

**MONITORING OF THE INCIPIENT BEARING DAMAGE
IN INDUCTION MOTORS USING INTELLIGENT
TECHNIQUES**

**M.Sc. Thesis by
Tayfun ŞENGÜLER, B.Sc.**

Department : Electrical Engineering

Programme: Electrical Engineering

Supervisor : Prof. Dr. Serhat ŞEKER

JANUARY 2008

**MONITORING OF THE INCIPIENT BEARING
DAMAGE IN INDUCTION MOTORS USING
INTELLIGENT TECHNIQUES**

**M.Sc. Thesis by
Tayfun ŞENGÜLER, B.Sc.
(504031031)**

Date of submission: 24 December 2007

Date of defence examination: 28 January 2008

Supervisor (Chairman): Prof. Dr. Serhat ŞEKER

Members of the Examining Committee: Assist. Prof.Dr. Levent OVACIK

Assist. Prof.Dr. Osman Kaan EROL

JANUARY 2008

**ASENKRON MOTORLADA GELİŞMEKTE OLAN
RULMAN ARIZA DURUMUNUN AKILLI
TEKNİKLERLE İZLENMESİ**

**YÜKSEK LİSANS TEZİ
Müh. Tayfun ŞENGÜLER
(504031031)**

**Tezin Enstitüye Verildiği Tarih : 24 Aralık 2007
Tezin Savunulduğu Tarih : 28 Ocak 2008**

**Tez Danışmanı : Prof.Dr. Serhat ŞEKER
Diğer Jüri Üyeleri Yrd. Doç. Dr. Levent OVACIK
Yrd. Doç. Dr. Osman Kaan EROL**

OCAK 2008

ACKNOWLEDGEMENT

I would like to thank to my family, my advisor Prof. Dr. Serhat ŞEKER, my colleague and dear friend Erinç KARATOPRAK and my future wife Melek ÇALIŞKAN for their helps and supports both mentally and technically.

December 2007

Tayfun ŞENGÜLER

TABLE OF CONTENTS	Page
LIST OF TABLES.....	vii
LIST OF FIGURES.....	viii
ABBREVIATIONS.....	x
SUMMARY.....	xi
1. INTRODUCTION.....	1
2. FEATURE EXTRACTION TECHNIQUES	4
2.1 Statistical Parameters	5
2.2 Fourier Transform.....	6
2.3 Short-Time Fourier Transform	9
2.4 Continuous and Discrete Wavelet Transform	13
2.5 Shannon’s Entropy Approach.....	15
3. DIMENSIONALITY REDUCTION AND MULTIVARIATE DATA ANALYSIS.....	16
3.1 Principal Components Analysis (PCA).....	16
3.2 Independent Component Analysis (ICA).....	17
3.2.1 Maximization of Nongaussianity	18
3.2.2 Minimization of the Mutual Information.....	20
3.2.3 Maximum Likelihood Estimation	21
3.3 K-means Clustering.....	22
4. CLASSIFICATION	23
4.1 Bayesian Decision Theory.....	23
4.2 Neural Networks	25
4.2.1 Multilayer Perceptrons (MLP).....	25
4.2.2 Probabilistic Neural Networks (PNN).....	27
4.2.3 Kohonen’s Self Organizing Map (SOM).....	28
5. APPLICATION	31
5.1 Induction Motors.....	31
5.2 Bearing Fluting Mechanism and Experimental Set-up	32
5.2.1 Bearing Fluting Mechanism.....	33
5.2.2 Experimental Setup for Aging Mechanism	33
5.2.3 Vibration Analysis For Bearing Damage Detection	35

5.2.4 Condition Monitoring Studies for Mixed Load Cases	36
5.2.4.1 Condition Monitoring Using Statistical Parameters and Entropy	37
5.2.4.2 Condition Monitoring Using MLP Output Error.....	41
5.2.4.3 Condition Monitoring Using ICA Based on WPD	45
5.2.4.4 Condition Monitoring Using SOM	46
6. CONCLUSION	52
REFERENCES	54
APPENDICES.....	59
VITA.....	65

LIST OF TABLES

	<u>Page No:</u>
Table 2.1 : Frequency Components of S_1 and S_2	7
Table 2.2 : Statistical Parameters of the Noisy Signals, S1 and S2	8
Table 2.3 : Frequency Components of S_1 and S_2	10
Table 5.1 : Nameplate Information of the 5-HP Motor	35
Table 6.1 : Comparing the Monitoring Methods	53

LIST OF FIGURES

Page No:

Figure 2.1 : Gizah Pyramids.....	4
Figure 2.2 : Noisy Signals, S1 and S2	7
Figure 2.3 : Power Spectral Densities of Noisy Signals, S1 and S2.....	8
Figure 2.4 : Stationary and Non-stationary Signals, S1 and S2	9
Figure 2.5 : PSD of Stationary and Non-stationary Signals, S1 and S2.	10
Figure 2.6 : Time-Frequency Representation of Stationary Signal S1	11
Figure 2.7 : Time-Frequency Representation of Non-stationary Signal S2.....	11
Figure 2.8 : Time-Frequency Representation of Non-stationary Signal S2 with Narrow Window Function.....	12
Figure 2.9 : Time-Frequency Representation of Non-stationary Signal S2 with Wide Window Function	13
Figure 2.10 : Signal Decomposition at the Nth stage.	14
 Figure 3.1 : The optimization landscape of kurtosis	19
 Figure 4.1 : Posterior probability values of two classes..	24
Figure 4.2 : A Simple Perceptron	25
Figure 4.3 : General Structure of MLP Neural Network..	26
Figure 4.4 : PNN Structure.....	27
Figure 4.5 : Kohonen Structure.	29
 Figure 5.1 : Induction Motor	31
Figure 5.2 : Induction Motor Cross-Section	31
Figure 5.3 : Schematic of the electrical motor bearing EDM setup.	33
Figure 5.4 : Motor load testing and data acquisition system.....	34
Figure 5.4a : Experimental set-up configuration.....	34
Figure 5.4b : Accelerometers on Cross-section (A-A') at short end	34
Figure 5.4c : Accelerometers on Cross-section (B-B') at pulley end.	34
Figure 5.5 : Spectra of the Motor Vibration Data for the Accelerated Bearing Damage	36
Figure 5.5a : Healthy Case.....	36
Figure 5.5b : Faulty Case	36
Figure 5.6 : Vibration data structure.....	36
Figure 5.7 : Monitoring Results of Statistical Parameters and Entropy of Vibration Sensor 10	37
Figure 5.8 : Monitoring Mean Values of Vibration Sensor 10	38
Figure 5.9 : Monitoring Standard Deviation of Vibration Sensor 10.....	39
Figure 5.10 : Monitoring Variance of Vibration Sensor 10.....	39

Figure 5.11 : Monitoring Skewness of Vibration Sensor 10.....	39
Figure 5.12 : Monitoring Kurtosis of Vibration Sensor 10.....	40
Figure 5.13 : Monitoring Entropy of Vibration Sensor 10	40
Figure 5.14 : MLP Training Scheme	41
Figure 5.15 : Monitoring System with MLP.....	42
Figure 5.16 : Real and Predictive Variance Values for Mixed Load Conditions.....	43
Figure 5.17 : Real and Predictive Entropy Values for Mixed Load Conditions	43
Figure 5.18 : Error and Squared Error between Real and Predictive Variance Values for Mixed Load Conditions	44
Figure 5.19 : Error and Squared Error Between Real and Predictive Entropy Values for Mixed Load Conditions	44
Figure 5.20 : Estimated 1st IC.....	45
Figure 5.21 : Estimated 2nd IC.....	46
Figure 5.22 : Estimated 3rd IC.	46
Figure 5.23 : Training of SOM.....	47
Figure 5.24 : Training Results of SOM	47
Figure 5.25 : Test Results of SOM for the first 4 cycles.	48
Figure 5.26 : Test Results of SOM for the first 6 cycles.	48
Figure 5.27 : Test Results of SOM for all cycles.	49
Figure 5.28 : Mean Quantization Error of SOM	49
Figure 5.29 : Mean Quantization Error of SOM in Logarithmic Scale	50
Figure 5.30 : Mahalanobis Distances.....	50
Figure 5.31 : Mahalanobis Distances in Logarithmic Scale	51

ABBREVIATIONS

PWM	: Pulse Width Modulation
PCA	: Principal Components Analysis
ICA	: Independent Component Analysis
PNN	: Probabilistic Neural Networks
MLP	: Multilayer Perceptron
SOM	: Self Organizing Map
HP	: Horse Power
STFT	: Short-Time Fourier Transform
PSD	: Power Spectral Density
APSD	: Auto-Power Spectral Density
CWT	: Continuous Wavelet Transformation
DWT	: Discrete Wavelet Transformation
MRA	: Multi-Resolution Analysis
PC	: Principal Component
IC	: Independent Component
ANN	: Artificial Neural Networks
BP	: Backpropagation
LMS	: Least-Mean Square
MQE	: Mean Quantization Error
EDM	: Electric Discharge Machining
AC	: Alternate Current
FFT	: Fast Fourier Transform
NFFT	: Number of Fast Fourier Transform
g	: Nominal acceleration due to gravity on Earth

MONITORING OF THE INCIPIENT BEARING DAMAGE OF THE INDUCTION MOTORS BY USING INTELLIGENT TECHNIQUES

SUMMARY

The traditional way to operate induction motors for a long time without any failure to provide reliability as well as all kinds of machinery stuff is to make periodic maintenances on a time schedule. Contemporary engineering approach shows a trend in which the predictive maintenance i.e., online monitoring and early detection of the failures take the place of the periodic maintenance.

It is stated that more than 50 % of the failures of induction motors are mechanical in nature, such as bearing, mechanical faults in rotors, imbalance, and alignment-related problems. Various methods were applied to induction motors for the bearing damage detection and characterization. Time-frequency domain analysis, unsupervised, supervised neural networks and Hidden Markov models were used for condition monitoring of induction motors.

In this study, a 5 horse power (HP) induction motor was subjected to the accelerated aging process. For this reason, a bearing fluting mechanism or Electric Discharge Machining (EDM) approach was considered to show the results of this event with an experimental study. There are eight measurement sets so that one healthy and seven aged cases. Five different load conditions, which are 0-25-50-75-100% respectively, were applied to the motor.

This study focused on tracking the bearing damage of the 5-HP induction motor. Statistical parameters, Shannon's Entropy, Multilayer Perceptron (MLP), Kohonen's Self Organizing Map (SOM) and Independent Component Analysis (ICA) techniques were used to monitor the bearing damage. In this manner, feature vectors were formed by Fourier transform and wavelet packet decomposition. Moreover, dimensionality reduction step was made by Principal Components Analysis (PCA) and ICA.

As a result of this study, pattern recognition techniques can be used for tracking the bearing damage. Bearing damage predictions were made by applying ICA, MLP and SOM. The aim of this study has been reached successfully by applying MLP neural networks because it was not affected by the load changes and development of the bearing damage was well tracked from increasing oscillations of MLP output error (variance error) and prediction of the bearing damage made in the cycle 4, two cycles before it started.

ASENKRON MOTORLADA GELİŞMEKTE OLAN RULMAN ARIZA DURUMUNUN AKILLI TEKNİKLERLE İZLENMESİ

ÖZET

Geleneksel yöntem olarak belirli zaman aralıklarında uygulanan periyodik bakım tüm makinelerde olduğu gibi asenkron motorlarında arıza yapmadan uzun süre çalışmasına yardımcı olur. Gelişmekte olan mühendislik yöntemleri gerçek zamanlı durum izleme ve arıza teşhisine dayalı öngörülü bakım uygulamalarının periyodik bakımın yerine aldığını göstermektedir.

Yapılan çalışmalar asenkron motorlardaki arızaların (rulman, mekanik rotor arızası, dengesizlik ve eksenel bozukluk gibi) %50'sinin mekanik kaynaklı olduğunu göstermiştir. Rulman arızalarının tespiti ve özellik çıkarımı için birçok yöntem kullanılmıştır. Zaman-frekans analizleri, denetimli, denetimsiz yapay sinir ağları ve saklı Markov modelleri asenkron motorlarda durum izleme amaçlı kullanılmıştır.

Bu çalışmada 5 beygir gücündeki bir asenkron motor hızlandırılmış yaşlandırma deneyine tabi tutulmuştur. Bu amaçla, Elektriksel Boşalma Sistemi (rulman yivlendime mekanizması) deney düzeni kurulmuştur. Biri sağlam yedisi yaşlanmış olmak üzere toplam sekiz durumda ölçümler yapılmıştır. Her durumda motorun sırasıyla %0–25–50–75–100 yük durumları gözönüne alınmıştır.

Çalışmada ana hedef olarak hızlandırılmış yaşlanma deneyine tabi tutulan 5 beygir gücündeki asenkron motorun rulman arızasının oluşma süreci takip edilmiştir. Bu amaçla İstatistiksel parametreler, Shannon entropisi, Çok Katmanlı Yapay Sinir Ağları, Kohonen'in Özdenetimli Ağı ve Bağımsız Bileşen Analizi rulman arızasının durumunu izleme amaçlı kullanılmıştır. Öznitelik vektörleri oluşturmak için Fourier dönüşümü ve Dalgacık Paket Dönüşümleri uygun görülerek boyut azaltma amaçlı olarak Tekil ve Bağımsız Bileşen Analizleri yapılmıştır.

Bu çalışmanın sonucu olarak, örüntü tanımlama yöntemlerinin rulman arızasını izlemede başarılığı olduğu gözlenmiştir. Rulmanın öngörülü bakım bilgisi için Çok Katmanlı Yapay Sinir Ağları, Kohonen'in Özdenetimli Ağı ve Bağımsız Bileşen Analizi kullanarak ömür tahmini yapılmıştır. Çalışmada ana hedefe Çok Katmanlı Yapay Sinir Ağları kullanılarak en başarılı şekilde ulaşılmıştır. Değişen yük durumları ağı performansı olumsuz etkilememiştir. Arızanın gelişmesi ağ çıkışındaki hatanın büyümesi ve salınımlarının artması olarak gözlemlenmiştir. Rulman arızası oluşumu gerçekleşmesinden iki yaşlandırma süreci önce tahmin edilebilmiştir.

1. INTRODUCTION

Induction motors have a wide usage in industrial applications since developments in power electronics technology enables easy control schemes such as speed control drives with pulse width modulation (PWM). There are varieties of applications of induction motors from nuclear plants to small household appliances [1].

The traditional way to operate induction motors for a long time without any failure to provide reliability as well as all kinds of rotating machinery is to make periodic maintenances on a time schedule. It is a must for the rotating machinery in factories, harbors, dams and electrical plants etc. where the operation is restless for all day and night.

Periodic maintenance itself may not be protective for induction motors against failures. Even if all maintenance schedules are followed properly, it does not mean that motors will not have a break down. Two equal induction motors running in the same factory do not always work with same loads during the operation hours. A failure is always inevitable for the motor working with higher loads than the other one.

Contemporary engineering approach shows a trend in which predictive maintenance i.e., online monitoring and early detection of failures take the place of periodic maintenance. There have been several studies for the early detection of anomalies and the isolation of the cause of failures in induction motors, with high level of accuracies [2-4]. It is stated that more than 50 % of failures are mechanical in nature, such as bearing, mechanical faults in rotors, imbalance, and alignment-related problems [5-9].

Products for condition monitoring, predictive maintenance are also available such as Artesis Technology System A.S in Turkey developed software and hardware-based products by using a model-based monitoring technology [Url 1].

Various methods were applied to induction motors for bearing damage detection and characterization. In terms of wavelet transform applications, continuous, discrete and

stationary wavelet transform show the similar bearing damage characterization which is indicated in the frequency interval of 2-4 kHz. This characterization can be interpreted as a common feature [10-11].

Bayesian decision algorithm and neural networks were successful in the detection of electrical and mechanical failures in electrical motors [12-13]. Moreover, neural networks were also accurate in the classification of detected failures [13].

Studies like [14] have shown that statistical modeling and monitoring of the statistical parameters can be used as a condition monitoring method since it provides information about the trend of developing failure.

Principle Components Analysis (PCA) approach was used to extract the aging trend and then an entropy function, which is Shannon's entropy, was defined to get an alarm level by means of the maximum entropy value. Also, interpreting the maximum entropy as a measure of the physical damage, the maximum entropy value could be defined the critical alarm level of the bearing damage developing [15].

Unsupervised, supervised neural networks and Hidden Markov models were used to monitor the induction motors for making life predictions [16-18].

In this study, a 5 horse power (HP) induction motor was subjected to the accelerated aging process. For this reason, a bearing fluting mechanism or Electric Discharge Machining (EDM) approach was considered to show the results of this event with an experimental study. Different monitoring approaches were combined with the pattern recognition techniques to track a good aging trend of the bearing failure and to be used in real time applications.

In Chapter 2, feature extraction techniques including time and frequency domain analysis are represented.

PCA, Independent Component Analysis (ICA) and K-means clustering are introduced in Chapter 3.

For classification techniques, Bayesian decision criteria, Probabilistic Neural Networks (PNN), Multilayer Perceptron (MLP) and Kohonen's Self Organizing Map (SOM) are explained in Chapter 4.

Chapter 5 gives the fundamental part of this study focused on tracking the bearing damage. Statistical parameters, Shannon's Entropy, MLP, SOM and ICA techniques

were used to monitor the bearing damage of the 5-HP induction motor. In this manner, feature vectors were Fourier transform and wavelet package analysis and dimensionality reduction was made by PCA and ICA.

The monitoring results are discussed and the applied methods are compared briely in Chapter 6, the final conclusion part.

2. FEATURE EXTRACTION TECHNIQUES

When someone is asked to describe his/her appearance, he/she will simply talk about his/her weight, height, hair color and style, eye color, skin color, voice tone and age etc., the special details which make him/her unique among many people. Actually, these special details are the features and the way of description is the feature extraction process whose method is up to his/her own description.

Feature extraction is a process that is made in everyday's life. When meeting someone new, a set of features of him/her is formed. This set is called the feature vector. Psychologically, human brain uses these feature vectors to identify the people. This process may also be applied to everything in real life. Because pyramids are a good feature to define Egypt, it is not allowed to say when playing Taboo. If someone talks about Gizah pyramids (see in Figure 2.1), it is easy to understand which country he/she means.



Figure 2.1 : Gizah Pyramids adapted from [Url-2].

When dealing with the signals like financial time series or vibration data, it is not so easy to get any particular information. Feature extracting from a data is made by applying different methods. There may be numerous methods which depend on the application. In general, statistical parameters, Fourier transform, short-time Fourier

transform (STFT) continuous and discrete wavelet transforms etc. are the mostly used methods for one-dimensional signals in time and frequency domains.

