

ISTANBUL TECHNICAL UNIVERSITY ★GRADUATE SCHOOL OF SCIENCE
ENGINEERING AND TECHNOLOGY

**SIGNAL BASED DATA MINING FOR FEATURE EXTRACTION AND FAULT
DETECTION**

Ph.D. THESIS

Selim GÜLLÜLÜ

Department of Electrical Engineering

Electrical Engineering Programme

SEPTEMBER 2012

ISTANBUL TECHNICAL UNIVERSITY ★GRADUATE SCHOOL OF SCIENCE
ENGINEERING AND TECHNOLOGY

**SIGNAL BASED DATA MINING FOR FEATURE EXTRACTION AND FAULT
DETECTION**

Ph.D. THESIS

Selim GÜLLÜLÜ
(504042001)

Department of Electrical Engineering

Electrical Engineering Programme

Thesis Advisor: Prof. Dr. Serhat ŞEKER

SEPTEMBER 2012

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

**ÖZELLİK ÇIKARIMI VE ARIZA TANISI İÇİN İŞARET TABANLI VERİ
MADENCİLİĞİ**

DOKTORA TEZİ

**Selim GÜLLÜLÜ
(504042001)**

Elektrik Mühendisliği Anabilim Dalı

Elektrik Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Serhat ŞEKER

EYLÜL 2012

Selim GÜLLÜLÜ, a Ph.D. student of ITU Graduate School of Science Engineering and Technology student ID 504042001, successfully defended the thesis entitled “**SIGNAL BASED DATA MINING FOR FEATURE EXTRACTION AND FAULT DETECTION**”, which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Prof. Dr. Serhat ŞEKER**
İstanbul Technical University

Jury Members : **Prof. Dr. Tülay YILDIRIM**
Yıldız Technical University

Prof. Dr. Melih GEÇKİNLİ
Beykent University

Prof. Dr. Canbolat UÇAK
Yeditepe University

Asst. Prof. Dr. Ramazan ÇAĞLAR
İstanbul Technical University

Date of Submission : 16 December 2011

Date of Defense : 19 September 2012

FOREWORD

I would like to express my deep appreciation and thanks for my advisor, Prof. Serhat Şeker. I also would like to thank to the University of Tennessee Nuclear Engineering Department, The Maintenance and Reliability Center and especially Prof. Belle. R. Upadhyaya for the motor vibration data which is used for test purposes in this study. This work is supported by ITU Graduate School of Science.

September 2012

Selim Güllülü
(M.Sc. Electronics and
Telecommunications Engineer)

TABLE OF CONTENTS

	<u>Page</u>
FOREWORD	v
TABLE OF CONTENTS.....	ix
ABBREVIATIONS	xi
LIST OF TABLES	xiii
LIST OF FIGURES	xv
SUMMARY	xix
ÖZET.....	xxi
1. INTRODUCTION.....	1
1.1 Purpose of Thesis	1
1.2 Background	1
1.2.1 General methods for data mining.....	1
1.2.2 Wavelet transform as a signal processing application in fault detection	4
1.2.3 Data mining and fault detection	7
1.3 Motivation Of The Thesis And Target Contribution	8
1.4 New Algorithm For Signal Based Data Mining Application.....	10
1.5 Organization Of The Thesis	11
2. MATHEMATICAL METHODS.....	13
2.1 Spectral Methods.....	13
2.1.1 Discrete fourier transform.....	13
2.1.2 Wavelet transform.....	14
2.1.3 Continuous wavelet transform	18
2.2 Artificial Neural Networks.....	19
2.2.1 Definition	19
2.2.2 Main concepts	21
2.2.3 Types of neural networks.....	22
2.2.4 Auto-associative neural network.....	24
3. DATA MINING.....	27
3.1 Data Mining Concept	27
3.2 Data Mining Input Data.....	29
3.3 Data Mining Functionalities.....	31
3.4 Data Mining Techniques	32
4. EXPERIMENTAL STUDY FOR VIBRATION MEASUREMENTS AND SENSOR VALIDATION.....	35
4.1 Motor Aging Process And The Creation Of The Artificial Bearing Fault.....	35
4.2 Motor Load Performance Test And Data Acquisition System.....	39
4.3 Sensor Validation via Neural Network	50
4.4 General Characterization Of Bearing Damage.....	56
5. SIMULATION FOR THE PROPOSED DATA MINING ALGORITHM AND ITS MATHEMATICAL INTERPRETATION.....	57
5.1 Simulation Data.....	57

5.2 Mathematical Interpretation of the CWT as a Correlation	60
6. SIGNAL BASED ALTERNATIVE APPROACH FOR FAULT DETECTION IN DATA MINING	65
6.1 Data Pre-Processing for Application	65
6.2 Neural Network Application	66
6.2.1 Neural network training and recall phase.....	67
6.2.2 Neural network test phase	72
6.2.3 Application on the vibration signal of the electric motor in healthy state	73
6.2.4 Application on simulation signal - I.....	79
6.2.5 Application on simulation signal - II.....	85
6.3 Summary of the Proposed Algorithm and Evaluation of Results.....	91
7. CONCLUSIONS AND DISCUSSIONS	93
REFERENCES	97
APPENDICES	105
APPENDIX A.	107
APPENDIX B.....	111
APPENDIX C.....	115
APPENDIX D	119
APPENDIX E	123
APPENDIX F	127
APPENDIX G	131
CURRICULUM VITAE	137

