox condition monitoring, and the use of optical measurement techniques for structural integrity assessment.

Appendix

Table 1: State-of-the-art-approaches with their use.

State-of-the-art approaches	Purpose of use	Advantages	Disadvantages
Time domain analysis	The interpretation of the signal is done through several parameters. Some of the parameters are: RMS, Crest factor, peak, probability density function, second, third and fourth order statistical moments, which can be extracted from vibration signal	1. Simple to analyze criteria like the amplitude and peak to peak value. 2. Calculation of parameters is not much involving and can be in-terpreted with simplicity. 3. Easy to implement in low cost online monitoring equipment. 4. Depending on the quality and sensitivity of the extracted features used in estimating the bearing condition,the accuracy of the fault can be determined.	1. Difficult to retrieve rele-vant information from time signal. 2. The parameters do not give nor provide any diagnostic information.
Frequency Analysis	In frequency domain-based analysis Fourier transformations are employed to transform time domain signals into frequency domain. To retrieve information from time domain signal e.g. with the use of Fast Fourier transform. This is often applied where there are harmonics in the periodicity of the signal.	1. frequency domain analysis is the most commonly used because of the simplicity of application of the Fourier transform and its ease of interpretation 2. The fact that in using spectrum analysis each discrete frequency can be monitored in contrast to the overall amplitude which is an advantage over the time domain analysis.	1.The major disadvantages of Fourier analysis considered here can be cited as information loss and diffic-ulty in interpreting the si-gnals when moving from time domain to frequency domain, particularly in non-stationary signals. 2.Frequency domain meth-ods however do require that the bearing defect fre-quencies must be known or es-timated, and interpre-tation becomes more difficult when the signal to noise ratio is low. 3.Frequency domain meth-ods also tend to average in transient vibrations and

			therefore becomes sensit-ive to background noise.
Time-frequency analysis	time-frequency analysis methods such as wavelet transform could be widely used to detect faults and can be applied to non-linear and non-stationary signal processing	1. It can be combined with high intelligence network to yield better results in situation where the signal is complex, non-linear and non-stationary. 2. It can be used to deal with small para-meter sginal which is suitable for online analysis.	1. some deficiencies which includes, border distortion, energy-leakage and interference terms. 2. There is an increase in computational cost.
parametric statistical method	parametric statistical method has the capability of pattern classification and is suitable for dynamic time series of signals that are non-stationary, and which has poor repeat ability and reproducibility. e.g. HMM	1. HMMs are often refer red to as the 'gold stan dard' for the diffi-cult task to perform speech recognition. 2. It can be combined w ith high intelligence net work to yield better res ults in situation where the signal is complex, non-linear and non-stationary. 3. When combined with high intelligence networ k, can be highly suitable for analyzing small data set hence good for online use.	1. On their own they usually involve large data set hence not suitable for online use.
Artificial intelligence	Is a promising field with recent advances in its model such as the deep neural network	1. The advent of deep neural network has help to overcome all the deficiencies asso-ciated with the use of ANN	1. Often involve the use of large data set. 2. ANN has deficiencies in diagnosing complex systems which are non-sta tionary and non-linear. 3. ANN have shallow arch -itectures thereby limiting it to learn complex non-li near relationships in fault diagnostics issues.