A feature vector set is the combination of the features extracted by a feature extraction method. Choosing the right method is always the main problem. One can combine the methods below to each other to build its own feature extraction mechanism.

2.1. Statistical Parameters

Statistical parameters are easily extracted in time domain and define the average properties of machinery data. The two main statistical parameters are the mean μ , and the standard deviation σ . For a given data set $\{x_i\}$, these two basic parameters are defined as follows:

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (2.1)$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (2.2)$$

where N is the number of the data points.

Another two parameters are skewness c , and kurtosis k , which reflects the departure from the Gaussian (normal) probability distribution. They are calculated as follows:

$$c = \frac{1}{N \sigma^3} \sum_{i=1}^N (x_i - \mu)^3 \quad (2.3)$$

$$k = \frac{1}{3N \sigma^4} \sum_{i=1}^N (x_i - \mu)^4 \quad (2.4)$$

For a perfect Gaussian distribution c is equal to zero. A negative value is due to skewness toward lower values, whereas a positive value indicates nonsymmetry toward high values. For small data sets, one often gets values that differ from zero. Kurtosis k , which is defined as a measure of the sharpness of the peak in the distribution function, is very close to three for a normal distribution. These statistical

parameters can be used to perform quick check of the changes in the statistical behavior of the signal [19].

2.2 Fourier Transform

In many signal analysis applications, the time domain representation of signals does not give you any relevant information. There are lots of representations of the signals, but the most popular is the Fourier transform. The frequency features of a given signal can easily be determined by computing the discrete Fourier transform.

For a given N samples data, the transform at frequency $m\Delta f$ is defined as:

$$X(m\Delta f) = \sum_{k=0}^{N-1} x(k\Delta t) \exp(-j2\pi km / N), \quad (2.5)$$

where Δf is the frequency resolution and Δt is the data-sampling interval. The auto-power spectral density (APSD) of $x(t)$ is estimated as:

$$S_{xx}(f) = \frac{1}{N} |X(m\Delta f)|^2, \quad f = m\Delta f, \quad (2.6)$$

The statistical accuracy of the estimate in Equation 2.6 increases as the number of data points increase [20].

The set of two random noisy signals S_1 and S_2 from two separate sources are shown in Figure 2.2. That is,

$$S_1 = \sin(2\pi f_0 t) + \sin(2\pi f_2 t) + \sin(2\pi f_4 t) + N(0,1) \quad (2.7)$$

$$S_2 = \sin(2\pi f_0 t) + \sin(2\pi f_1 t) + \sin(2\pi f_3 t) + N(0,1) \quad (2.8)$$

where t is time, f_i is frequency and $N(0,1)$ denotes the random Gaussian noise with zero mean and unit standard deviation. Sampling frequency (f_s) was chosen to be 1000 Hz with data length of 512 points. Frequency components are represented in Table 2.1.

Table 2.1 : Frequency Components of S_1 and S_2 .

Frequencies	Value (Hz)
f_0	50
f_1	100
f_2	150
f_3	200
f_4	250

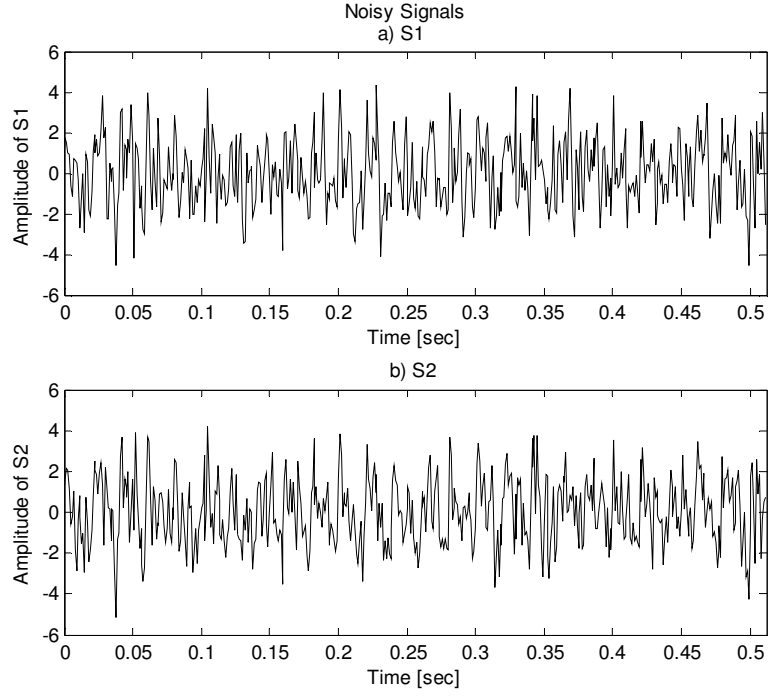


Figure 2.2 : Noisy Signals, S1 and S2

Statistical parameters are represented in Table 2.2, and it is clear that they are not easily separable even if they are from different sources.

Table 2.2 : Statistical Parameters of the Noisy Signals, S1 and S2.

Statistical Parameters	S1	S2
<i>Min</i>	-5.2299	-4.9598
<i>Max</i>	4.2003	4.4973
<i>Mean</i>	-0.0255	-0.0248
<i>Kurtosis</i>	2.7611	2.7257
<i>Skewness</i>	0.0766	0.0574
<i>Std</i>	1.6108	1.6220
<i>Range</i>	9.4303	9.4571

After applying Fourier transform to S_1 and S_2 noisy signals, the frequency spectrum is shown in Figure 2.3. According to their power spectral densities (PSD), it can be said that S_1 has a 50 Hz frequency component and its third and fifth harmonics. Similarly, S_2 has a 50 Hz frequency component and its second and forth harmonics. At this point, it is clearly seen how Fourier transform is useful to see the different characteristic features of signals in frequency domain.

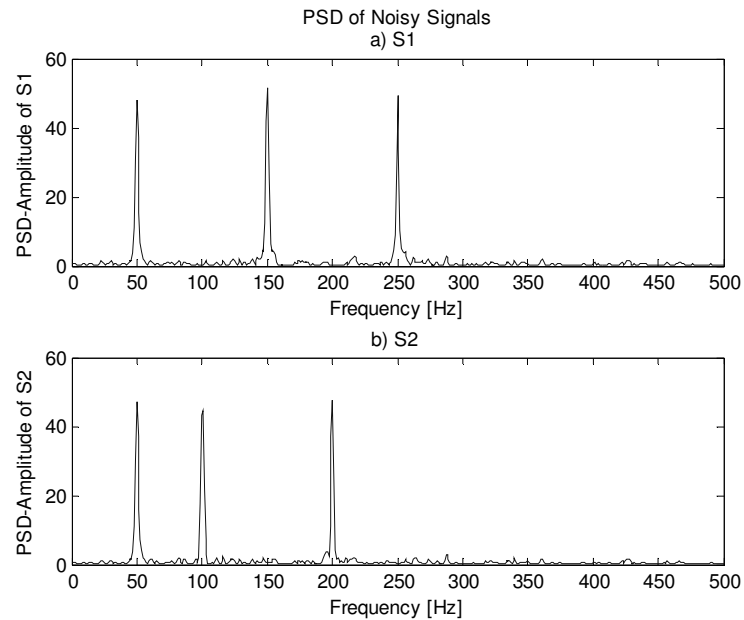


Figure 2.3 : Power Spectral Densities of Noisy Signals, S1 and S2.

2.3 Short-Time Fourier Transform

In the last section, it is suggested that Fourier transform extracts the frequency features for a given signal. Before moving one step forward to STFT, an illustrated example is given below.

In Figure 2.4, S_1 and S_2 are the stationary and the non-stationary signals respectively. That is,

$$S_1 = \sin(2\pi f_0 t_1) + \cos(2\pi f_1 t_1) + \cos(2\pi f_2 t_1) \quad (2.9)$$

$$S_2 = \begin{cases} \sin(2\pi f_0 t) & 0 \leq t \leq 0.127 \\ \cos(2\pi f_1 t) & 0.128 \leq t \leq 0.254 \\ \cos(2\pi f_2 t) & 0.255 \leq t \leq 0.381 \end{cases} \quad (2.10)$$

where t is time, f_i is frequency. Sampling frequency (f_s) was chosen to be 1000 Hz with data length of 128 points. t_i is 0.383 second and frequency components are represented in Table 2.3.

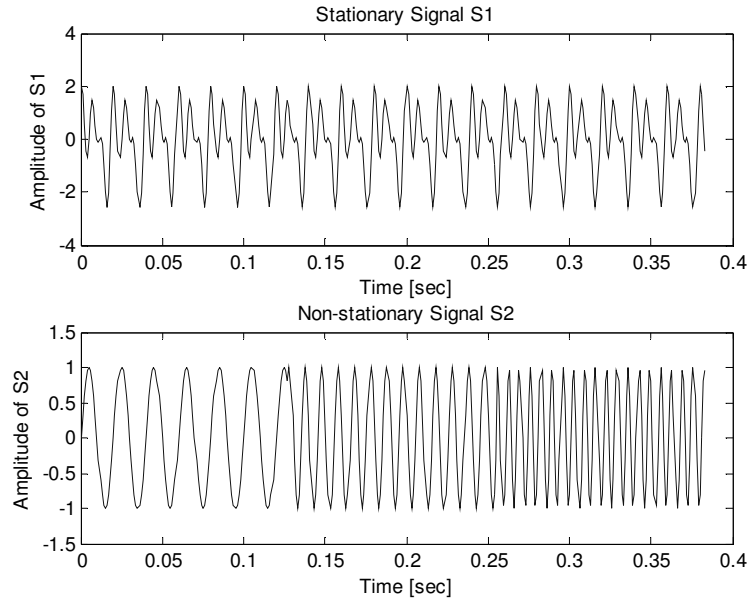


Figure 2.4 : Stationary and Non-stationary Signals, S1 and S2.

Table 2.3 : Frequency Components of S_1 and S_2 .

Frequencies	(Hz)
f_0	50
f_1	100
f_2	150

Even though S_1 and S_2 are completely different signals, PSDs of both signals in Figure 2.5 carry exactly the same frequency components if you ignore their amplitudes. From above example, it is clearly understood that Fourier transform are not successful to distinguish two different signals which carry the same frequency components in different time intervals.

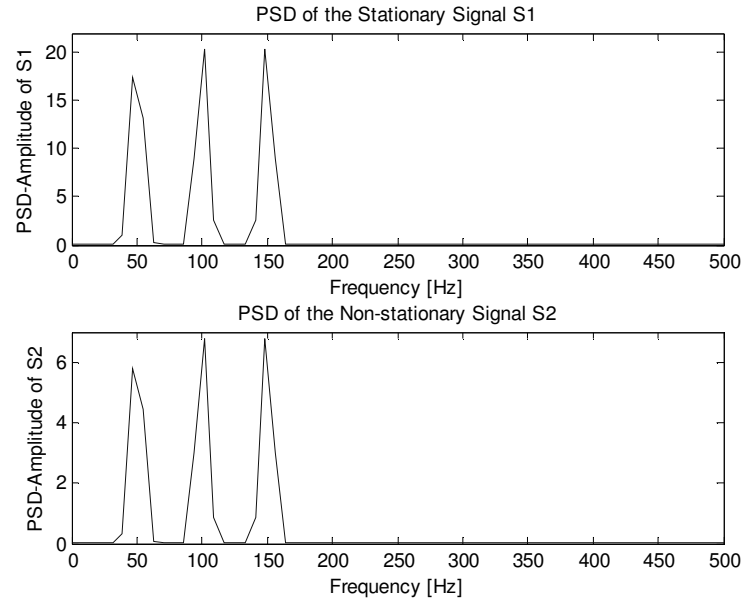


Figure 2.5 : PSD of Stationary and Non-stationary Signals, S1 and S2.

Short-time Fourier transform (STFT) by Gabor in 1946 was introduced which provides time-frequency representation of the signal set to determine the time interval of frequency components of given signals [21]. STFT spectrum is obtained by windowing the signal through a fixed dimension window. The signal may be considered approximately stationary in this window.

Let $x(t)$ is a signal with the assumption that it is stationary when it is windowed through a fixed dimension window $g(t)$, centered at time location τ . Fourier transform of the windowed signal yields STFT.

$$STFT(\tau, f) = \int_{-\infty}^{+\infty} x(t) g(t - \tau) \exp[-2\pi f t] dt \quad (2.11)$$

Equation (2.11) maps the signal into a two-dimensional function in time-frequency (t, f) plane. The analysis depends on the chosen window $g(t)$. Once the window $g(t)$ is chosen, the STFT resolution is fixed over the entire time-frequency plane [22].

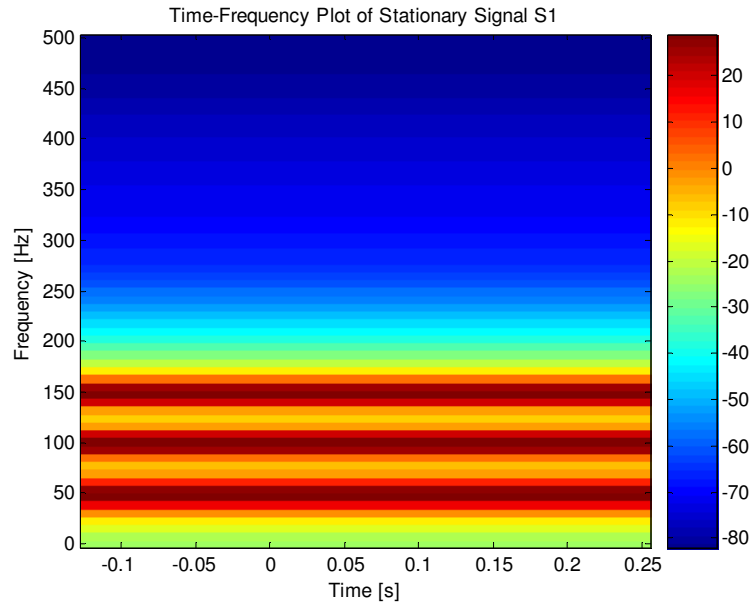


Figure 2.6 : Time-Frequency Representation of Stationary Signal S1.

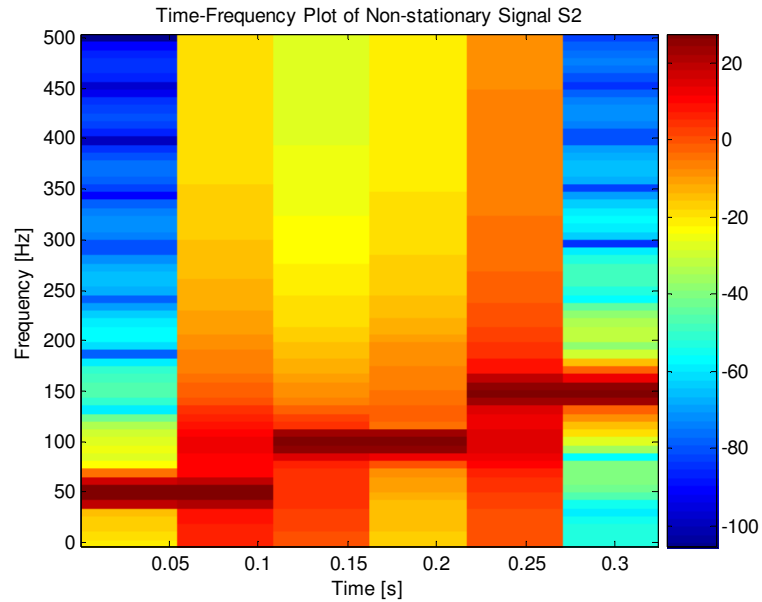


Figure 2.7 : Time-Frequency Representation of Non-stationary Signal S2

The STFT spectrum of the stationary signal S_1 in Figure 2.6 shows that three frequency components (at 50Hz, 100Hz and 150Hz) exist in all time intervals. Also, the STFT spectrum of the non-stationary signal S_2 by using an optimum hanning window function in Figure 2.7 shows that three frequency components exist in different time intervals in the signal. There seems to be some other frequency bands around 50Hz, 100Hz, and 150Hz which are related with the resolution problem of STFT.

If a narrow window function $g(t)$ is used for STFT, frequency resolution decreases as time resolution increase. Reversely, wide window function makes time resolution increase while frequency resolution is poorer. Figure 2.8 and 2.9 shows how wide and narrow hanning window functions affect the frequency and time resolution.

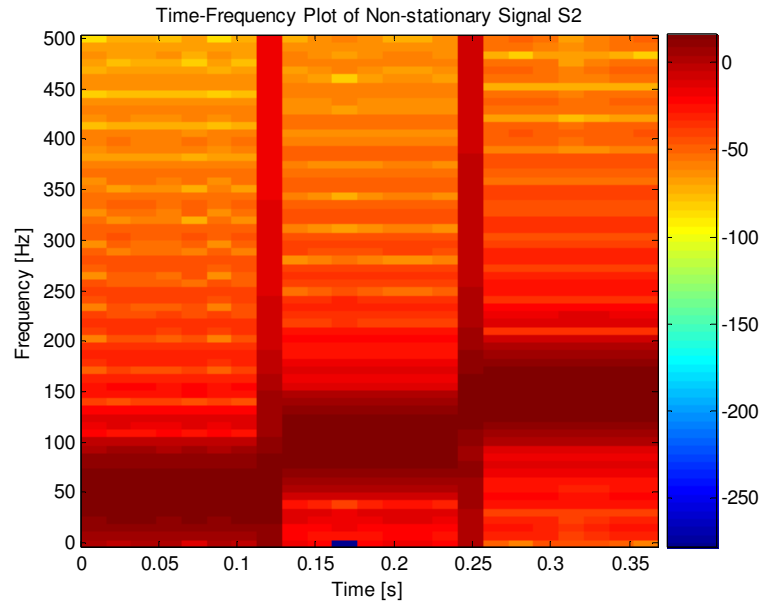


Figure 2.8 : Time-Frequency Representation of Non-stationary Signal S2 with Narrow Window Function

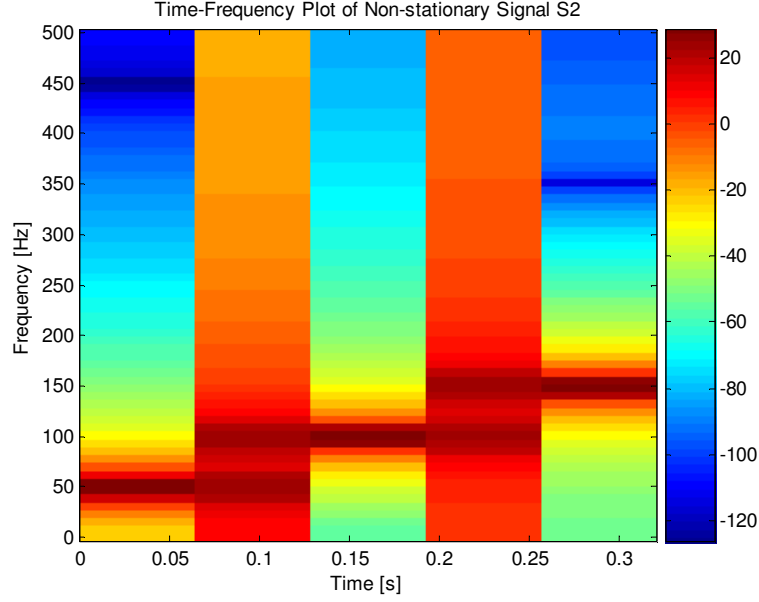


Figure 2.9 : Time-Frequency Representation of Non-stationary Signal S2 with Wide Window Function

2.4 Continuous and Discrete Wavelet Transform

As an alternative method to time-frequency localization in the STFT, wavelet transform is introduced in time-scale plane to overcome the resolution problem.