ABBREVIATIONS

AA	: Auto-Associative
ANN	: Artificial Neural Network
AANN	: Auto-Associative Neural Network
AF	: Activation Function
AHP	: Analytic Hierarchy Process
ANN	: Artificial Neural Network
App	: Appendix
BP	: Back propagation
CWT	: Continuous Wavelet Transform
DFT	: Discrete Fourier Transform
DWT	: Discrete Wavelet Transform
FFT	: Fast Fourier Transform
GA	: Genetic Algorithm
GDM	: Gradient Descent Momentum
GUI	: Graphical User Interface
HA	: Hetero Associative
KDD	: Knowledge Discovery in Databases
LMA	: Levenberg – Marquardt Algorithm
MSE	: Mean Square Error
NN	: Neural Network
MLP	: Multi-layer Perceptron
NPSD	: Normalized Power Spectral Density
PCA	: Principal Component Analysis
PCWT	: Psd of the Continuous Wavelet Transform
PSD	: Power Spectral Density
RBP	: Resilient Backpropagation
RPM	: Revolutions Per Minute
SLP	: Single Layer Perceptron
SNR	: Signal Noise Ratio
SOM	: Self Organizing Maps
SQL	: Structured Query Language
SSE	: Sum of Squared Error
SVM	: Support Vector Machines
STFT	: Short Time Fourier Transform
TSDM	: Time Series Data Mining
WD	: Wigner Distribution
WT	: Wavelet Transform

LIST OF TABLES

	<u>Page</u>
Table 4.1 :Plate data of the motor	36
Table 4.2 :Dimensions of the motor.....	36
Table 5.1 :Statistical comparison of the signals.....	59
Table 6.1 :AANN training and test results for the vibration signal	76
Table 6.2 :AANN training and test results for the simulation signal I.....	83
Table 6.3 :AANN training and test results for the simulation signal II	89

LIST OF FIGURES

	<u>Page</u>
Figure 2.1 : Morlet and Daubechies 4 Wavelets.....	15
Figure 2.2 : Symlet 4 Wavelet.	15
Figure 2.3 : Wavelet band-pass filters on a logarithmic frequency scale.	16
Figure 2.4 : Typical wavelet family in time and frequency domains.	16
Figure 2.5 : Time-frequency plane showing resolution cells for wavelet transform.	17
Figure 2.6 : Structure of a biological neuron.	20
Figure 2.7 : Structure of an artificial neuron.....	20
Figure 2.8 : Threshold function.	21
Figure 2.9 : Sigmoid function.	21
Figure 2.10 : Example AANN topology with three hidden layers.....	25
Figure 2.11 : Example AANN topology with single hidden layer.....	25
Figure 3.1 : Knowledge discovery process.	28
Figure 3.2 : Global temperature change 30	30
Figure 3.3 : Seismic wave measurement via a sismogram 30	30
Figure 3.4 : Data mining technologies based on retention of data..... 32	32
Figure 3.5 : Proposed methodology in the thesis..... 33	33
Figure 4.1 : Motor load testing and data acquisition system a) experimental setup configuration; b) Cross section (A-A') at short end; c) cross section (B- B') at pulley end 38	38
Figure 4.2 : Schematic of the electrical motor bearing EDM setup..... 38	38
Figure 4.3 : First 8 channels' output..... 40	40
Figure 4.4 : Last 8 channels' output. 41	41
Figure 4.5 : Vibrations of both sensors PE2 and PE10 stage 0. 42	42
Figure 4.6 : Vibrations of both sensors PE2 and PE10 stage 1..... 42	42
Figure 4.7 : Vibrations of both sensors PE2 and PE10 stage 2. 43	43
Figure 4.8 : Vibrations of both sensors PE2 and PE10 stage 3..... 43	43
Figure 4.9 : Vibrations of both sensors PE2 and PE10 stage 4..... 44	44
Figure 4.10 : Vibrations of both sensors PE2 and PE10 stage 5. 44	44
Figure 4.11 : Vibrations of both sensors PE2 and PE10 stage 6. 45	45
Figure 4.12 : Vibrations of both sensors PE2 and PE10 stage 7. 45	45
Figure 4.13 : Spectra of the vibration sensors PE2 and PE10, stage 0. 46	46
Figure 4.14 : Spectra of the vibration sensors PE2 and PE10, stage 1. 46	46
Figure 4.15 : Spectra of the vibration sensors PE2 and PE10, stage 2. 47	47
Figure 4.16 : Spectra of the vibration sensors PE2 and PE10, stage 3. 47	47
Figure 4.17 : Spectra of the vibration sensors PE2 and PE10, stage 4. 48	48
Figure 4.18 : Spectra of the vibration sensors PE2 and PE10, stage 5. 48	48
Figure 4.19 : Spectra of the vibration sensors PE2 and PE10, stage 6. 49	49
Figure 4.20 : Spectra of the vibration sensors PE2 and PE10, stage 7. 49	49
Figure 4.21 : Spectra of the motor vibration data from accelerometer #9. 50	50

Figure 4.22 : Spectra of the motor vibration data from accelerometer #10.	51
Figure 4.23 : Neural network #1.	52
Figure 4.24 : Neural network #2.	52
Figure 4.25 : PE10 and PE2 training data.	53
Figure 4.26 : PE10 and PE2 test data.	54
Figure 4.27 : PSD of accelerometers PE10 and PE2 in log scale.....	54
Figure 4.28 : NN1 output error.	55
Figure 4.29 : NN2 output error.	56
Figure 5.1 : Vibration data simulation.....	57
Figure 5.2 : Vibration signal in healthy case (measured data).	58
Figure 5.3 : Simulation data and measured vibration signal.	58
Figure 5.4 : PSD of the real vibration (black) and simulation data (red).....	59
Figure 5.5 : Spectral comparisons: PSD (blue), PSD of CWT (red).....	60
Figure 5.6 : Scale-frequency band relationship.....	62
Figure 6.1 : Data pre-processing	66
Figure 6.2 : 3-layer AANN topology... ..	67
Figure 6.3 : 5-layer AANN topology.....	68
Figure 6.4 : 5-layer AANN convergence	68
Figure 6.5 : 5-layer AANN learning.	69
Figure 6.6 : Slow learning with GDM.	70
Figure 6.7 : Fast learning with RBP.....	70
Figure 6.8 : Block diagram of the algorithm and its training.....	71
Figure 6.9 : Block diagram of the signal based data mining algorithm and its test...72	
Figure 6.10 : Vibration signal of the electric motor in healthy state.	73
Figure 6.11 : CWT of the vibration signal between 0-0.25 seconds.....	74
Figure 6.12 : Normalized PSD of the CWT at the first scale.....	74
Figure 6.13 : Normalized PSD of the raw signal between 0-0.25 seconds.....	75
Figure 6.14 : Error variation for the training sample Pcwtc3 – run #8.....	77
Figure 6.15 : Error variation for the recall sample Pcwtc20 – run #8.....	77
Figure 6.16 : Error variation for the test sample PSD20 – run #8.....	78
Figure 6.17 : Learning curve with RBP for run #8.....	78
Figure 6.18 : Simulation signal I.	79
Figure 6.19 : Vibration signal (red) and simulation (black).....	80
Figure 6.20 : PSD of the vibration signal (black) and simulation I (red).....	80
Figure 6.21 : CWT of the simulation signal between 0-0.25 seconds.....	81
Figure 6.22 : Normalized PSD of the CWT at the first scale.....	81
Figure 6.23 : Error variation for the training sample pcwtc3 , run #12.....	82
Figure 6.24 : Error variation for the recall sample pcwtc20, run #12.....	84
Figure 6.25 : Error variation for the test sample PSD20, run #12.....	84
Figure 6.26 : Learning curve with RBP for run #12.....	85
Figure 6.27 : Simulation signal II.	86
Figure 6.28 : PSD of the vibration signal (black) and simulation II (red).....	86
Figure 6.29 : CWT of the simulation signal II between 0-0.25 seconds.....	87
Figure 6.30 : Normalized PSD of the CWT at the first scale.....	87
Figure 6.31 : Error variation for the training sample pcwtc3, run#10.....	88
Figure 6.32 : Error variation for the recall sample pcwtc20, run #10.....	90
Figure 6.33 : Error variation for the test sample PSD20, run #10.....	90
Figure 6.34 : Learning curve with RBP, run #10.	91

Figure A.1 : Pcwtc1 – Pcwtc5	107
Figure A.2 : Pcwtc6 – Pcwtc10	108
Figure A.3 : Pcwtc11 – Pcwtc15	109
Figure A.4 : Pcwtc16 – Pcwtc20	110
Figure B.1 : PSD1-PSD5	111
Figure B.2 : PSD6-PSD10	112
Figure B.3 : PSD11-PSD15	113
Figure B.4 : PSD16-PSD20	114
Figure C.1 : Pcwt1-Pcwt5	115
Figure C.2 : Pcwt6-Pcwt10	116
Figure C.3 : Pcwt11-Pcwt15	117
Figure C.4 : Pcwt16-Pcwt20	118
Figure D.1 : PSD1 – PSD5	119
Figure D.2 : PSD6 – PSD10	120
Figure D.3 : PSD11 – PSD15	121
Figure D.4 : PSD16 – PSD20	122
Figure E.1 : Pcwt1 – Pcwt5	123
Figure E.2 : Pcwt6 – Pcwt10	124
Figure E.3 : Pcwt11 – Pcwt15	125
Figure E.4 : Pcwt16 – Pcwt20	126
Figure F.1 : PSD1 – PSD5	127
Figure F.2 : PSD6 – PSD10	128
Figure F.3 : PSD11 – PSD15	129
Figure F.4 : PSD16 – PSD20	130
Figure G.1 : Graphical user interface of application	131
Figure G.2 : Vibration signal	132
Figure G.3 : PSD of the vibration signal	132
Figure G.4 : CWT at 1st scale	133
Figure G.5 : CWT at multiple scales	133
Figure G.6 : CWT at 1st scale	134
Figure G.7 : PSD of the CWT at 1st scale	134
Figure G.8 : Training of the AANN	135
Figure G.9 : Test of the AANN	135