The use of wavelet transform is particularly appropriate since it gives information about the signal both in frequency and time domains. Let $f(x)$ be the signal, the continuous wavelet transform (CWT) of $f(x)$ is then defined as:

$$W_f(a, b) = \int_{-\infty}^{+\infty} f(x) \psi_{a,b}(x) dx \quad (2.12)$$

where

$$\psi_{a,b} = \frac{1}{\sqrt{|a|}} \psi\left(\frac{x-b}{a}\right) \quad a, b \in \mathbb{R}, a \neq 0 \quad (2.13)$$

Also, it provides the admissibility condition as below:

$$C_\psi = \int_0^{+\infty} \frac{|\psi(\omega)|^2}{\omega} d\omega < \infty \quad (2.14)$$

and for this reason, it is:

$$\int_0^{+\infty} \psi(x) dx = 0 \quad (2.15)$$

Here $\psi(\omega)$ stands for the Fourier transform of $\psi(x)$. The admissibility condition implies that the Fourier transform of $\psi(x)$ vanishes at the zero frequency. Therefore ψ is called as a wave or the mother wavelet and it has two characteristic parameters, namely, dilation (a) and translation (b), which vary continuously. The translation parameter, “ b ”, controls the position of the wavelet in time. A “narrow” wavelet can access high-frequency information, while a wide wavelet can access low-frequency information. This means that the parameter “ a ” varies with different frequency. The parameters “ a ” and “ b ” take discrete values, $a = a_0^j$, $b = nb_0 a_0^j$, where $n, j \in \mathbb{Z}$, $a_0 > 1$, and $b_0 > 0$. The discrete wavelet transformation (DWT) is defined as [23, 24].

$$DWT[j, k] = \frac{1}{\sqrt{a_0^j}} \sum_n f[n] \psi \left[\frac{k - na_0^j}{a_0^j} \right] \quad (2.16)$$

S. Mallat introduced an efficient algorithm to perform the DWT known as the multi-resolution analysis (MRA) [24]. The MRA is similar to a two-channel sub-band coder in high-pass (H) and low-pass (L) filters, from which the original signal can be reconstructed [24]. Figure 2.10 shows the frequency decomposition of the signal schematically. The low frequency sub-band is referred to as ‘approximation Ca_i ’ and the high-frequency sub-band by ‘detail Cd_i ’.

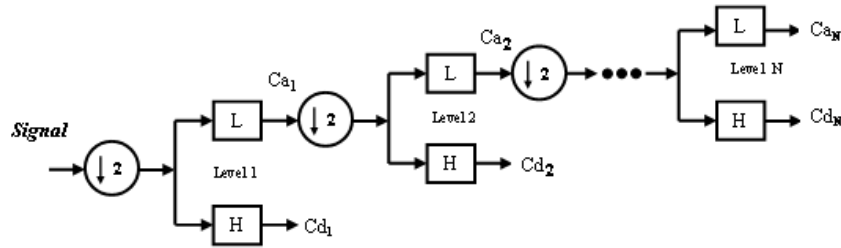


Figure 2.10 : Signal Decomposition at the N^{th} stage.

2.5 Shannon's Entropy Approach

Further being a transform, entropy, the measure of uncertainty, may be a useful tool for extracting information from a set of signals. For a given signal X , the Shannon's entropy is defined as:

$$H_i = -p_i \log_2(p_i + \varepsilon_i) \quad (2.17)$$

where p_i is the probability of x_i for set $X=\{x_1, x_2, \dots, x_n\}$ and ε_i is very small value [25-27].

3. DIMENSIONALITY REDUCTION AND MULTIVARIATE DATA ANALYSIS

By dealing with the n -dimensional feature vectors, some features may reveal classification performances of classifiers decrease. To eliminate the features that affects classification performances negatively and to form low dimensional feature vector sets ($m < n$), the methods below are introduced. Reduced feature vectors in dimension m carry the most important aspects of the original data in dimension n .

3.1 Principal Components Analysis (PCA)

PCA is a kind of multivariable analysis method, which is also called matrix data analysis. By using variable transform, correlated variables are changed into uncorrelated new variables, which is useful to data analysis. So it is used in multi-dimension analysis widely [28].

Let \hat{x} be an n dimensional feature vector element. $X = \{\hat{x}_t\}$, $t=1,2,\dots,m$ and generally, $n < m$. Mean value of the feature vectors is:

$$\hat{\mu} = \frac{1}{m} \sum_{t=1}^m \hat{x}_t \quad (3.1)$$

Then, covariance matrix of these feature vectors is given by:

$$\mathbf{C} = \frac{1}{m} \sum_{t=1}^m (\hat{x}_t - \hat{\mu})(\hat{x}_t - \hat{\mu})^T \quad (3.2)$$

The principal components (PCs) are computed by solving the eigen value problem of the covariance matrix \mathbf{C} :

$$\mathbf{C} \hat{v}_i = \lambda_i \hat{v}_i \quad (3.3)$$

where λ_i ($i=1,2,\dots,n$) are the eigen values and they are sorted in descending order, \hat{v}_i ($i=1,2,\dots,n$) are the corresponding eigenvectors.

To represent the raw feature vectors with low-dimensional ones, what needs to be done is to compute the first k eigenvectors ($k \leq n$) which correspond to the k largest eigen values. In order to select the number k , a threshold θ is introduced to denote the approximation precision of the k largest eigenvectors. That is:

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n \lambda_i} \geq \theta \quad (3.4)$$

Given the precision parameter θ , the number of eigenvectors k can be decided [30].

Let $V = [\hat{v}_1, \hat{v}_2, \dots, \hat{v}_k]$, and $\Lambda = \text{diag} [\lambda_1, \lambda_2, \dots, \lambda_k]$. After the matrix \hat{V} is decided, the low-dimensional feature vectors, named PC, of raw ones are determined as follows:

$$\hat{P} = V^T \hat{x}_t \quad (3.5)$$

The PCs of PCA have three properties: (1) they are uncorrelated; (2) they have sequentially maximum variances; (3) the mean-squared approximation error in the representation of the original vectors by the first several PCs is minimal [28,29].

3.2 Independent Components Analysis (ICA)

Independent components analysis is also a multivariate data analysis method which has been developed to deal with the problems like “cocktail party effect” [30, 31]. Cocktail party problem is simply to separate two voice signals independently from their mixtures that are recorded by two microphones in different positions.

Assume that two signals, $x_1(t)$ and $x_2(t)$, are collected from two separate sensors which are the combination of two independent sources, $s_1(t)$ and $s_2(t)$, that is:

$$x_1(t) = a s_1(t) + b s_2(t) \quad (3.6)$$

$$x_2(t) = c s_1(t) + d s_2(t) \quad (3.7)$$

Equation (3.6) and (3.7) are considered as 2-D model and can also be expressed in matrix form:

$$x = A s \quad (3.8)$$

where $x=(x_1, x_2)^T$, $s=(s_1, s_2)^T$, and A is the mixing matrix. By expressing s as:

$$s = W x \quad (3.9)$$

where $W=A^{-1}$, real problem of ICA is estimation of W matrix coefficients. The simple solution to this problem is found by assuming the source signals are statistically independent and nongaussian [30].

There are some methods to estimate coefficient matrix W as maximization of nongaussianity, minimization of mutual information, and maximum likelihood estimation.

3.2.1 Maximization of Nongaussianity

In order to maximize the nongaussianity, two important subjects, kurtosis and negentropy, can be introduced as the measures of nongaussianity.

Kurtosis is a statistical parameter which is called as the forth-order cumulant and it is classically defined in Section 2.1 by Equation (2.4). It is equal to zero for Gaussian variables.

$$kur = \frac{1}{3N\sigma^4} \sum_{i=1}^N (x_i - \mu)^4 \quad (3.10)$$

Using Equation (3.10), kurtosis equals to nonzero for most of the nongaussian random variables [30]. The following equation can be written for two independent random variables, s_1 and s_2 :

$$kur(s_1 + s_2) = kur(s_1) + kur(s_2) \quad (3.11)$$

and,

$$kur(\alpha s_1) = \alpha^4 kur(s_1) \quad (3.12)$$

where α is a scalar parameter.

To illustrate how independent components are found by kurtosis minimization or maximization, the optimization landscape is introduced. The 2-D model in Equations

(3.6) and (3.7) are considered where $x = As$. It is assumed that the signals have unit variances and kurtosis values of s_1 and s_2 , $kur(s_1)$ and $kur(s_2)$, are nonzero. One of the independent component can be defined as $y = b^T x$. If the transformed vector is $q = A^T b$, then, $y = b^T x = b^T A s = q^T s = q_1 s_1 + q_2 s_2$. By using the Equation (3.11) and (3.12):

$$kur(y) = kur(q_1 s_1) + kur(q_2 s_2) = q_1^4 kur(s_1) + q_2^4 kur(s_2) \quad (3.13)$$

Since the assumption of y has a variance value of 1, the second moment (variance) can be written as:

$$E(y^2) = q_1^2 + q_2^2 = 1 \quad (3.14)$$

which tells that the vector q is on the unit circle of the 2-D plane. Finally, the optimization problem is to find the maxima of the below function, that is:

$$|kur(y)| = |q_1^4 kur(s_1) + q_2^4 kur(s_2)| \quad (3.15)$$

By simplifying kurtosis equal to 1:

$$F(q) = q_1^4 + q_2^4 \quad (3.16)$$

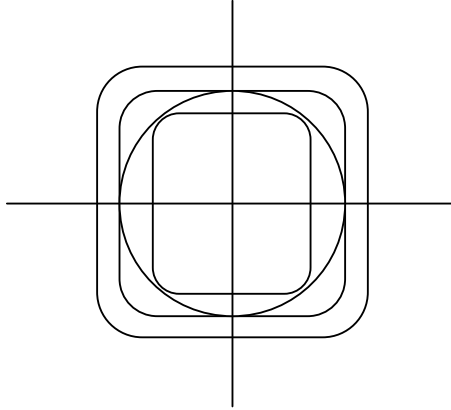


Figure 3.1 : The optimization landscape of kurtosis.

Figure 3.1 shows the curves of some contours of $F(q)$ function which gives the optimization landscape of the problem. The maxima are at the points where one of the elements of the q vector is zero and the other is nonzero. By using the unity circle

constraint, the nonzero element is +1 or -1 which makes y equal to one of the components $\pm s_i$ and the problem is over.

Entropy is defined in Section (2.5). Similarly, the differential entropy H of a given continuous random signal s with the density $f(s)$ is expressed as:

$$H(s) = -\int f(s) \log f(s) ds \quad (3.17)$$

According to information theory, a Gaussian variable has the largest entropy value among all variables of equal variance [32]. A modified version of the differential entropy called negentropy is introduced as follows:

$$J(s) = H(s_{gauss}) - H(s) \quad (3.18)$$

where s_{gauss} is a Gaussian random signal. Negentropy is zero for a Gaussian distribution of s and always non-negative [31].

Cumulant-based measures like kurtosis and entropy-based measures like negentropy define the nongaussianity which help to estimate the independent components. This can be done with gradient methods and fixed point algorithms [31].

3.2.2 Minimization of the Mutual Information

Mutual information is the measure of independence between random variables and works as the same way of finding nongaussian directions. By using the differential entropy, it is expressed as:

$$I(s_1, s_2, \dots, s_n) = \sum_{i=1}^n H(s_i) - H(s) \quad (3.19)$$

where I is the mutual information and s_i , $i=1, \dots, n$ is the random variables. For finding the independent components, an important property of the mutual information [32] can be defined for a linear invertible transform $s=Wx$:

$$I(s_1, s_2, \dots, s_n) = \sum_{i=1}^n H(s_i) - H(x) - \log |\det W| \quad (3.20)$$

If s_i is assumed to be uncorrelated and has a unit variance, $\det W$ must be constant which leads negentropy differ by a constant. That is:

$$I(s_1, s_2, \dots, s_n) = K - \sum_{i=1}^n J(s_i) \quad (3.21)$$

where K is constant and it does not depend on W . Equation (3.21) is quite familiar with the concept of maximizing the nongaussianity. The gradient methods and fixed point algorithm can also used for the minimization mutual information [31].

3.2.3 Maximum Likelihood Estimation

The maximum likelihood estimation which is related with the infomax principle can be used to estimate the independent components. If $W^T = (w_1, w_2, \dots, w_n)$ for A^{-1} , the log-likelihood can be interpreted:

$$L = \sum_{t=1}^T \sum_{i=1}^n \log f_i(w_i^T s(t)) + T \log |\det W| \quad (3.22)$$

where f_i is the density functions of x_i , and $s(t)$, $t=1, 2, \dots, T$ are observations of s .

The infomax principle which is a maximization process of entropy also helps to find the independent components. If s is the input to a neural network, the output equals to $\Phi_i(w_i^T s)$ where Φ_i are well chosen activation functions. The entropy of the outputs:

$$L_2 = H(\Phi_1(w_1^T s), \dots, \Phi_n(w_n^T s)) \quad (3.23)$$

By using the classical entropy formula:

$$H(\Phi_1(w_1^T s), \dots, \Phi_n(w_n^T s)) = H(s) + E\{\log |\det \frac{\partial F}{\partial W}|\} \quad (3.24)$$

where $F(s) = \Phi_i(w_i^T s)$ denotes the activation functions. The derivative expression is simply obtained by:

$$E\{\log |\det \frac{\partial F}{\partial W}|\} = \sum_i E\{\log \Phi'_i(w_i^T s)\} + \log |\det W| \quad (3.25)$$

Equation (3.25) proves that the maximum likelihood equals to infomax.

Gradient methods and fixed point algorithms are also used for the maximum likelihood estimation as the maximization of nongaussianity and the minimization

the mutual information which have quite relative background for interpreting the independent components [31].

3.3 K-Means Clustering

Clustering means organizing and categorizing data, thus it is helpful for data compression and model construction. It is the partition of a data set into groups. The similarity within a group is larger than that among groups.

K-means algorithm divides n vector $x_i, i=1,2,\dots,n$ are into k groups $C_j, j=1,2,\dots,k$ and finds a cluster centroid (center) for each group by minimizing a distance function J .

$$J = \sum_{j=1}^k J_j = \sum_{j=1}^k \left(\sum_{m, x_m \in C_j} (x_m - k_j)^2 \right) \quad (3.26)$$

J_i is the Euclidean distance function within group j in Equation (3.26).

It is an iterative algorithm and an optimum solution cannot be guaranteed. The performance of the algorithm depends on the initialized cluster centroids [33, 34].

4. CLASSIFICATION

The best computer in real life is the human brain. It is considered to be slower than microprocessor based computers, which is a wrong belief. In many applications, human brain is faster than computers. For example, it classifies human voices in less than a second whether it belongs to a man or woman, a child or an adult, it has anger or happy sound etc., but it may take minutes or even hours for an average speed computer to perform such a task.

Computers cannot make any classification unless they are taught how to do it. To overcome this problem, a training set and a classifier are needed. The training set is the group of the feature vectors formed by a specific method depending on the application and it is used for the learning process of selected classifier. A well chosen feature vector set is the most important thing in classification to get the best performance. The classifier makes decisions among classes by using its classification algorithm.

There are several classification methods in pattern recognition related problems because of the variety of applications from financial applications to biomedical operations. Classification provides critical and online information to make decisions strategically.

4.1 Bayesian Decision Theory

Bayesian decision theory is an essential application in statistical pattern recognition. Bayesian classifier chooses the class with highest the posterior probability by combining the prior probability and what the data tells. The posterior probability after having seen the observation, x :

$$posterior = \frac{prior \times likelihood}{evidence} \quad (4.1)$$

Prior is the knowledge of the class before looking the data set, x . Likelihood is the conditional probability that an event belonging to the associated observation value x and the evidence is the marginal probability [35]. The posterior probability where C_i $\{i=1,2,\dots,n\}$ is the number of classes and x is measured values can be calculated as:

$$P(C_i/x) = \frac{P(x/C_i) P(C_i)}{P(x)} \quad (4.2)$$

where,

$$P(x/C_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad (4.3)$$

and,

$$P(x) = \sum_{i=1}^n P(C_i/x) P(C_i) \quad (4.4)$$

Equation (4.3) is for Gaussian distributed independent variables and n shows the number of classes in Equation (4.4).

For a two class case, $P(C_1)$ and $P(C_2)$ shows prior probabilities. $P(C_1/x)$ and $P(C_2/x)$ can be plotted according to Equation (4.2).

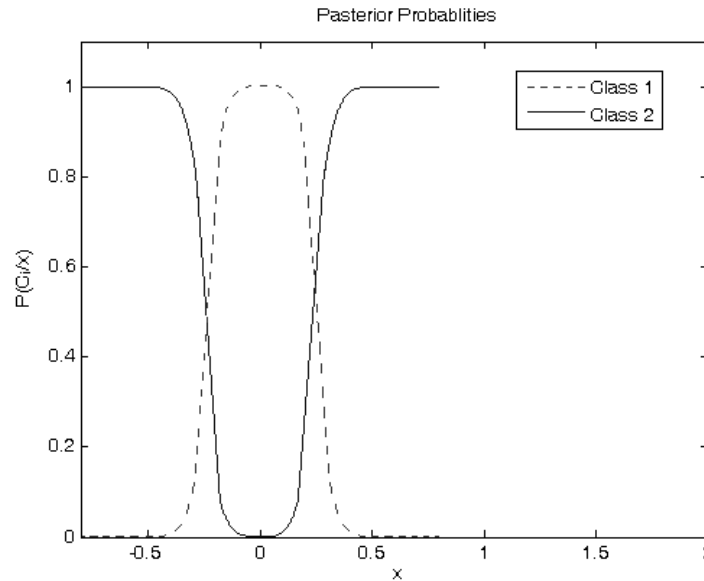


Figure 4.1 : Posterior probability values of two classes.