SIGNAL BASED DATA MINING FOR FEATURE EXTRACTION AND FAULT DETECTION

SUMMARY

In this study, a new method for feature extraction and fault detection is introduced by using signal processing and data mining techniques. The application is mainly consisted of two parts: data pre-processing and artificial neural network. At the data pre-processing phase, the vibration signal from the healthy state of an electric motor is used. At this phase, firstly, the sub-signal at the first scale is obtained by applying continuous wavelet transform (CWT) to the vibration signal. The computation which is done for the first scale presents maximum correlation in maximum frequency range. After continuous wavelet transform, Fourier transform (power spectral density –PSD) is applied to this sub-signal. At the neural network phase, the normalized PSD of the CWT at the first scale is given for training an auto-associative neural network (AANN). The reason of using an AANN is its identity mapping property. After training, the neural network is recalled by the same type of data set (PSD of the CWT at the first scale). During the recall phase, a threshold value is identified by the error variation at the output. During the test phase, the PSD of the raw signal is used. The error deviation which appears at the output of the neural network at the test phase forms a feature vector. This vector is compared with the threshold identified at the recall phase of the neural network. The patterns exceeding the threshold are identified as faults. According to the results, it is clearly seen that the neural network creates higher amplitudes between 2 – 4 kHz interval. This property is well known in literature and is a characteristic of bearing damages which occur during the aging process of the electric motors. Hence, it can be said that the proposed method in this study can predict the manufacturing defects by using the healthy state data only. The method is also tested on two other simulation signals. One of the simulation signals is similar to the vibration signal obtained from the electric motor. The results for this simulation data is matching with the results of the vibration signal, showing a faulty state at the 2-4 kHz frequency band. The second simulation signal which is tested by the method has different frequency properties than the vibration signal. The results for the second simulation shows a novelty state in the 4-6 kHz frequency band. The hidden property of the simulation signal is extracted by the proposed method.

ÖZELLİK ÇIKARIMI VE HATA TESPİTİ İÇİN İŞARET TABANLI VERİ MADENCİLİĞİ

ÖZET

Bu doktora çalışmasında veri madenciliği ve işaret işleme teknikleri kullanılarak özellik çıkarımı ve arıza tanısına yönelik yeni bir yöntem geliştirilmiştir. Çalışma ana olarak veri ön-işleme ve yapay sinir ağı uygulaması olmak üzere iki kısımdan oluşmaktadır. Çalışmada ham veri olarak asenkron elektrik motorunun sağlam durumundaki titreşim işareti kullanılmıştır. Veri ön-işleme aşamasında ham veri olarak kullanılmakta olan titreşim işaretine sürekli dalgacık dönüşümü uygulanmaktadır. Bu aşamada ana dalgacık olarak symlet 4 ana dalgacığı kullanılmıştır. Sürekli dalgacık dönüşümü uygulandıktan sonra elde edilen yeni işaretin birinci ölçekteki alt-işareti uygulamanın devamında kullanılmaktadır. Dalgacık analizi, analiz edilen işaret ile analiz eden fonksiyon (dalgacık) arasındaki korelasyona dayanmaktadır. Bu nedenle birinci ölçekte geniş frekans aralığında işaret içeriği gösterildiğinden korelasyon ilişkisi de geniş bir frekans aralığını kapsamaktadır. Sürekli dalgacık dönüşümünün birinci ölçeğinde temsil edilen bu alt işarete daha sonra Fourier dönüşümü (güç spektrumu yoğunluğu) ve normalizasyon işlemleri uygulanmıştır. Bu sayede işaretin frekans domenindeki gösterimi elde edilmektedir. Bu dönüşüm sonunda elde edilen güç spektrumu yoğunluğu ile ham işaretin güç spektrumu yoğunluğu karşılaştırıldığında yüksek frekans bölgesinde belli frekanslardaki genliklerin sürekli dalgacık dönüşümünün korelasyon özelliği sayesinde kuvvetlendiği açıkça gözlemlenebilmektedir. Veri ön-işleme kısmından sonra elde edilen bu işaret bir öz-ilişkili yapay sinir ağının eğitiminde kullanılmaktadır. Öz-ilişkili yapay sinir ağları giriş ve çıkış katmanlarında aynı sayıda nöron bulunduran ve genellikle boyut azaltımı, temel bileşen analizi gibi işlemlerde kullanılmakta olan bir yapay sinir ağı tipidir. Giriş ve çıkış katmanlarındaki nöron sayılarının aynı olmasının sebebi, ağın girişine ve çıkışına aynı işaretin verilmesini sağlayabilmektir. Giriş ve çıkış katmanları arasında bir ya da üç adet gizli katman bulunduran bu ağ tipinde gizli katmanlardaki nöron sayısı her zaman giriş ve çıkış katmanlarındaki nöron sayısından azdır. Bunun sebebi girişe verilen işaretin önemli bileşenlerinin ağ tarafından öğrenilmesini ve işaretin çıkış katmanında yeniden oluşturulabilmesini sağlamaktır. Ham veri olarak kullanılan elektrik motorunun sağlam durumundan elde edilen titreşim verisi 12 kHz frekansında örneklenmiştir. Toplam veri süresi 10 saniyedir. Bu da toplam 120.000 noktada ölçüm değeri olduğunu göstermektedir. Uygulamada bu 10 saniyelik verinin 5 saniyelik kısmı kullanılmaktadır. Yapay sinir ağının eğitiminde kullanılan işaret kümesi 0.25 saniyelik bölümlere ayrılmış, her kümenin sürekli dalgacık dönüşümünün birinci ölçekteki temsili hesaplanmış, ardından güç spektrumu yoğunluğu alınmış ve normalize edilmiştir. Bu sayede eğitim kümesi için 20 adet örnek oluşturulmuştur. Güç spektrumu yoğunluğu 512 noktada hesaplandığından, yapay sinir ağının giriş ve çıkış katmanlarının her birinde 512 adet nöron bulunmaktadır. Yapay sinir ağının gizli katmanı için sırasıyla 50, 100 ve 150 adet

nöron kullanılarak, çalışma bu üç farklı yapay sinir ağı yapısı için ayrı ayrı uygulanmıştır. Eğitim için oluşturulan 20 örnekten oluşan küme sırasıyla 4-16, 8-12, 12-8 ve 16-4 şeklinde bölümlenerek eğitim ve geri çağırma kümeleri oluşturulmuştur. 4 farklı eğitim-geri çağırma kümesi ve 3 farklı ağ yapısı için toplam 12 adet deneme yapılmaktadır. Yapay sinir ağı eğitim kümesindeki örneklerle eğitildikten sonra geri çağırma kümesindeki örnekler yapay sinir ağına verilmiş ve çıkışta oluşan hata kaydedilmiştir. Geri çağırma kümesi ile çıkışta oluşan hataların maksimumu hesaplanarak 12 adet denemenin her biri için bir eşik değeri hesaplanmıştır. Eğitim ve geri çağırma aşaması bu şekilde tamamlanan yapay sinir ağına test aşamasında aynı elektrik motoru titreşim işaretinin güç spektrumu yoğunluğu uygulanmaktadır. Eğitimdeki gibi bir bölümlenme yapılarak 0.25 saniyelik titreşim işareti örneklerinin direkt olarak güç spektrumu yoğunluğu hesaplanarak (sürekli dalgacık dönüşümü uygulanmadan) ardından normalizasyon uygulanmaktadır. Bu şekilde test kümesi 20 adet örnekle oluşturulmaktadır. 12 denemenin her biri için ayrı eğitim kümesiyle eğitilmiş olan ve ayrı geri çağırma kümesi ile arıza eşiği hesaplanmış yapay sinir ağının girişine test kümesi girilmektedir. Test kümesi için çıkışta oluşan hata, geri çağırma aşamasında hesaplanmış olan eşik değeri ile kıyaslanmaktadır. Eşik değerini geçen hata değerleri, anormallik veya bozukluk olarak nitelendirilmektedir.

Elektrik motorunun sağlam durumuna ait titreşim verisinin ham veri olarak kullanıldığı bu uygulamanın sonucu değerlendirilecek olursa, yöntemin artan eğitim kümesi sayısı ile beraber anormallik veya bozukluk olarak belirlediği durum sayısının da artıyor olduğu görülmektedir. Yapay sinir ağının çıkışında elde edilen hatalar incelendiğinde, anormalliğin olduğu genlik değerlerinin 2-4 kHz arasında olduğu gözlemlenebilir. 2-4 kHz arasındaki bu yüksek genlikler elektrik motorlarında yaşlanmayla birlikte oluşan rulman bozulmalarına ait bir karakteristiktir. Bu arızalı durum tezin veri toplama ve motor aşındırma aşamalarının anlatıldığı 4. bölümde gösterilmiş ve ilgili referanslar paylaşılmıştır.

Tezde kullanılan yöntem elektrik motorunun sağlam durumundan elde edilen titreşim verisinde uygulandıktan sonra iki farklı simülasyon verisinde de denenmiştir.

Birinci simülasyon verisi elektrik motorundan elde edilen titreşim işaretine benzeyecek şekilde oluşturulmuştur. Yapay sinir ağı, eğitim, geri çağırma ve test kümeleri birinci simülasyon işareti için de aynı şekillerde oluşturulmuştur. Yöntemin uygulanmasının sonucunda beklendiği şekilde 2-4 kHz arasında anormalliklerin olduğu gözlemlenmiştir.

İkinci simülasyon verisi, elektrik motorundan elde edilen titreşim verisinden de birinci simülasyon verisinden de farklı karakteristiklere sahiptir. Yöntemin bu farklı işarete 4-6 kHz arasında anormallikler tespit ettiği görülmüştür. Bu anormallikler simülasyon işaretinde gizli olarak bulunan özelliklerdir ve işaretin güç spektrumu yoğunluğunda gözlemlenmemesine rağmen tezde önerilen yöntem sayesinde açığa çıkarılabilmektedir.

Kullanılan yöntemin hassasiyetini ifade edebilmek amacıyla işaret gürültü oranları göz önüne alınmış ve anormalliklerin tespit edilebildiği maksimum işaret gürültü oranı iterasyon yoluyla belirtilmiştir.