The intersection points of the posteriors of two classes can be determined easily from Figure 4.1. These points are the threshold for decision. x values in the interval of $x_1 < x < x_2$ and having the probability of $P(C_1/x)$ belongs to the class C_1 ; x values having the probability of $P(C_2/x)$ and for $x < x_1$ and $x_2 < x$ belongs to the class C_2 [12].

4.2 Neural Networks

Artificial neural networks (ANN) are one of the most popular applications among classification methods as they have numerous topologies. They are capable of classifying specific classes among a large data set by having a non-linear structure and decision surfaces.

4.2.1 Multilayer Perceptron (MLP) Neural Networks

Multilayer perceptron neural networks are formed by the number of perceptrons which are called as neurons. Inputs of a neuron are transformed by an activation function like linear, logistic, tangent, hardlim, gaussian...etc, which determines the type of neuron.

A basic perception (see in Figure 4.2) is introduced first by Rosenblatt in 1950's for the binary classification problem [36].

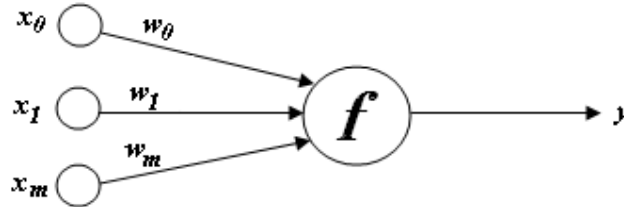


Figure 4.2 : A Simple Perceptron

$x_i \{i=1,2,\dots,m\}$ is the input vector and x_0 is the bias value, usually equals to +1. $w_i \{i=0,1,2,\dots,m\}$ is the weights of neuron and f is the activation function. The output is calculated as:

$$y = \sum_{i=1}^m f(x_i * w_i) + f(w_0) \quad (4.5)$$

Generally, MLP neural networks consist of three major parts called layers which are shown in Figure 4.3. First one is input layer where input vectors are applied to

network. The input layer neurons usually have linear activation functions. The second one is middle layer which is mostly mentioned as hidden layer. There may be more than one hidden layer in a network and each layer has its own activation function. The last layer is output layer where the outputs of network represented. Activation functions can be linear or logistic in the neurons of the output layer [37]. The target values are compared in the output layer and the weights of network are updated by back-propagation (BP) algorithm.

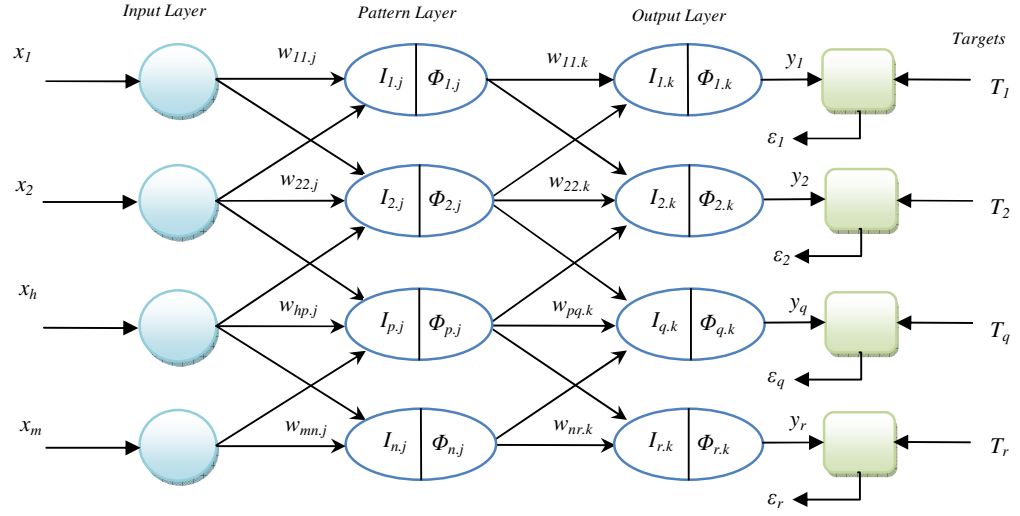


Figure 4.3 : General Structure of MLP Neural Network.

The chain rule is used by deriving the BP algorithm which includes the derivatives of activation functions [38]. Linear, logistic and tangent activation functions which are mostly used in neural network approaches because of their easy derivative forms as represented below.

$$\text{Linear} \quad \Phi(I) = I \quad \dot{\Phi}(I) = 1 \quad (4.6)$$

$$\text{Logistic} \quad \Phi(I) = \frac{1}{1 + e^{-\alpha I}} \quad \dot{\Phi}(I) = \alpha \Phi(I)(1 - \Phi(I)) \quad (4.7)$$

$$\text{Tanh} \quad \Phi(I) = \frac{e^{\alpha I} - e^{-\alpha I}}{e^{\alpha I} + e^{-\alpha I}} \quad \dot{\Phi}(I) = \alpha(1 - \Phi(I)^2) \quad (4.8)$$

Alpha (α) above is the slope parameter and usually chosen as 1 [38].

Back-propagation is the most popular learning algorithm of neural networks among variety of other methods. It is a kind of gradient-decent algorithm where error is fed back through network and the weights are updated during the process. Calculation of the weights in the output layer and the hidden layer are given by appendices A.1 and A.2.

The delta learning rule, which is also called Widrow-Hoff delta rule or least mean square (LMS) rule, is applied to adjust the weights [39]. If it is applied to the simple perceptron shown Figure 4.2 where ε denotes the error and T is the desired output, the change in weight vector is calculated as:

$$\Delta w_i = -K \frac{\partial \varepsilon^2}{\partial w_i} = K 2(T - y) x_i = 2 K \varepsilon x_i \quad (4.9)$$

where K is a constant of proportionality. The minus sign indicates the minimization process [38].

There are also several training algorithms (Levenberg-Marquardt, Quasi-Newton, and Resilient Backpropagation etc.) which have various topologies, training times and learning rules because of the real world applications requires different processes.

4.2.2 Probabilistic Neural Networks (PNN)

Probabilistic neural networks are mostly used for statistical pattern classification problems. It can be also used in real-time applications. PNN is actually a Bayesian classifier in the form of a neural network structure which consists of four layers as it is shown in Figure 4.4 [38, 40, 41].

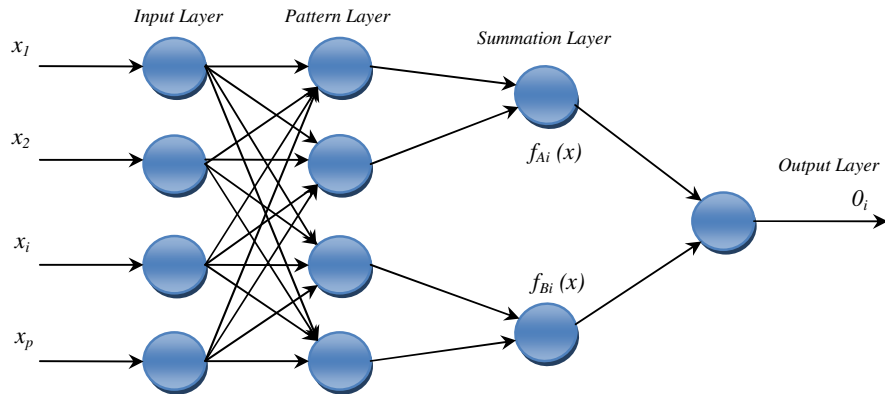


Figure 4.4 : PNN Structure.

The first layer is input layer which is also called “fanout” or “buffer” layer where the inputs are applied. The second is pattern layer where the inputs are multiplied by weights and the weighted sums of the inputs to each neuron are transformed by an exponential activation function which is Gaussian weighting function:

$$\Phi(I_i) = \frac{1}{n} \sum_{i=1}^n \exp[-\|(I - I_i)\|^2 / 2 \sigma^2] \quad (4.10)$$

where I is the weighted input to the pattern layer neurons, I_i is the specific case in a class, n is the number of classes, and σ is the smoothing parameter that defines the decision boundary surfaces of classes [38, 41].

The third layer is summation layer where the sum of the outputs of the pattern layer is transmitted to a single neuron. The weights equal to 1 and there is only one neuron for each class. The last layer is output layer which receives the outputs of the summation layer neurons and decides the class by producing a binary output [38].

The weights of the output layer are calculated as:

$$w_i = -[h_A / h_B][l_A / l_B][n_A / n_B] \quad (4.11)$$

where h refers to the prior probabilities, l is the losses which are associated with the inaccurately identified patterns, and n is the number of the training patterns from each class. The training process is completed by setting the training patterns as the weights of the pattern layer and connecting the pattern layer to the summation layer [38].

4.2.3 Kohonen Neural Networks

Kohonen neural networks, which are also called self organizing maps (SOM), can also be applied to pattern classification problems. It has two layers, input layer is where the inputs are represented and Kohonen layer is where the weights are placed with the labeled neurons of each class, as shown in Figure 4.5.

The weights in Kohonen layer are generated randomly in order to the dimension of input vector with a structure as grid, random, or horizontal. Figure 4.5 shows a grid Kohonen layer with three labeled classes in different colors.

When an input vector element is applied to Kohonen network, Euclidian distances between the input and each of the weights are calculated. The weight having the

closest distance to the input element is the winner neuron and its neighbors are updated as:

$$\Delta w_i = \eta [x_i - w^{old}] \quad (4.12)$$

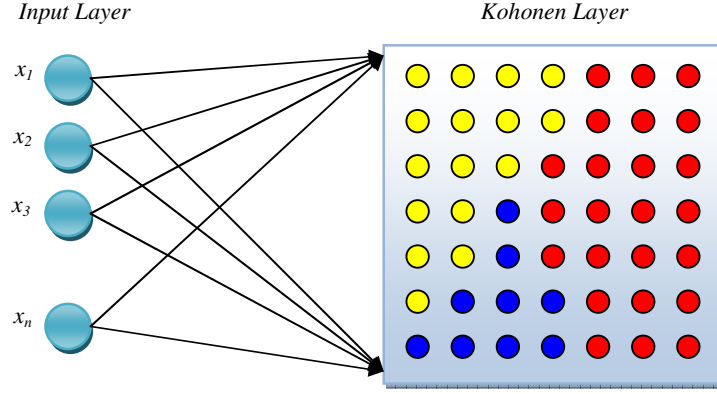


Figure 4.5 : Kohonen Structure.

where η the learning constant which varies from 0 to 1, x_i is the input vector and w_i denotes the weight vector. After each input vector elements are applied to the network, the training process ends and the weights form clusters due to the classes [38].

The training process is followed by a labeling procedure in which the clustered weights are named by each class. The labeling of Kohonen network allows dealing with classification problems. The training vectors from each class are again applied to the trained network and the winner neurons are marked with classes. The weights having the highest number of marked class are attended as that class label. After completing the labeling procedure, if any test vector is applied to the network, Euclidian distances are calculated and the input vector element is attended as the name of the winning neuron's class.

Mean quantization error (MQE) of SOM is a quality index for the vectors applied to the trained Kohonen map. It is the average distance between each vector and its best matching unit on the map. Let SOM consists i units and each unit i is assigned by a weight vector w_i . If a data x_k is applied to the map, its best matching unit is calculated as:

$$c_k = \arg \min_i \|x_k - w_i\| \quad (4.13)$$

Mean quantization error of the map is calculated as:

$$MQE_i = \frac{1}{U_i} \sum_{k \in U_i} \|x_k - w_i\|, \quad U_i = \{k | c_k = i\} \quad (4.14)$$

U -matrix is $U=u_{ij}$ where u_{ij} is the distance between two neighboring weights [Url-3].

Mahalanobis distance can also be used to calculate the quality of Kohonen map. It is the similarity of an unknown data set to a known one. It differs from Euclidian distance and takes into account the correlations of a given data set. Let x and y be two random vectors, Mahalanobis distance between x and y is expressed as:

$$d(x, y) = \sqrt{(x - y)^T P^{-1} (x - y)} \quad (4.15)$$

where P is the covariance matrix [42].

5. APPLICATION

In this section, various pattern recognition methods are combined to track the bearing failure of an induction motor of 5 HP, which is subjected to the accelerated aging process.

5.1 Induction Motors

Induction motors (Figure 5.1) are the most popular type of alternate current (AC) motors because it is easy to operate them. They do not have a separate field circuit and they work as a rotating transformer to induce voltages and currents in their field circuits. Their equivalent circuits are quite similar to transformers, except for the effects of varying speed [43].



Figure 5.1 : Induction Motor

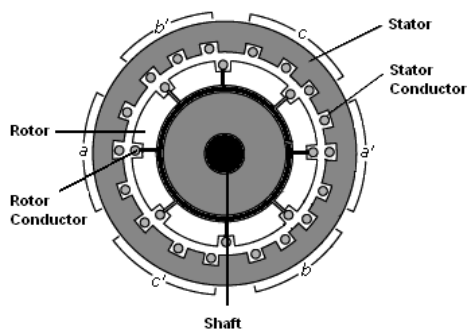


Figure 5.2 : Induction Motor Cross-Section

If a three-phase voltage is applied to the stator windings, three phase of stator currents starts to flow which produce a magnetic B_s . The speed of the rotating magnetic field:

$$n_{sync} = \frac{120f_e}{P} \quad (5.1)$$

where f_e is the frequency of the voltage and P is the poles of the induction machine. A voltage is induced in the rotor windings (or conductor as in Figure 5.2) due to the magnetic field B_s . That is:

$$e_{ind} = (v \times B) \cdot l \quad (5.2)$$

where v is the velocity of the rotor bar, B is the magnetic flux, and l is the length of the conductor in the magnetic field. The rotor voltage induced by the relative motion between the rotor and the stator magnetic field produces a rotor current, and that rotor current flow produces a stator magnetic field B_r . The induced torque in the machine:

$$\tau_{ind} = k B_r \times B_s \quad (5.3)$$

which accelerates the motor [44].

5.2 Bearing Fluting Mechanism and Experimental Set-up

In recent years, it is seen that a common cause of the bearing damage is shaft currents produced by power electronics drivers used in most industrial applications [45]. For this reason, a bearing fluting mechanism or Electric Discharge Machining (EDM) approach was considered to show the results of this event with an experimental study [46]. That experimental accelerated aging and fault diagnosis study of induction motors was held in the University of Tennessee, Nuclear Engineering Department by A. S. Erbay and B. R. Upadhyaya in 1999. The section 5.2.2 presents the experimental set-up in detail.

In this study, vibration data was considered from the accelerated aging experiment in [46] to implement new methods for condition monitoring and fault diagnosis of rotating machinery.

5.2.1 Bearing Fluting Mechanism

The rotor is supported by bearings with a grease film that is not conductive. At high speeds, an even distribution of the grease film exists between the rolling elements and the inner and outer races. The rotor voltage can increase with respect to ground. When this voltage builds to a level capable of breaking down the grease film, arcing occurs and a discharge mode current flows through the bearing elements. At low speeds, the grease film is minimized. The balls often make contact with the race, and the rotor voltage does not build up. The current flows through the bearing in a conductive mode. The bearing current thus has two modes, conduction and discharge. Conduction mode bearing currents exhibit continuous flow through the bearings. This form of bearing current does not result in premature bearing failure because current flows continuously without arching. Discharge mode bearing currents occur at random when the grease film momentarily breaks down. The damage mechanism may show pitting, fluting, and other forms of damage [47].

5.2.2 Experimental Setup for Aging Mechanism

In order to simulate the electrical discharge from the shaft to the bearing, a special test setup was designed [46]. A schematic presentation of this Electrical Discharge Machining (EDM) for the bearing elements is shown in Figure 5.3. At each aging cycle, the motor run at no load condition for 30 minutes, with an externally applied shaft current of 27 Amperes at 30 Volts AC which are the IEEE standard values [48].

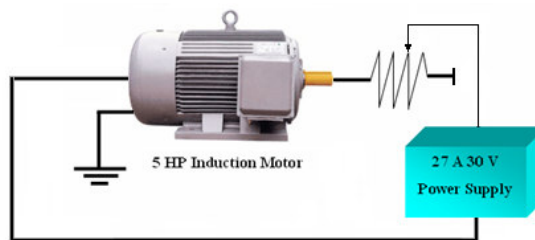


Figure 5.3 : Schematic of the electrical motor bearing EDM setup.

The EDM aging was followed by thermal aging in order to accelerate the aging process. After each cycle of accelerated aging, the test motor was put on a motor performance test platform. From the experimental setup, high frequency data with a sampling frequency of 12 kHz was acquired for the motor currents and voltages,

rotor speed, torque and six vibration measurements. There are eight measurement sets so that one healthy and seven aged cases.

The motor load test system is shown in Figure 5.4. Figure 5.4(a) shows the experimental setup for processing the measured electrical and vibration data and then transferring them to a host personal computer. There are six accelerometers used in this set-up for independent vibration measurements. As shown in Figure 5.4(b) sensors 6 (Short-End Horizontal), 7 (Short-End Axial), and 8 (Short-End Vertical) are placed in plane A-A', and sensors 9 and 10 are placed in plane B-B'. The placements of accelerometers 9 (Process-end 10) and 10 (Process-end 2) are shown in Figure 5.4(c). The last accelerometer 5 is on the cover and it is pointed in Figure 5.4 (a).