Tezde önerilen yöntemin arıza tespiti ve özellik çıkarımına ait getirdiği yeni yaklaşım sürekli dalgacık dönüşümünün korelasyon özelliğini kullanarak birinci ölçekte temsil edilen işaretin, ana işaretin yüksek frekans bölgesinde mevcut olan zayıf bileşenleri kuvvetlendirmesi ve bu sayede muhtemel arızalı durumları belirleyebilmesidir. Bu özelliğe ek olarak literatürde ele alınan benzer arıza tespiti uygulamalarıyla kıyaslandığında yöntem, sisteme ait (örn. elektrik motoru) sağlam

durum verisini kullanarak potansiyel arızaları belirleyebilmektedir. Literatürde büyük çoğunlukta kullanılan arıza tespit yöntemi, arızalı durumun yapay zeka uygulamasına (örn. yapay sinir ağı) öğretilmesinden sonra bilinmeyen durumların yapay zeka uygulamasına sorulması ve sonuçlara göre sınıflandırma/kümeleme yapılmasına dayanmaktadır. Bu açıdan bakıldığında tezde önerilen yöntem sadece sağlam durum verisinin farklı biçimleriyle yapay sinir ağını eğitip test etmekte ve sonuca ulaşmaktadır. Bu özellik de veri madenciliğinin veri kümesinde gizli halde bulunan değerli, önemli bilgiye ulaşma amacıyla birebir örtüşmektedir.

1. INTRODUCTION

In this introduction section, the general outline of the thesis is given. First, the purpose of the study is explained. In the background section, the earlier studies about the related topics are explained with references. The motivation and the contribution target of the thesis is explained after the literature survey. The new algorithm of the study is introduced in the following section. Finally, the organization of all other sections are described.

1.1 Purpose of Thesis

Fault detection and feature extraction are important aspects for ensuring the safe running of the machines. Signal processing and data mining are two different fields of study which can be used on condition monitoring and fault diagnostics with different perspectives. The purpose of this study is to introduce a new method which combines data mining and signal processing techniques for solving certain problems like feature extraction and fault detection. The recent studies related to these topics have been analysed accordingly and are described in the next section. Two different sets of data are used in this study: a collection of vibration signals obtained from an electric motor and simulation data. However, this new approach can be applied in other domains and systems which generate, or include, similar time series signals.

1.2 Background

In this section, the related literature search is given in three perspectives. First is describing the main methods of data mining. In the second part, the studies related to fault detection via signal processing, especially wavelets, are given. In the last part, recent data mining applications for fault detection are discussed.

1.2.1 General methods for data mining

The concept of data mining has become very popular during the last decades in many different domains like finance, industrial processes, medical, marketing, etc. Since

data mining is initially defined as a part of knowledge discovery in databases (KDD) process, it generally needs refined, clear and reliable data to work properly. Consequently, the main application areas of data mining have widened to certain domains, which contain “easy to use” data.

There is a variety of problems that data mining techniques try to solve. The most common problems can be listed as: characterization, discrimination, association analysis, classification, clustering, prediction and pattern extraction. These problems appear in several domains.

Related to these data mining problems, a huge number of studies have been done in literature. While speaking of a classification problem in data mining, the issue is to be able to automatically classify a set of objects (a document, a web page, a transaction, a person, a disease, etc.) according to the known samples. In other words, classification needs supervision. For categorization & classification, several techniques can be used like: support vector machines (SVM) [1], k-nearest neighbor method [2], genetic algorithms (GA) [3], fuzzy logic [4] or neural networks [5].

Even though the clustering problem is similar to classification, according to its nature, it is an unsupervised process. In the problem of clustering, just like classification, similar sets of objects are to be analyzed. But in this case, there are no known classes. Consequently, the objects are clustered according to certain criteria. The selection of these criteria constitutes the classification techniques. By the help of clustering, brain images can be analyzed for identifying abnormalities by using self organizing maps (SOM) [6], stock market analysis can be done by k-means method [7], users of social networks can be evaluated and assessed by k-means and hierarchical spectral clustering methods [8], university majors can be clustered by k-means and analytic hierarchy process (AHP) [9].

Another popular field of data mining is the association analysis. In the World Wide Web, we come across the usage of association analysis nearly every day. During online shopping, the suggestions that the web site introduces [10,11] are the outcomes of association analysis. The detection of frauds in credit card usages by the *support* and *confidence* calculations [12], subscription frauds in telecommunication [13], identifying associations in criminal networks [14] are other examples of association analysis.

Data mining is also used for predictive purposes. Churn prediction is a common study field. In the telecommunications sector, it is very important to predict churn subscribers; by using decision trees, one can calculate the risk of losing mobile users [15,16]. Financial distress prediction is also a popular field of study [17].

The data used by the data mining techniques vary according to the field of study. The data may be a collection of tables in a relational database, transactions in a transactional database, or even flat text files obtained from documents. Among these types of data, time series signals form a different field because of their nature. The data mining applications that deal with such data is called time series data mining (TSDM). Time series data is generally huge in size and have a high-dimensional structure [18-25]. Therefore they form one of the most challenging data types for data mining techniques [26–29]. Another main problem with the time series is that they generally include a noise pattern inside them, a widely used technique is to use low pass filters to eliminate the noise and make the data ready to be processed by the data mining techniques [30, 31]. Signal processing techniques like discrete wavelet transform (DWT) and discrete Fourier transform (DFT) are also used for reducing the dimension of the time series data [23].