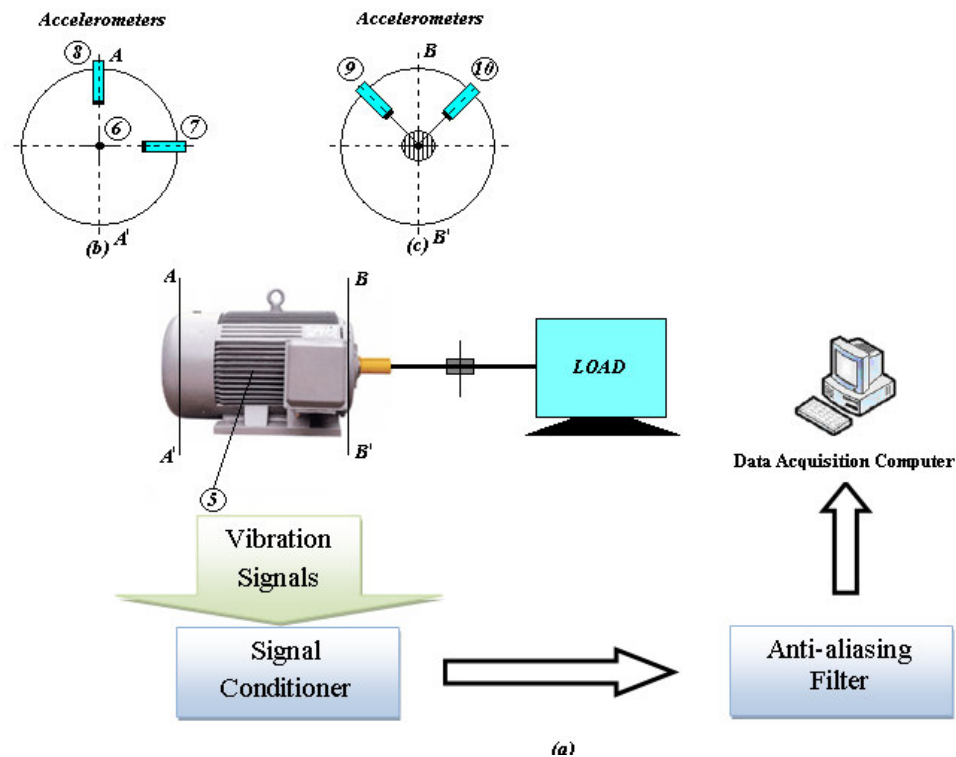


Figure 5.4 : Motor load testing and data acquisition system: a) Experimental set-up configuration; b) Accelerometers on Cross-section (A-A') at short end; c) Accelerometers on Cross-section (B-B') at pulley end.

For further information, nameplate information of 5-HP induction motor is given in following table, Table 5.1.

Table 5.1 : Nameplate Information of the 5-HP Motor [46]

Type of Motor	Premium Efficiency Motor
Manufacturer	US Electrical Motors Division of Emerson Electric Co. St. Louis, MO
Model Number	7965B
Horse Power	5
Motor speed (rpm)	1750
Phase	3
Frame number	184T
Frame Type	TCE
Volts	230 / 460
Amperes	13.0 / 6.5
Service Factor	1.15
Hertz	60
NEMA Design	B
Insulation Class	F
Enclosure	TE
Maximum Ambient Temperature	40 °C
LR KVA Code	J
Duty	Continuous
Power Factor	82.5
Efficiency	90.2
Maximum KVAR	1.4
Shaft End Bearing Type	6206-2Z-J/C3
Drive End Bearing Type	6205-2Z-J/C3
ID #	Z08Z177R190F /

5.2.3 Vibration Analysis for Bearing Damage Detection

The vibration analysis for the rotating machinery diagnosis is very popular because bearing damages usually reveal high frequency components which are due to the frosting on the surface of the bearing elements and it is easy to see the defects by spectral domain analysis [46].

For vibration analysis, accelerometer 10 (at process-end 2) was chosen because it is closer to the aged bearing. Figure 5.5 shows the healthy and the faulty case spectra of the bearing vibrations from accelerometer 10 for the accelerated bearing damage at 100% load conditions. The high frequency components which occurred after the bearing damage can easily be seen from Figure 5.5 (b) between 2-4 kHz. These high frequency band components give the characteristic behavior of the bearing damage for the vibration data.

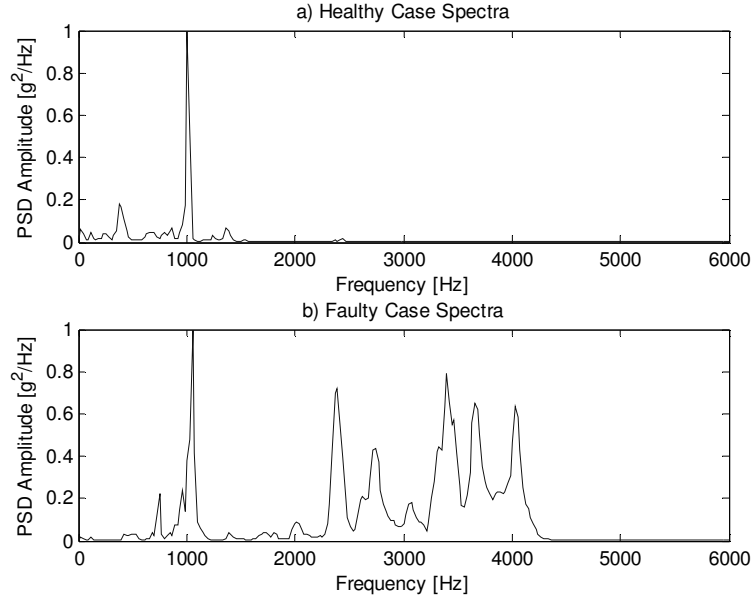


Figure 5.5 : Spectra of the Motor Vibration Data from Sensor 10 for the Accelerated Bearing Damage a) Healthy Case b) Faulty Case

5.2.4 Condition Monitoring Studies for Mixed Load Cases

The existence of bearing damage is a random process. Real purpose of condition monitoring is to capture a trend of aging for different applications.

Online condition monitoring can access the information real time and measured signal is processed without having a delay from system. In this study, variance monitoring, entropy monitoring, supervised and unsupervised networks (MLP and SOM) are used to track the developing of the accelerated aging process.

In this manner, the induction motor was subjected to five different load conditions in each aging cycle. Loads applied in each cycle are %0-25-50-75-100 respectively. Representation of vibration data structure is demonstrated in Figure 5.6.

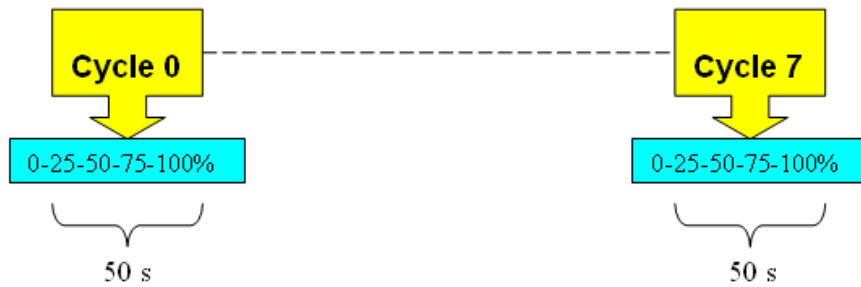


Figure 5.6 : Vibration data structure

For each load the data is taken 10 seconds which means 50 seconds for one aging cycle. Vibration data was used in real time and measured in every 0.5 seconds to be analyzed for mixed load conditions. Bearing was completely damaged during the aging cycle 6 (at 330.5th second) at % 75 load.

In the following sections, different monitoring approaches are combined with pattern recognition techniques to track a good trend of the bearing failure.

5.2.4.1 Condition Monitoring Using Statistical Parameters and Entropy

Five statistical parameters (mean, standard deviation, variance, skewness and kurtosis) and entropy value are calculated in every 0.5 second for mixed load conditions to monitor the bearing failure status. In Figure 5.7, monitoring results of five statistical parameters and entropy values are shown. Mean, standard deviation, variance values make a peak in 330.5th second when the complete bearing damage occurs while entropy has its minimum value. Skewness and kurtosis values are around 0 and 3 respectively before just the bearing damage.

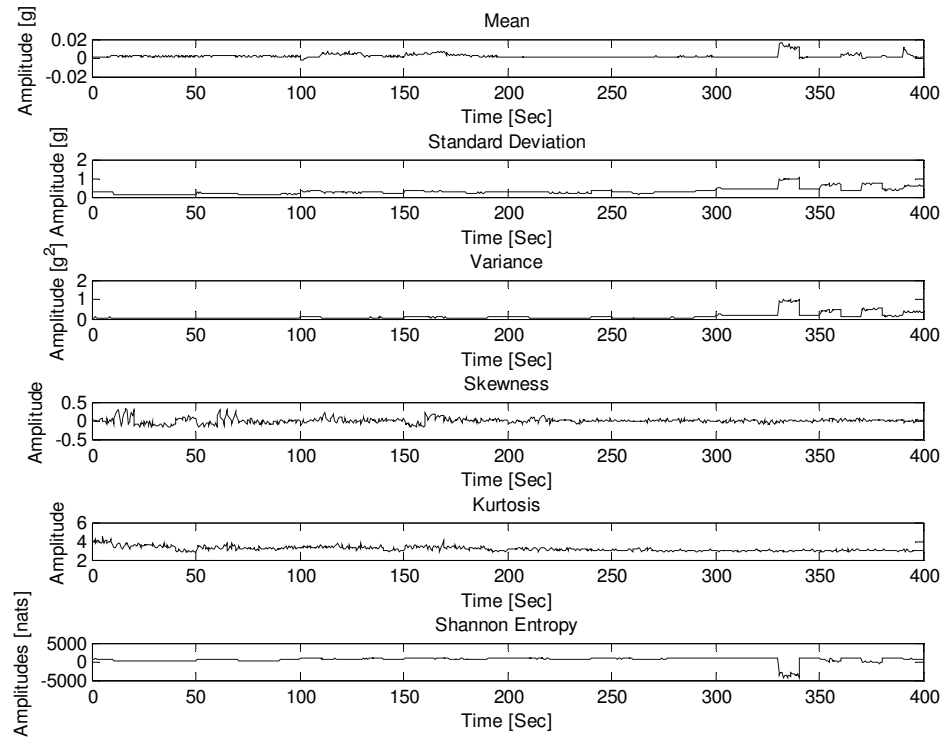


Figure 5.7 : Monitoring Results of Statistical Parameters and Entropy of Vibration Sensor 10

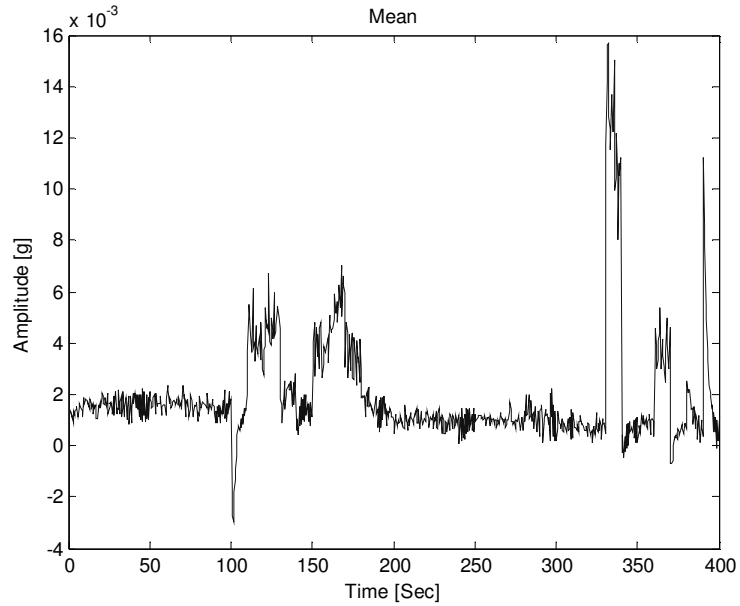


Figure 5.8 : Monitoring Mean Values of Vibration Sensor 10

The mean values are almost the same for the first two cycles 0 and 1 as shown in Figure 5.8. Aging starts to increase the vibrations of 2nd and 3rd cycles. The actual damage is observed in the 6th cycle, 330.5th second.

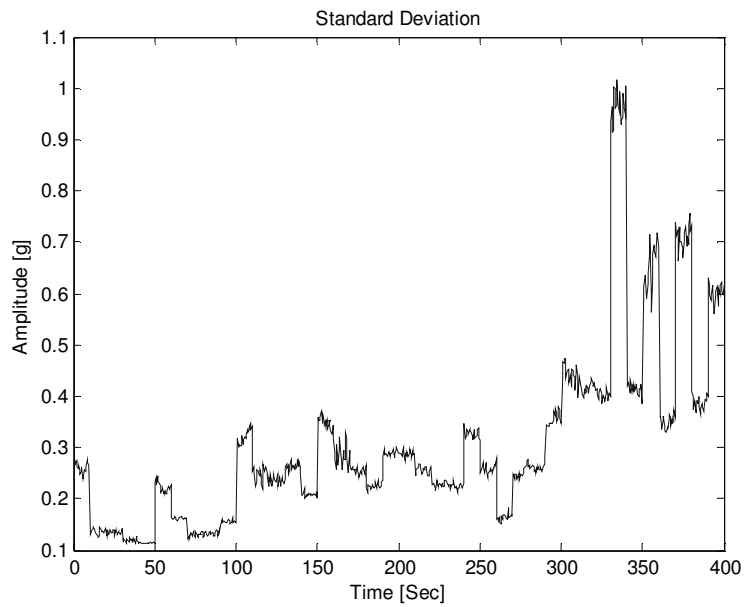


Figure 5.9 : Monitoring Standard Deviation of Vibration Sensor 10

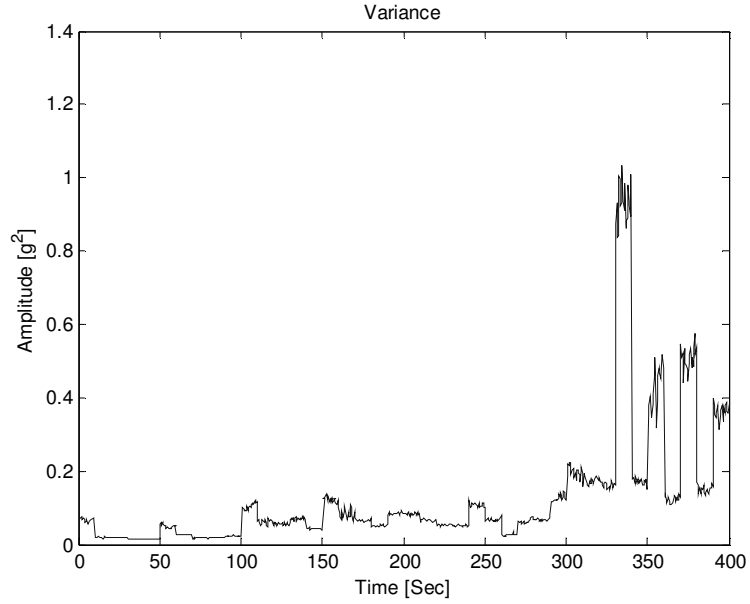


Figure 5.10 : Monitoring Variance of Vibration Sensor 10

From Figures 5.9 and 5.10, same aging trend for the bearing failure is observed, that is, amplitudes rise as the aging time increase. For the first 4 cycle which are 0, 1, 2, and 3 respectively, amplitudes of no-load conditions are higher than the loaded conditions. This trend is disordered in cycle 4 and 5 and the complete damage is observed in cycle 6, 330.5th second.

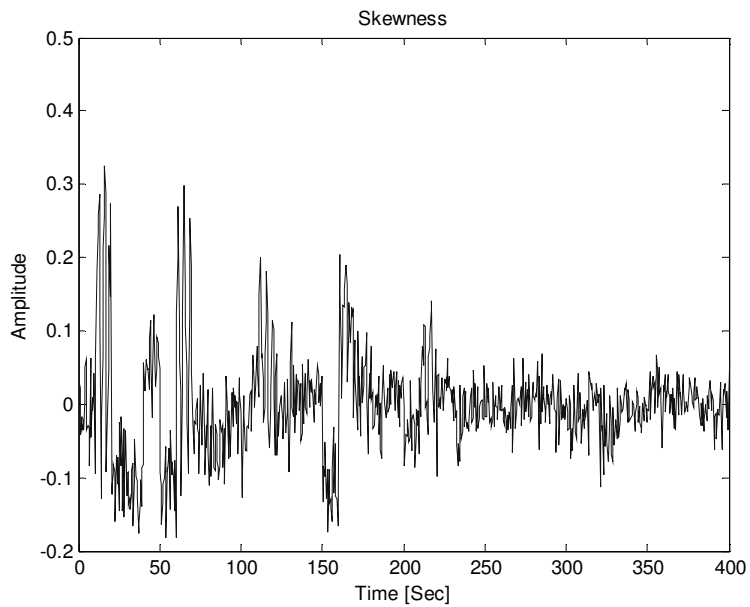


Figure 5.11 : Monitoring Skewness of Vibration Sensor 10

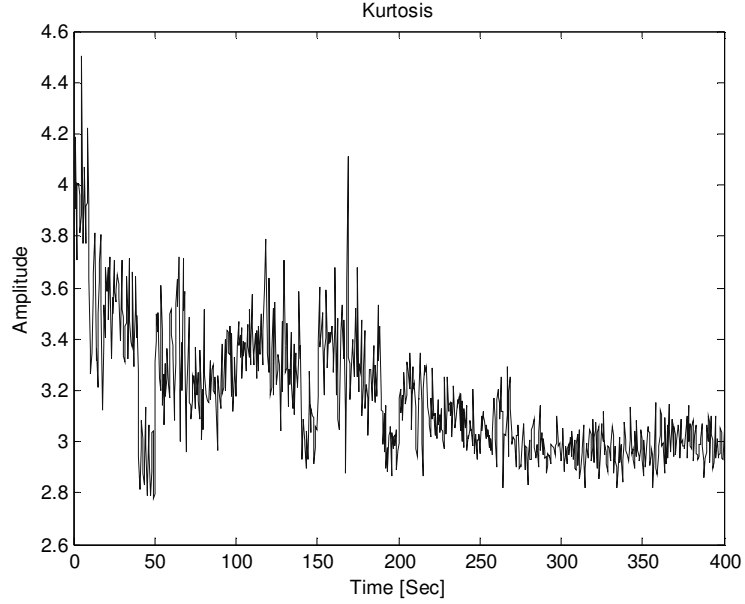


Figure 5.12 : Monitoring Kurtosis of Vibration Sensor 10

It seems like big oscillations in the amplitudes of both skewness and kurtosis in Figure 5.11 and 5.12. As the aging cycle increase, oscillations in kurtosis and skewness decrease having the values 3 and 0 respectively. From this results it can be interpreted that signal becomes Gaussian as the damage occurs. Moreover, there is no clear sign of the bearing failure in the cycle 6, 330.5th second except Gaussianity of an aged signal.

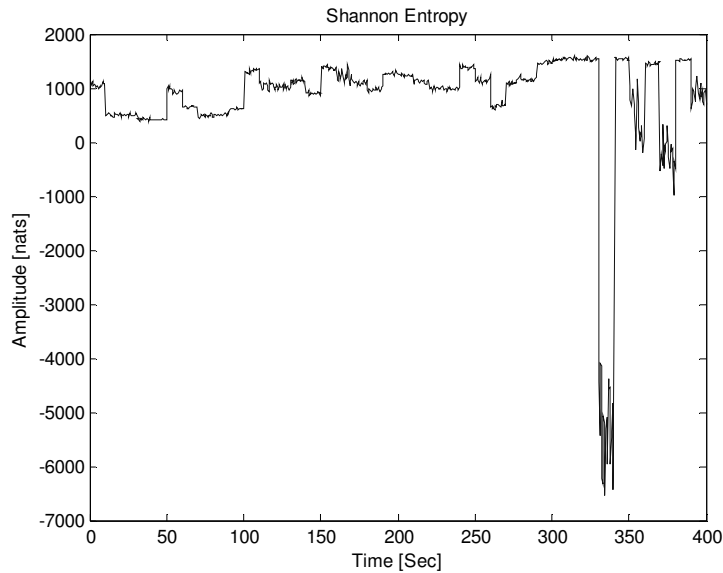


Figure 5.13 : Monitoring Entropy of Vibration Sensor 10

In Figure 5.13, amplitude of entropy raises as the aging time increase. For the first 4 cycle, amplitudes of no-load conditions are higher than loaded conditions and this trend is disordered in cycle 4 and 5 which is also observed from standard deviation and variance values. Actual damage was seen in cycle 6, 330.5th second with a big down fall of the entropy values.

5.2.4.2 Condition Monitoring Using MLP Output Error

A MLP neural network was used for monitoring of the incipient bearing damage. It was trained with the healthy vibration signals of the induction motor for mixed load conditions. It was observed that error between the output of the trained network and real values could track a good trend for accelerated aging of the bearing.

In this manner, vibration signals were transformed with the Fast Fourier Transform (FFT) to the frequency domain, and their power spectral densities (PSD) were determined as the inputs of the network. The target values were taken as variance and Shannon's entropy.

Figure 5.14 shows the training scheme of the MLP neural network.

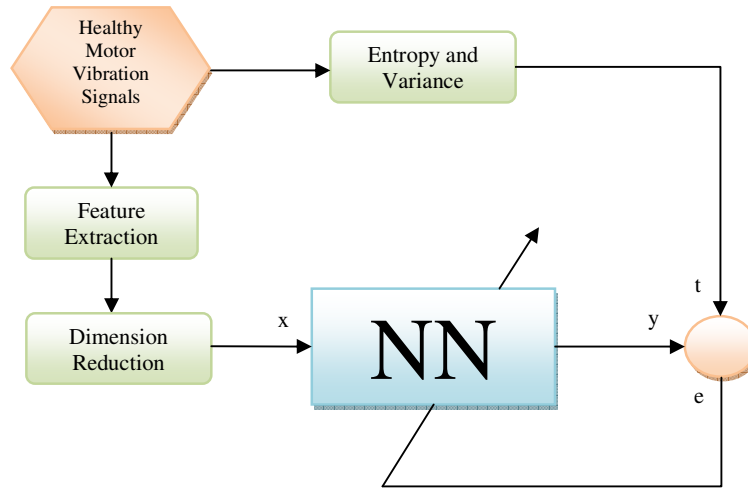


Figure 5.14 : MLP Training Scheme

Vibration signal's data length was 50 seconds for training data consisting of mixed load conditions in healthy motor. They were transformed to frequency domain with the number of FFT (NFFT) points equal to 128 and feature vectors were formed by 65 PSD coefficients. Dimensionality of the feature vectors were reduced to 10 from

65 by principal components analysis (PCA). Tests were showed that reducing the dimension did not affect the performance of the results.

The trained network was used for condition monitoring as it is seen on the Figure 5.15. Only one hidden layer was used with 50 nodes having logistic activation functions. At the output node, the linear activation was used. For the training algorithm, `trainlm` and `traingdx` were chosen in MATLAB environment which are Levenberg-Marquardt backpropagation and gradient descent with momentum and adaptive learning rate backpropagation respectively. They both gave similar results but the `trainlm` was mostly applied in experiments because of its fast training time. After the training, whole eight cycles of motor vibration data with mixed load conditions were applied to the network in each 10 seconds. It was observed that the error between output and real values captured a good aging trend as the accelerated aging process time increases.

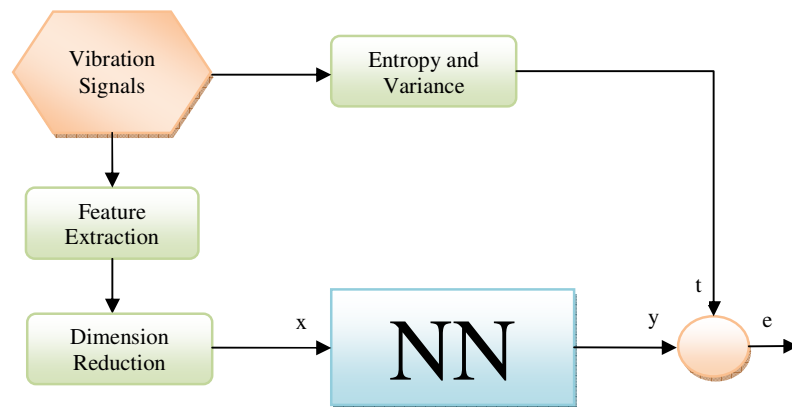


Figure 5.15 : Monitoring System with MLP

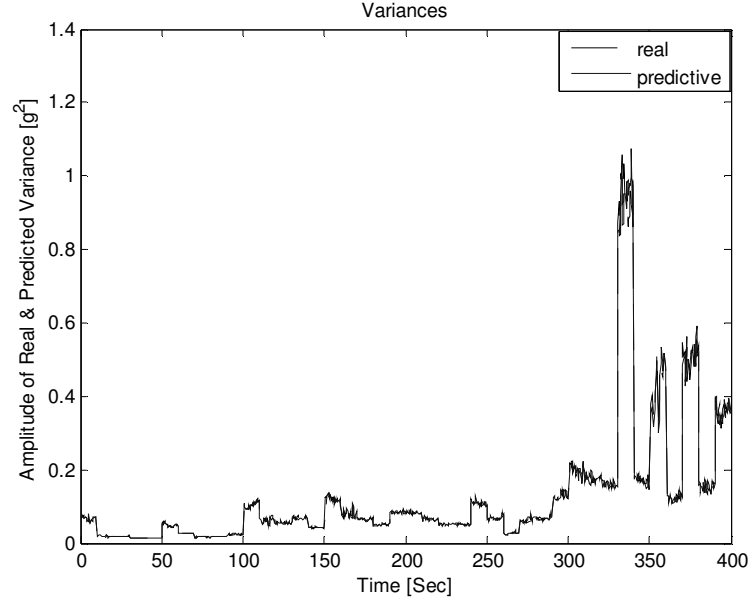


Figure 5.16 : Real and Predictive Variance Values for Mixed Load Conditions

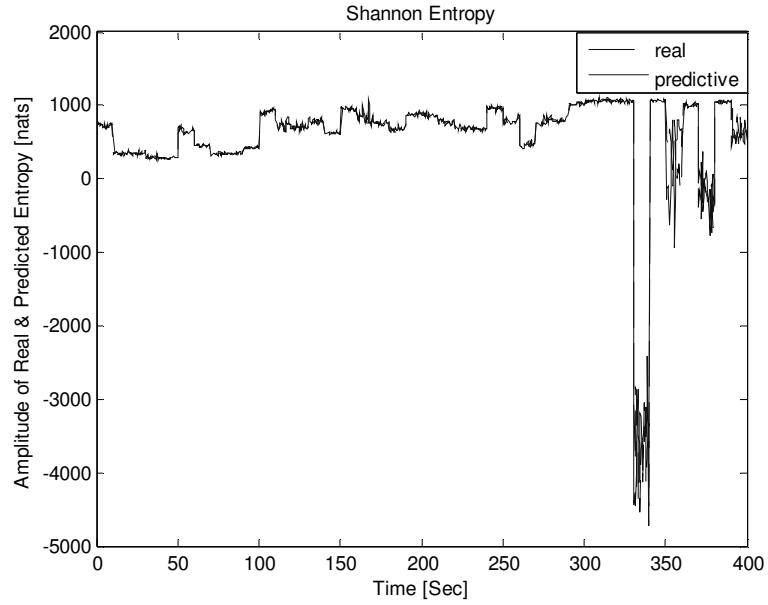


Figure 5.17 : Real and Predictive Entropy Values for Mixed Load Conditions

MLP output is compared with the real variance and entropy values in Figure 5.16 and 5.17 for mixed load case. It is seen that the predicted and real values are exactly the same. As the bearing damage starts, the difference between the real and predicted variance and entropy values begins to increase.

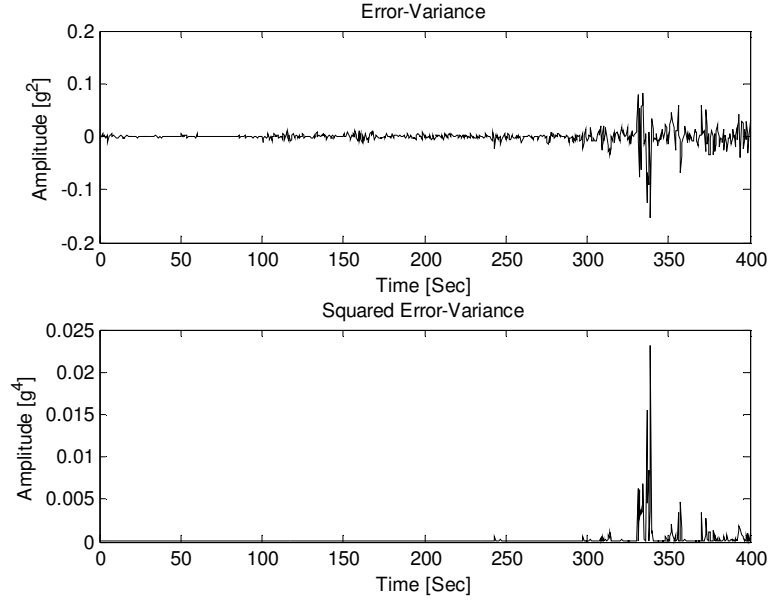


Figure 5.18 : Error and Squared Error Between Real and Predictive Variance Values for Mixed Load Conditions

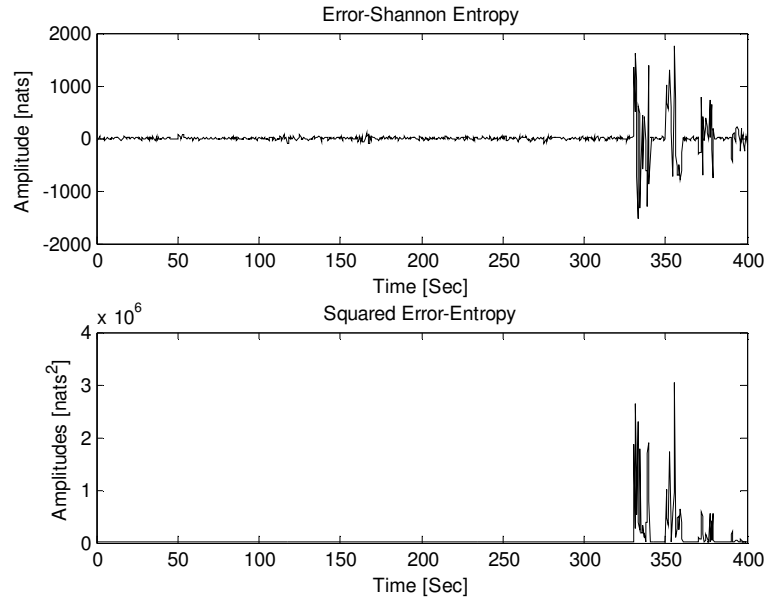


Figure 5.19 : Error and Squared Error Between Real and Predictive Entropy Values for Mixed Load Conditions

Figure 5.18 and 5.19 shows error values at the output of the MLP neural network. Variance and entropy errors rise instantly at the 330.5th second when the bearing damage occurs. Also oscillations in variance error can be used for alarm level information which is defined at the start of the cycle 4.

5.2.4.3 Condition Monitoring Using ICA Based on WPD

The usage of applying ICA to one dimensional data is impossible. So that the vibration signals in healthy case with all load conditions were subjected to the wavelet packet decomposition assuming that the vibration signals were taken from different sensors. Three level WPD consisting of high frequency band was made and variances in each frequency band were calculated as features.

From ICA, three independent components (ICs) and the mixing matrix W were estimated in healthy case which may also be used for dimensionality reduction purposes. WPD was then applied to the rest of the signal in every 0.5 seconds and variance values were calculated in each frequency band. By multiplying the mixing matrix W and variance values of each band, a good aging trend was captured. The aging trend is shown in Figure 5.20, 5.21 and 5.22.

Three estimated ICs are exactly the same ignoring the amplitudes and negative values of the first IC. Three ICs follow a good aging trend without disturbance of the changing load conditions and can be used for monitoring. Bearing damages starts around cycle 4-5 and actual fault can be seen in cycle 6, 330.5th second.

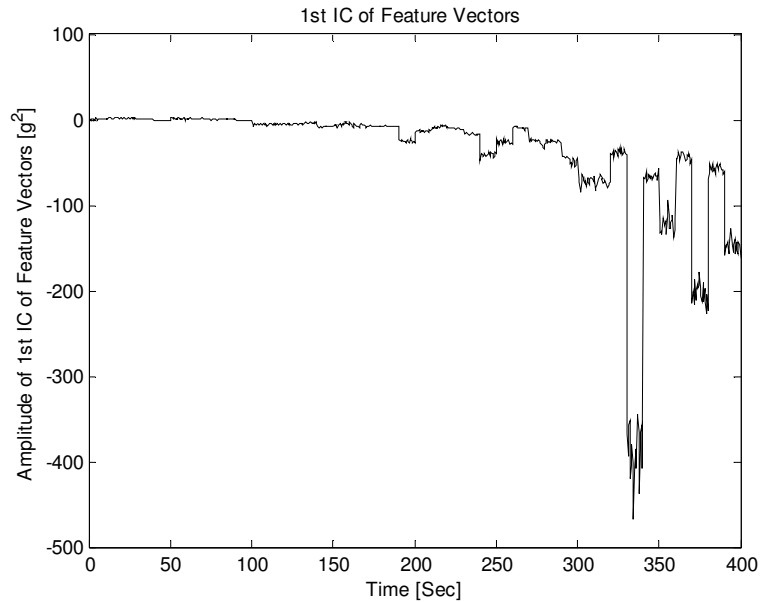


Figure 5.20 : Estimated 1st IC.

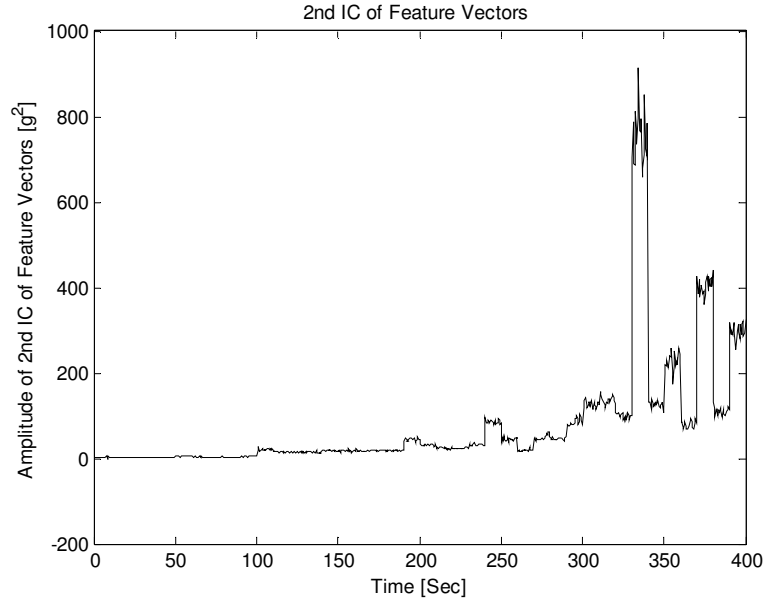


Figure 5.21 : Estimated 2nd IC.

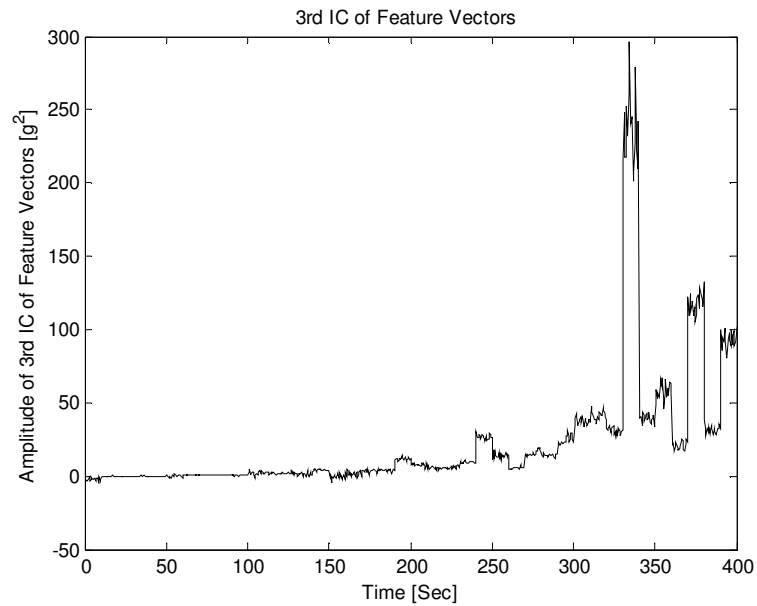


Figure 5.22 : Estimated 3rd IC.

5.2.4.4 Condition Monitoring Using SOM

Kohonen's SOM which is unsupervised network is used for monitoring the bearing damage. It is trained for healthy motor vibration data at mixed load conditions as in Figure 5.23.



Figure 5.23 : Training of SOM

For feature extraction method, WPD is applied to vibration data. Variance of each frequency band was calculated to form feature vectors. Dimension of the feature vectors was reduced to 3 from 15 by ICA. After training SOM with healthy motor signals for mixed load conditions, whole motor data was monitored real time by analyzing data in each 0.5 second.