Because of its computational advantages, artificial neural networks (ANN) are widely used in data mining. ANNs have a variety of applications, i.e. classification, clustering and prediction. They can assist to solve criminal cases and make better decisions [32]. They can help us build up better approaches in e-commerce [33], [34]. Better diagnosis techniques can be created in the medical domain [35,16]. In stock market analysis, ANNs and data mining techniques are commonly used [36], [37]. Meteorological and network traffic forecasting are also areas that ANNs are used for data mining and prediction [38,39].

In literature, it can easily be seen that many articles consider neural networks to be a promising data mining tool. ANNs offer qualitative methods for business and economic systems that traditional quantitative tools in statistics and econometrics cannot quantify due to the complexity in translating the systems into precise mathematical functions. Hence, the use of neural networks in data mining is a promising research field especially in the presence of large amount of data sets. In addition to this, neural networks are able to detect and assimilate relationships

between a large number of variables. In most cases neural networks perform as well or better than the traditional statistical techniques to which they are compared [40].

The main goal of feature extraction is to obtain more detailed information contained in the measured data than had been previously possible. Feature extraction also helps reducing the size of the input data which has redundancies in it. Depending on the type of the data, different feature extraction methods can be used. Standard digital signal processing techniques, such as time series statistics, correlation analysis, and fast Fourier transform (FFT) are widely used for feature extraction purposes. Also, at last decade, wavelet transforms (WT) revealed as one of the most popular methods in this field. A widely used area of feature extraction is the fault detection and diagnostic studies in electric motors by using vibration and current signals [41-44]. Vehicle engine sounds can be analyzed by the help of wavelet transform for identifying certain faults [45]. The assessment of wavelet transform (WT) as a feature extraction method can be used in representing the electrophysiological signals [46]. Facial features can be extracted by using discrete wavelet transforms (DWT) [47]. Wavelet transforms are also used for identifying faults in electric motors [48]. The details about the usage of the wavelet transform in fault detection are described in the next section.

1.2.2 Wavelet transform as a signal processing application in fault detection

Condition monitoring, fault diagnosis, early fault detection, fault prediction are very popular research topics especially for the industrial processes. In many electro-mechanical systems, several signal processing applications are used for the safe running of the devices in the system. The raw signals that are obtained from the systems can be used for such purposes like fault detection, feature extraction as well. However, in this case, the process is going to be very complicated and there will be other aspects that need to be taken into consideration, like the effect of noise during the measurement of the signal, the sensitivity and the validation of the sensors, etc. To avoid such problems, advanced signal processing techniques are used. By using these techniques, the domain can be transformed from time to frequency or other domain and the analysis can be done on more information-efficient data & platform. By using these transformed data/signals, it is easier to do the interpretation and analysis. One of the mostly used methods for this purpose is the Fast Fourier

Transform (FFT). However, under some certain conditions FFT will not be efficient. One of these conditions would be the application on the non-stationary signals. In the presence of the non-stationary signals, it is more convenient to use time-frequency analysis methods. First example of such a method would be the Short Time Fourier Transform (STFT).

Du and Wang [66] have used STFT for fault detection in high voltage inverters. Koo and Kim [67] used both Wigner Distribution (WD) and STFT for the vibration monitoring of the reactor coolant pump in the nuclear power plant. They also have used back-propagation neural network for classifying the status of the pump. Several states of the data have been provided to the neural network (normal conditions, shaft bow data, misalignment data, etc.) for training. After this process, the accuracy of the system was tested by other similar data. Methods like STFT and WD provide a mapping from time domain to time-frequency plane. Main problem with the STFT is that it uses a constant window through all frequency components of the signal. This property does not provide the possibility to obtain good frequency resolution for the analysis of low-frequency components. Additionally, it is not possible to obtain a good time resolution for high-frequency components. This problem is handled by using wavelet analysis instead of STFT. Wavelets can extract the features of a signal thanks to its multi scale analysis property. For this reason, the wavelets are more suitable for analyzing non-stationary signals [68]. Wavelets are described in detail in the second section of this study.

For fault diagnosis, wavelet applications are widely used. The applications are used on several areas where certain measurements can be obtained from the systems. These magnitudes can be vibration signals from the gears, rotating machines and other industrial systems; sound signals from moving parts of machinery; current or voltage signals from the electrical components of the systems, etc.

Early wavelet applications on fault diagnosis mainly analyze vibration signals. In 1993, vibration signals of the gearboxes have been analyzed by wavelets and it is shown that wavelet analysis was able to find the mechanical faults on the gearboxes [69]. Similar studies are still in place in the recent years; wavelets are used for denoising and extracting the pulse feature of the faulty signal [70]. The analysis of vibration signals were deeply investigated by Newland with several articles in 1994 [71-74]. After these studies, wavelets have been more popular in vibration analysis

and machine fault diagnosis. Vibration signals from the gears have been mostly investigated by wavelet analysis.

A study comparing several techniques for identifying gear damages showed that wavelet transform performed better than other techniques and it has less sensitivity to the load, speed and frequency of bandwidth utilized [75]. Other similar studies are done with wavelet applications for the tooth defects in gear systems [76-77].

Different types of cracks are other interest areas for wavelet analysis. Rotor cracks have been analyzed and it is found that these cracks are directly related with the changes in the vibration amplitudes [78]. Cracks in a dam can be identified and monitored by using wavelet transform [79].