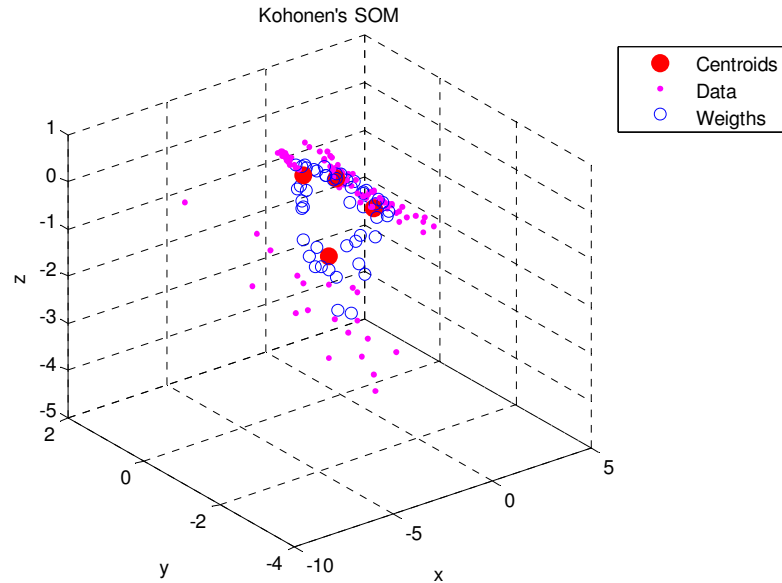


Figure 5.24 : Training Results of SOM

Training results of SOM is seen from Figure 5.24. Training data with the size of 100x3 denoted as small dots. Map size determined as 6x8, and batch training algorithm was used to update the weights. Updated weights are indicated by circles. K-means algorithm was applied to the updated weights to determine the cluster centroids of the map. The centroids are denoted as big dots in Figure 5.24. After that the whole motor data was applied to SOM. Mean squared error and Mahalanobis distances to the determined centroids were calculated.

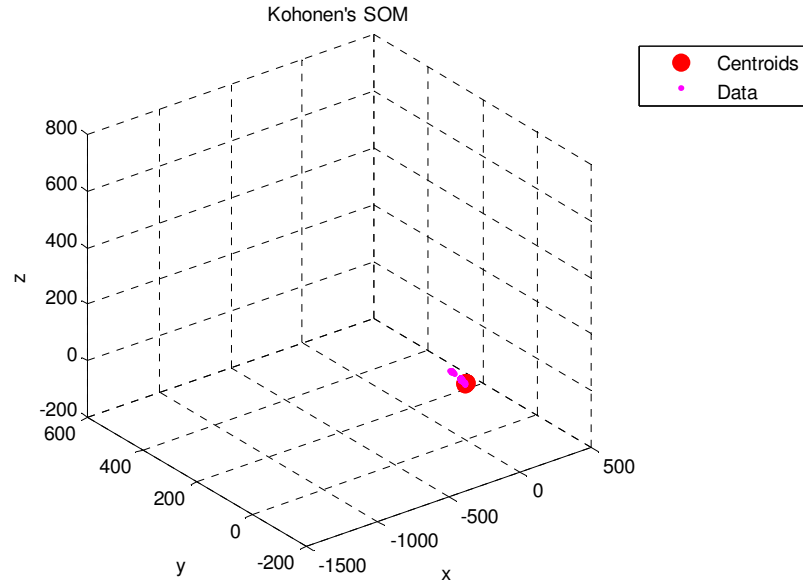


Figure 5.25 : Test Results of SOM for the first 4 cycles.

Distances from cluster centroids are seen in Figure 5.25, 5.26, and 5.27. As the bearing damage occurs, processed data goes further from centroids. Figure 5.24 shows the first 4 cycle data and 4 cluster centroids.

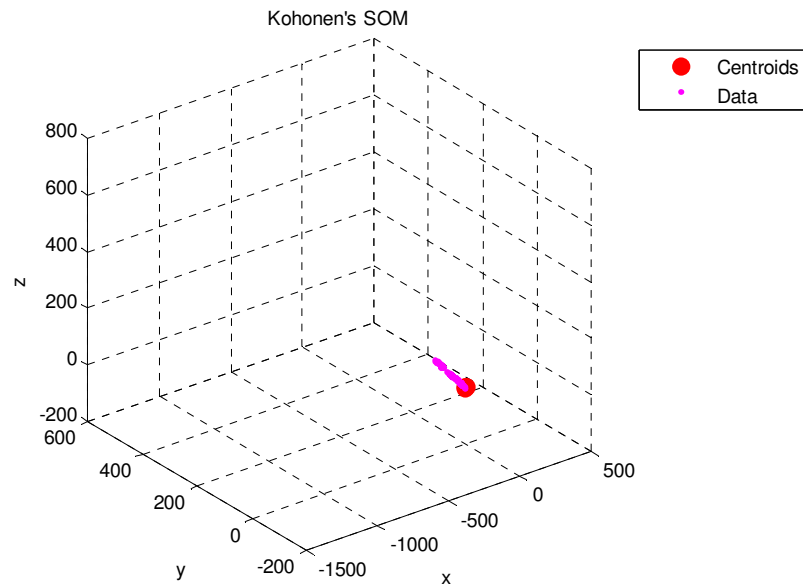


Figure 5.26 : Test Results of SOM for the first 6 cycles.

Figure 5.26 shows the first 6 cycle data and 4 cluster centroids.

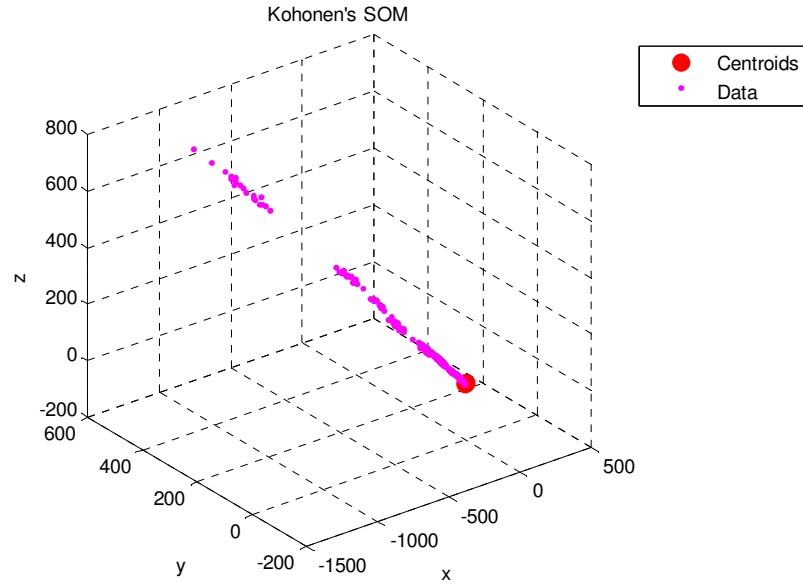


Figure 5.27 : Test Results of SOM for all cycles.

Figure 5.27 shows all eight cycle data and 4 cluster centroids. Actually damage is observed in cycle 6 and it is clearly seen that data goes further from the centroids of healthy case as the damage occurs.

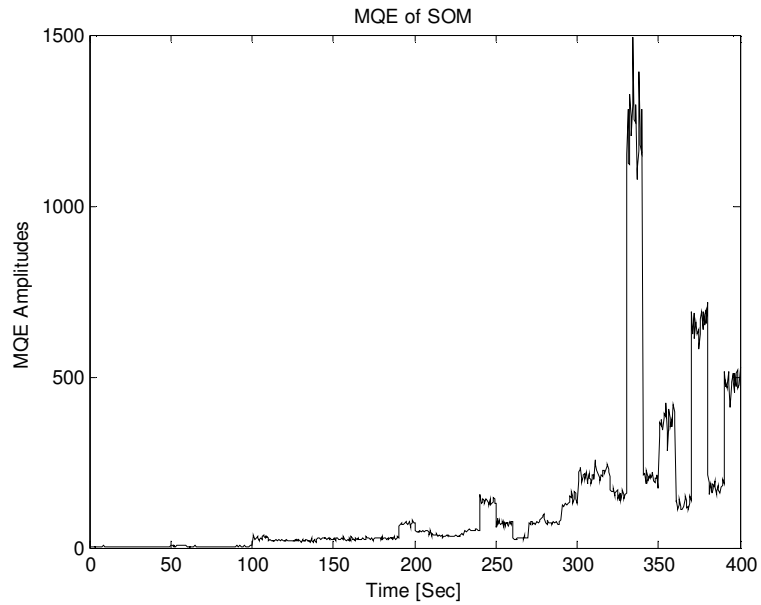


Figure 5.28 : Mean Quantization Error of SOM

In Figure 5.28, MQE has a better trend than variance tracking. Bearing damage starts around 4-5 cycles, and complete bearing failure is seen in cycle 6, 330.5th second.

Figure 5.29 shows MQE in logarithmic scale. As the aging increases, the oscillations stops and amplitudes rise. MQE is maximum when the bearing damage occurs.

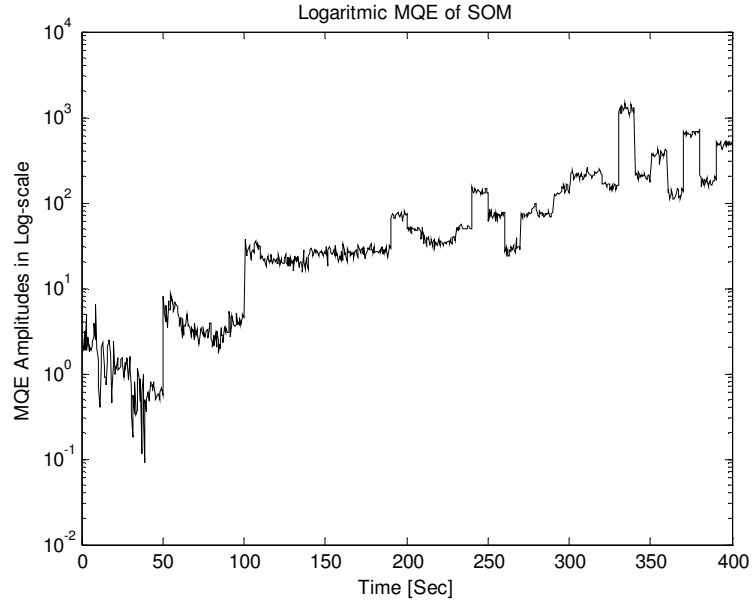


Figure 5.29 : Mean Quantization Error of SOM in Logarithmic Scale

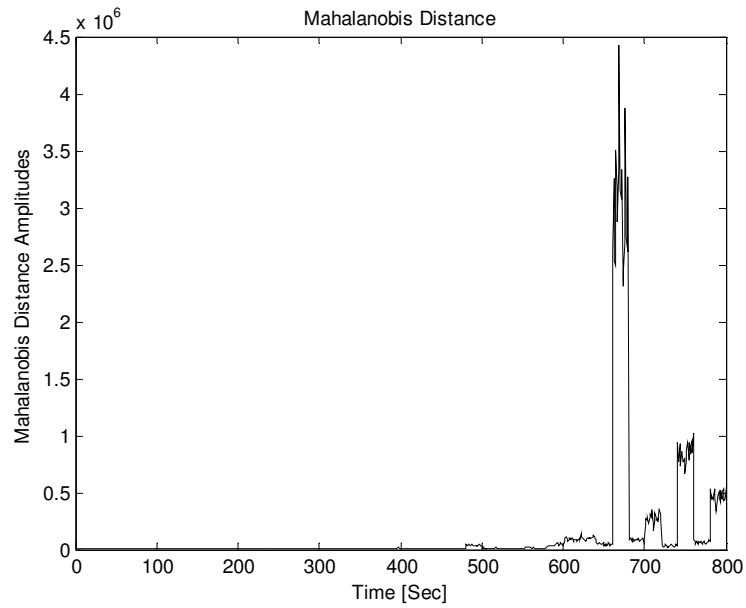


Figure 5.30 : Mahalanobis Distance

By observing Mahalanobis distances, the bearing damage is tracked in cycle 6, 330.5th second in Figure 5.30. Figure 5.31 shows the Mahalanobis distances in logarithmic scale. Like MQE, the oscillations stops and amplitudes rise as the aging

increases. The highest amplitude occurs in cycle 6, 330.5th second denoting the bearing damage.

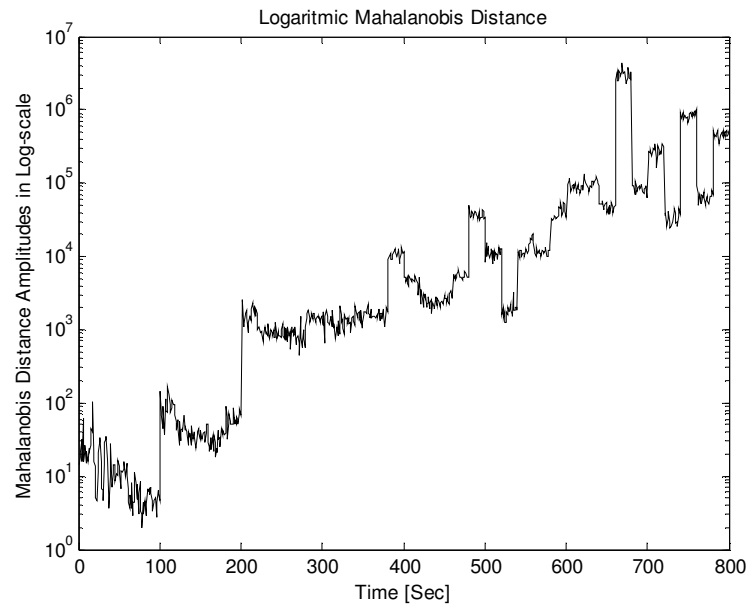


Figure 5.31 : Mahalanobis Distances in Logarithmic Scale

6. CONCLUSION

Different methods using pattern recognition techniques were combined to capture an informative aging trend of the bearing failure and to make life predictions for the predictive maintenance information.

Comparison of the applied methods are illustrated in Table 6.1 according to the best performance in MLP output error of variance and listed below:

Table 6.1 : Comparison of Monitoring Methods (X = NOT, $\sqrt{}$ = OK)

Monitoring Methods	Disturbance of Load Changes	Prediction of Bearing Damage	Bearing Damage Detection
Mean	X	X	$\sqrt{}$
Standard Deviation	$\sqrt{}$	X	$\sqrt{}$
Variance	$\sqrt{}$	X	$\sqrt{}$
Skewness	X	X	$\sqrt{}$
Kurtosis	X	X	$\sqrt{}$
Entropy	$\sqrt{}$	X	$\sqrt{}$
ICA	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$
MLP Error (Variance)	X	$\sqrt{}$	$\sqrt{}$
MLP Error (Entropy)	X	X	$\sqrt{}$
SOM-MQE	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$
SOM-Mahalanobis Distance	X	X	$\sqrt{}$

- From all methods bearing damage was exactly detected at the same time and cycle, that is, cycle 6th 330.5th second.
- Mean values did not give any informative aging trend.
- Standard deviation was affected by the changing load conditions and gave a bad aging trend.
- Variance was also affected by the changing load conditions but the trend is better than of standard deviation.

- By monitoring kurtosis and skewness, motor vibration signals were observed to gain Gaussian characteristics as damage occurs. Non-Gaussian behaviour of vibration signals in early cycles skipped to Gaussian characteristic as the damage developed. As considering last cycles where the damage occurred, vibration signals were completely Gaussian without any disturbance of changing load conditions. This behaviour has been noted to be studied deeply as a future work.
- Entropy had a good aging trend as well as variance and it was also affected by the change of load conditions.
- MLP output error tracked the bearing damage using variance and entropy. Variance error captured the best trend of aging by rise of oscillations as the cycle number increase. Entropy error had no trend but it only detected the bearing fault.
- ICA captured better trends than variance and entropy and it did not affected by the changing load conditions in early cycles. A prediction of bearing could successfully be made by observing ICs.
- Monitoring results of MQE of SOM was also better than variance and entropy. It is not affected by the change of load conditions in early cycles but it did not have a smooth trend like ICA.
- Mahalanobis distances using SOM did not have a good aging trend and it was also accurate for detection of the bearing damage as the other methods.

As a result of this study, pattern recognition techniques can be used for tracking the bearing damage. Bearing damage predictions were made by applying ICA, MLP and SOM. According to the above results, the aim of this study has been reached successfully by applying a MLP neural network. The MLP network was not affected by the load changes and development of the bearing damage was well tracked from increasing oscillations of the output error (variance error) and prediction of the bearing damage made in the cycle 4, two cycles before the actual damage in cycle 6th.

REFERENCES

- [1] **Krause P.C.**, 1987, Analysis of Electric Machinery, McGraw-Hill International Editions, Electrical Engineering Series, pp. 164-165.
- [2] **Bonnet A. H. and Soukup G. C.**, 1992, Cause and analysis of stator and rotor failures in three phase squirrel cage induction motors, *IEEE Transactions on Industry Applications*, vol. **28**, no. 4, pp. 921-937.
- [3] **Kliman G. B. and Stein J.**, 1992, Methods of motor current signature analysis, *Electric Machines and Power Systems*, vol. **20**, pp. 463-474.
- [4] **Anders G. J., Endrenyi J., Ford G. L., and Stone G.C.**, 1990, A probabilistic model for evaluating the remaining life of electrical insulation in rotating machines, *IEEE Transactions on Energy Conversions*, vol. **5**, no. 5, pp. 761-766.
- [5] **Schoen R. R., Habetler T. G., Kamran F., and Bartheld R. G.**, 1994, Motor Bearing Damage Detection Using Stator Current Monitoring, *IEEE Industrial Applications Meeting*, vol. **1**, pp. 110-116.
- [6] **Castello M. J.**, 1993, Shaft voltages and rotating machinery, *IEEE Transactions on Industry Applications*, vol. **29**, no. 2, pp. 419-425.
- [7] **Yazici B. and Kliman G. B.**, 1999, An adaptive statistical time-frequency method for detection of broken bars and bearing faults in motors using stator current, *IEEE Trans. Industry Appl.*, vol. **35**, no. 2, pp. 442-452
- [8] **Nicholas J.R.**, 1993, Predictive Condition Monitoring of Electric Motors, P/PM Technology, pp. 28-32.
- [9] **Bisbee G.A.**, 1994, Why Do Motor Shaft and Bearing Fail, *TAPPI Journal*, vol. **77**, No. 9, pp. 251-252

- [10] **Seker S., Karatoprak E., Kayran A. H., and Senguler T.**, 2007, Stationary Wavelet Transform for Fault Detection in Rotating Machinery, *SPIE Optics East 2007*, Boston, MA, USA.
- [11] **Karatoprak E., Senguler T., Ayaz E., and Seker S.**, 2007, Comparisons of the Continuous and Discrete Wavelet Transforms for Potential Failure Detection in Electric Motors”, *ELECO 5th International Conference on Electrical and Electronics Engineering*, Bursa, Turkey
- [12] **Karatoprak E., Seker S., Cataltepe Z. and Senguler T.**, 2007, Using Bayes Decision Rule for Motor Fault Detection, *SIU IEEE 15th Signal Processing and Communication Applications Conference* , Eskisehir, Turkey.
- [13] **Karatoprak E., Senguler T., and Seker S.**, 2006, PMDC Motor Modelling, Control, Fault Detection and Classification Based on Artificial Neural Networks, *Turkish National Automatic Control Committee Meeting*, Ankara, Turkey.
- [14] **Karatoprak E., Senguler T., Ayaz E., Caglar R., and Seker S.**, 2007, Spectral and Statistical Based Modeling for Bearing Damage in Induction Motors, *The 6th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*, Cracow, Poland.
- [15] **Senguler T., Karatoprak E., Ayaz E., Gullulu S., and Seker S.**, 2007, Entropy Approach Using PCA for Sequential Accelerated Aging Processes in Electrical Motors, *The 6th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*, Cracow, Poland.
- [16] **Penman J. and Yin C.M.**, 1994, Feasibility of Using Unsupervised Learning, Artificial Neural Networks For The Condition Monitoring of Electrical Machines, *IEEE Proc.-Electr. Power Appl.*, vol. **141**, No. 6.
- [17] **Smyth P.**, 1993, Hidden Markov Models and Neural Networks for Fault Detection Indynamic Systems, *Neural Networks for Signal Processing, IEEE Proceedings*, IEEE-SP Workshop Volume, pp. 582-592.

- [18] **Hatzipantelis E., Murray A., and Penman J.**, 1995, Comparing Hidden Markov Models With Artificial Neural Network Architectures for Condition Monitoring Applications, *Fourth International Conference on Artificial Neural Networks*, Volume , Issue , 26-28, pp:369 – 374.
- [19] **Taylor J. K.**, 1990, Statistical Techniques for Data Analysis, Lewis Published.
- [20] **Vaseghi V. S.**, 1996, Advanced Signal Processing and Digital Noise Reduction. John Wiley, New York, pp. 38-42.
- [21] **Gabor D.**, 1946, Theory of Communication, *IEEE*, vol. **93**, no. 3, pp. 429-457.
- [22] **Seker S., and Ayaz E.**, 2002, A Study of Condition Monitoring for Induction Motors Under Accelerated Aging Process, *IEEE Power Engineering Review*, vol. **22**, no. 7, pp. 35-36.
- [23] **Daubechies I.**, 1990, The Wavelet Transform, Time-Frequency Localization and Signal Analysis, *IEEE Trans. on Information Theory*, vol. **36**, No.5.
- [24] **Mallat S.**, 1997, A Wavelet Tour of Signal Processing, Academic Press, New York.
- [25] **Cover T. M. and Thomas J. A.**, 1991, Elements of Information Theory, Wiley, New York.
- [26] **Shannon C. E.**, 1948, A Mathematical Theory of Communication, The Bell System Technical Journal, vol.27, pp.379-423,623-656.
- [27] **Schneier B.**, 1996, Applied Cryptography, John Wiley and Sons, pp.234.
- [28] **Shlens J.**, 2005, A tutorial on Principal Component Analysis, Systems Neurobiology Laboratory, Salk Institute for Biological Studies and Institute for Nonlinear Science, University of California, San Diego.
- [29] **Smith L. I.**, 2006, A Tutorial on Principal Component Analysis, University of Otago, New Zealand.
- [30] **Cherry E. C.**, 1955, Some Experiments on the Recognition of Speech with One and with Two Ears”, *J. Acoust. Soc. Amer.*, vol. **25**, pp. 975-979, 1955.

- [31] **Hyvärinen A., Karhunen J., and Oja E.**, 2001, Independent Component Analysis, John Wiley & Sons, Inc., NY.
- [32] **Papoulis A.**, 1991, Probability, Random Variables, and Stochastic Processes, McGraw-Hill.
- [33] **Duda R. O., Hart P. E., and Stork D. G.**, 2000, Pattern Classification, Wiley.
- [34] **Jang J. R., Sun C., and Mizutan E.**, 1997, Neuro-Fuzzy and Computing, A Computational Approach to Learning and Machine Intelligence, Prentice Hall, pp. 423-425.
- [35] **Alpaydin E.**, 2004, Introduction to Machine Learning, The MIT Press, Cambridge MA, pp. 69-73.
- [36] **Rosenblatt F.**, 1959, The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, *Psychological Review*, vol. **65**, pp. 386-408.
- [37] **Hines J. W.**, 1997, Matlab Supplement to Fuzzy and Neural Approaches in Engineering, John Wiley & Sons, Inc., pp 71-88
- [38] **Tsoukalas L. H., and Uhrig R. E.**, 1997, Fuzzy and Neural Approaches in Engineering, John Wiley & Sons, Inc., pp 229-287.
- [39] **Hecht-Nielsen R.**, 1989, Neurocomputing, Addison-Wesley Longman Publishing Co., Inc., pp 58-61.
- [40] **Rutkowski L.**, 2004, Adaptive Probabilistic Neural Networks for Pattern Classification in Time-Varying Environment” *IEEE Trans. on Neural Networks*, vol. **15**, No.4, pp. 811-827.
- [41] **Specht D. F.**, 1998, Probabilistic Neural Networks for Classification, Mapping, or Associative Memory”, *IEEE International Conference on Neural Networks*, vol. **1**. pp. 525-532.
- [42] **Mahalanobis P.C.**, 1936, On the Generalized Distance in Statistics, *Proceedings of the National Institute of Science*, India, pp. 49-55.
- [43] **Chapman S. J.**, 1999, Electric Machinery Fundamentals, McGraw Hill International Editions, Third Edition, pp. 357-448.

- [44] **Ayaz E., Ozturk A., and Seker S.**, 2006, Continuous Wavelet Transform for Bearing Damage Detection in Electric Motors, *Electrotechnical Conference*, Melecon 2006, IEEE Mediterranean, pp. 1130-1133.
- [45] **Chen S., Lipo T.A.**, 1998, Bearing Currents and Shaft Voltages of Induction Motors Under Hard- and Soft-Switching Inverter Excitation. , *IEEE Transactions on Industry Applications*, vol. **34**, no. 5, pp. 1042-1048.
- [46] **Erbay A. S., Upadhyaya B. R.**, 1999, Multi-sensor Fusion for Induction Motor Aging Analysis and Fault Diagnosis, *Research Report*, The University of Tennessee, Nuclear Engineering Department, Knoxville, Tennessee, 37996-2300, US.
- [47] **Donner G. L., Subler W. L., Evon S. T.**, A Motor Primer-Part I, *IEEE Transactions on Industry Applications*, vol. **38**, no. 5, pp. 1455-1465.
- [48] **IEEE Std. 117-1974**, IEEE Standard Test Procedure for Evaluation of Systems of Insulation Materials for Random-wound AC Electric Machinery.
- [Url-2] <<http://www.artesis.com/>>
- [Url-2] <http://en.wikipedia.org/wiki/Image:All_Gizah_Pyramids.jpg>
- [Url-3] <<http://cis.hut.fi/projects/somtoolbox/>>

APPENDICES

APPENDIX A.1 : Calculation of the weights in the output layer of MLP.

Error is calculated by subtracting the output of the neuron in layer k from its desired target value T :

$$\varepsilon = \varepsilon_q = [T_q - \Phi_{q,k}] \quad (\text{A.1})$$

$$\varepsilon^2 = \varepsilon_q^2 = [T_q - \Phi_{q,k}]^2 \quad (\text{A.2})$$



Figure A.1 : Representation of the adjusting the weights of the output layer neurons.

The delta rule is applied for calculating the change in the weight with respect to the square error:

$$\Delta w_{pq,k} = -\eta_{p,q} \frac{\partial \varepsilon_q^2}{\partial w_{pq,k}} \quad (\text{A.3})$$

Here, η stands for the *learning rate*. By using the chain rule, the evaluation of the partial derivative becomes:

$$\frac{\partial \varepsilon_q^2}{\partial w_{pq,k}} = \frac{\partial \varepsilon_q^2}{\partial \Phi_{q,k}} \frac{\partial \Phi_{q,k}}{\partial I_{q,k}} \frac{\partial I_{q,k}}{\partial w_{pq,k}} \quad (\text{A.4})$$

The partial derivative of the error from Equation (A.2) is calculated as:

$$\frac{\partial \varepsilon_q^2}{\partial \Phi_{q,k}} = -2[T_q - \Phi_{q,k}] \quad (\text{A.5})$$

It is assumed that the neurons in the output layer have the logistic activation function. From Equation (4.7), we get:

$$\frac{\partial \Phi_{q,k}}{\partial I_{q,k}} = -\alpha \Phi_{q,k} [1 - \Phi_{q,k}] \quad (\text{A.6})$$

From Figure A.1, it is seen that the $I_{q,k}$ is the sum of the weighted inputs from the hidden layer, that is:

$$I_{q,k} = \sum_{p=1}^n w_{pq,k} \Phi_{p,j} \quad (\text{A.7})$$

The partial derivative of $I_{q,k}$ is then:

$$\frac{\partial I_{q,k}}{\partial w_{pq,k}} = \Phi_{p,j} \quad (\text{A.8})$$

By combining the partial derivatives above, Equation (A.4) can be written as:

$$\frac{\partial \epsilon_q^2}{\partial w_{pq,k}} = -2\alpha [T_q - \Phi_{q,k}] \Phi_{q,k} [1 - \Phi_{q,k}] \Phi_{p,j} = -\delta_{pq,k} \Phi_{p,j} \quad (\text{A.9})$$

where $\delta_{q,k}$ is denoted as:

$$\delta_{pq,k} \equiv 2\alpha [T_q - \Phi_{q,k}] \Phi_{q,k} [1 - \Phi_{q,k}] \quad (\text{A.10})$$

$$\delta_{pq,k} = 2\epsilon_q \frac{\partial \Phi_{q,k}}{\partial I_{q,k}} \quad (\text{A.11})$$

Finally, the change in the weight vector can be written as:

$$\Delta w_{pq,k} = -\eta_{p,q} \frac{\partial \epsilon_q^2}{\partial w_{pq,k}} = -\eta_{p,q} \delta_{pq,k} \Phi_{p,j} \quad (\text{A.12})$$

The new weight vector:

$$w_{pq,k}(N+1) = w_{pq,k}(N) - \eta_{p,q} \delta_{pq,k} \Phi_{p,j} \quad (\text{A.13})$$

Here, N is the number iteration involved. Then, the adjusted weight is defined in Equation (A.13).

APPENDIX A.2 : Calculation of the weights in the hidden layer.

The delta rule can also be defined for the change of the weight vector of the neurons in the hidden layer:

$$\Delta w_{hp,j} = -\eta_{h,p} \frac{\partial \mathcal{E}^2}{\partial w_{hp,j}} = -\eta_{h,p} \sum_{q=1}^r \frac{\partial \mathcal{E}_q^2}{\partial w_{hp,j}} \quad (\text{A.14})$$

The total mean square error is represented as:

$$\mathcal{E}^2 = \sum_{q=1}^r \mathcal{E}_q^2 = \sum_{q=1}^r [T_q - \Phi_{q,k}]^2 \quad (\text{A.15})$$

since the several output errors are involved.

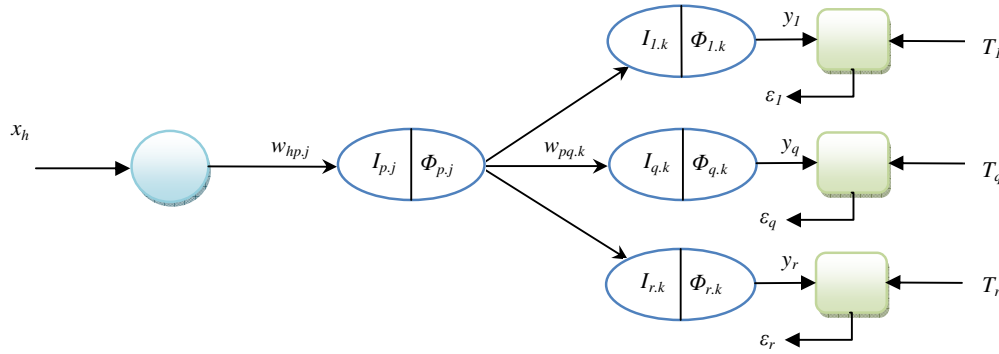


Figure A.2 : Representation of the adjusting the weights of the hidden layer neurons.

The partial derivative of the total mean square error with respect to the $w_{hp,j}$ is given:

$$\frac{\partial \mathcal{E}^2}{\partial w_{hp,j}} = \sum_{q=1}^r \frac{\partial \mathcal{E}_q^2}{\partial \Phi_{q,k}} \frac{\partial \Phi_{q,k}}{\partial I_{q,k}} \frac{\partial I_{q,k}}{\partial \Phi_{p,j}} \frac{\partial \Phi_{p,j}}{\partial I_{p,j}} \frac{\partial I_{p,j}}{\partial w_{hp,j}} \quad (\text{A.16})$$

The first two partial derivatives in Equation (A.16) are already given in Equations (A.5) and (A.6) which are:

$$\frac{\partial \mathcal{E}_q^2}{\partial \Phi_{q,k}} = -2(T_q - \Phi_{q,k}) = 2\mathcal{E}_q \quad (\text{A.17})$$

$$\frac{\partial \Phi_{q,k}}{\partial I_{q,k}} = -\alpha \Phi_{q,k} (1 - \Phi_{q,k}) \quad (\text{A.18})$$

The partial derivative of $I_{q,k}$ is then:

$$I_{q,k} = \sum_{p=1}^n w_{pq,k} \Phi_{p,j} \quad (\text{A.19})$$

$$\frac{\partial I_{q,k}}{\partial \Phi_{p,j}} = w_{pq,k} \quad (\text{A.20})$$

It is also assumed that the neurons of the hidden layer have the logistic activation function. The partial derivative of $\Phi_{p,j}$ is given:

$$\frac{\partial \Phi_{p,j}}{\partial I_{p,j}} = -\alpha \Phi_{p,j} (1 - \Phi_{p,j}) \quad (\text{A.21})$$

The sum of the weighted inputs of the input layer is:

$$I_{p,j} = \sum_{h=1}^m w_{hp,j} x_h \quad (\text{A.22})$$

The partial derivative of $I_{p,j}$ is also calculated as:

$$\frac{\partial I_{p,j}}{\partial w_{hp,j}} = x_h \quad (\text{A.23})$$

By combining the partial derivative expressions above, Equation (A.16) can be written as:

$$\frac{\partial \mathcal{E}^2}{\partial w_{hp,j}} = \sum_{q=1}^r (-2) \alpha (T_q - \Phi_{q,k}) [\Phi_{q,k} (1 - \Phi_{q,k})] w_{pq,k} \alpha [\Phi_{p,j} (1 - \Phi_{p,j})] x_h \quad (\text{A.24})$$

By using the $\delta_{q,k}$ from equation (A.11):

$$\frac{\partial \mathcal{E}^2}{\partial w_{hp,j}} = - \sum_{q=1}^r \delta_{pq,k} w_{pq,k} \frac{\partial \Phi_{p,j}}{\partial I_{p,j}} x_h \quad (\text{A.25})$$

By adding the $\delta_{hp,j}$ to the equation (A.15):

$$\delta_{hp,j} \equiv \delta_{pq,k} w_{pq,k} \frac{\partial \Phi_{p,j}}{\partial I_{p,j}} \quad (\text{A.26})$$

The error is given as:

$$\frac{\partial \varepsilon^2}{\partial w_{hp,j}} = - \sum_{q=1}^r \delta_{hp,j} x_h \quad (\text{A.27})$$

Finally, the change in the weight vector can be expressed as:

$$\Delta w_{hp,j} = -\eta_{h,p} \frac{\partial \varepsilon^2}{\partial w_{hp,j}} = -\eta_{h,p} \sum_{q=1}^r \delta_{pq,k} w_{pq,k} \frac{\partial \Phi_{p,j}}{\partial I_{p,j}} x_h \quad (\text{A.28})$$

$$\Delta w_{hp,j} = -\eta_{h,p} x_h \sum_{q=1}^r \delta_{hp,j} \quad (\text{A.29})$$

The updated weight vector of the neuron in the hidden layer is given as:

$$w_{hp,j}(N+1) = w_{hp,j}(N) - \eta_{p,q} x_h \sum_{q=1}^r \delta_{hp,j} \quad (\text{A.30})$$

Backpropagation algorithm has a weakness with the local minima. The local minima prevent the minimization of the error. A momentum term is introduced to reach the global minimum.

$$\Delta w_{pq,k}(N+1) = -\eta_{p,q} \delta_{pq,k} \Phi_{p,j} + \mu w_{pq,k}(N) \quad (\text{A.31})$$

$$w_{pq,k}(N+1) = w_{pq,k}(N) + \Delta w_{pq,k}(N+1) \quad (\text{A.32})$$

Here, μ is the momentum coefficient which provides the stability for the training process [22].

VITA

Candidate's full name: Tayfun ŞENGÜLER

Place and date of birth: Istanbul - 18/08/1980

Permanent Address: Istanbul Technical University, Faculty of Electrical & Electronic Engineering, Department of Electrical Engineering, MASLAK, ISTANBUL

Universities and Colleges attended: Istanbul Technical University

Publications:

- Karatoprak E., Şengüler T., Ayaz E., Çağlar R., Şeker S., 2007. Spectral and Statistical Based Modeling for Bearing Damage in Induction Motors, The 6th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, Cracow-Poland.
- Şengüler T., Karatoprak E., Ayaz E., Güllülü S., Şeker S., 2007. Entropy Approach Using PCA for Sequential Accelerated Aging Processes in Electrical Motors, The 6th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, Cracow-Poland.
- Şeker S., Karatoprak E., Kayran A. H., Şengüler T., 2007. Stationary wavelet transform for fault detection in rotating machinery, SPIE Optics East 2007, Seaport World Trade Center, Boston, MA, USA.
- Karatoprak E., Şengüler T., Ayaz E., Şeker S., 2007. Comparisons of the continuous and Discrete Wavelet Transforms for Potential Failure Detection in Electric Motors, ELECO 5th International Conference on Electrical and Electronics Engineering, Bursa, TURKEY
- Karatoprak E., Şeker S., Çataltepe Z., Şengüler T., 2007. Using Bayes Decision Rule for Motor Fault Detection, SIU (IEEE 15th Signal Processing and Communication Applications Conference), Eskişehir, Turkey.
- Karatoprak E., Şengüler T., Şeker S., 2006. PMDC Motor Modeling, Control, Fault detection and Classification Based on Artificial Neural Networks, Turkish National Automatic Control Committee 2006 Meeting Ankara.