Another property of wavelets in fault detection is the feature extraction possibility that they provide. In a narrower context of fault detection, the features are the characteristic properties of the faults that are analyzed. The main issue during the feature extraction with wavelets is the way to choose the wavelet coefficients that are representing the feature. Several studies have selected different ways of choosing these coefficients. One method is to identify a threshold function and only use the coefficients that are exceeding the threshold and deleting the rest [80]. Statistics based criteria is also used to identify the features [81].

After the identification of the features, the next step in such applications is the classification of the faults according to the extracted features. At this stage, several methods can be used, i.e. neural networks, Bayes classifiers, nearest neighbor methods.

Other signal processing techniques can also be used for feature extraction. Several studies are done in order to compare the effectiveness of the methods. In a study related with a power distribution system, both FFT and wavelets were used for identifying the location of the faults [82]. The results showed that wavelets were performing better than FFT for feature extraction. Another comparison about the performances of wavelet transform, FFT and Hartley transform showed that wavelets performed the best among the three techniques [83]. Techniques like principal component analysis (PCA) can help reducing the size of the feature space identified by the wavelet transform [84].

PCA is most widely used for projecting such coefficients to a lower-dimension space. In another example, normalization and PCA are used as pre-processing elements for feature extraction via wavelet transform [85]. Nonlinear principal component analysis can be done by using auto associative networks [86, 65].

Another application of wavelets and auto associative neural networks is the fault diagnosis method which is using vibration data of a gearbox casing as analysis material [87]. The DWT technique is used to obtain the detail coefficients. These coefficients are used as the training set for the AANN. The test set consists of both damaged and undamaged data. According to the output of the NN, the system is automatically classifying the data as damaged or not damaged [87].

1.2.3 Data mining and fault detection

Fault detection is a very early application field in data mining related studies. In 1993, Apte et al. [88] worked on a process improvement technique which was related to disk drive manufacturing. The main argument in the study was related to whether all or a portion of the disk drives could be identified as faulty before performing high-cost testing on them. They used a number of values obtained in the high-cost testing system called RunIn. These values and their corresponding failure types are known by the help of a system named RAES [89] and human analysts. These values were then applied to certain methods like k-nearest neighbour, linear discriminant, tree classification, and rule induction for classifying the type of the faulty component in the disk drive. The results for classification were far away from their objective. Additionally, feature extraction was done according to the rule induction techniques which identified very few numbers of key fields around 600 features. This part of the study was focusing on whether the disk drive is faulty or not. A rule was created as a result which mentioned that if certain key fields had a value bigger than 0.05, then the disk would be faulty with a certain probability where the probability would be calculated on all the analyzed samples. The study managed to reduce the time spent on high cost testing because of the physical structure used in the test environment [88].

Another study that is used for improving process quality and fault prediction using data mining techniques is focused on a production process in LCD manufacturing [90]. Here, a transactional database is used which contains information about the

products, the process that it is involved in, the start and end time of the processes. Besides this information, it also includes error codes if there are any. By using this information, certain association rules are created which combine the process numbers with process times and tries to predict which fault can occur with a certain probability. This is a sort of classification problem which uses rule creation in a database.

A study focusing on detecting welding flaws uses two clustering methods, i.e. fuzzy k-neighbor and fuzzy c-means and makes the comparison between them according to the success rates [91]. An extension of this study is using multi-layer perceptron (MLP) for classification of the welding data [92]. The data set is consisted of a number of samples which include 25 numeric attributes that were extracted in the authors' previous study [91]. The MLP structure that they used was consisted of 3 layers, having 25 nodes in both input and hidden layers and a single output node which is indicating whether the welding is faulty or not.

Reference [93] uses time-series data mining technique to identify and classify the faulty states of a poly-phase induction motor. It uses a technique called "Time – Stepping Coupled Finite-Element-State-Space method" for the simulation of the faulty case data (phase current waveform and time-domain torque profiles). By using this data, faulty cases are classified according to their characteristics using the time-series data mining technique. In the next phases of the study, real data is used as a proof of the time series data mining technique. A so-called time-delay embedding is used for transforming the torque time series to a reconstructed state space. After this transformation, radius of gyration is used for the distinguishment between the healthy and fault states. This paper is also focusing on the classification of the data states and checking whether the motor is faulty or healthy. It can easily be said that the problem that is trying to solve is classification. The authors have another study which focuses on a similar problem with the same technique [94].

1.3 Motivation Of The Thesis And Target Contribution

As seen in the previous section, the studies related to fault detection, feature extraction, data mining, signal processing are present in literature with several combinations. The main usage of data mining for fault detection is focusing on distinguishing the faulty states, devices, etc. from the healthy ones. Additionally, in

case of faulty scenario, the type of the faults is to be identified. These studies are dealing with either classification or clustering problems based on the available information. If the researcher has the classes matching with the data, the problem becomes a classification problem, if not, it is clustering. For classification, the methodology followed by these studies is using artificial intelligence systems which are capable of learning a certain status of the information i.e. the healthy state, and being tested on the faulty state or both healthy and faulty states.

Signal processing, especially wavelet transform, has several important roles in fault detection problem. One of them is data dimension reduction and data transformation. By the help of wavelet or other signal processing techniques, a different presentation of the original time signal is obtained. The other and important use of wavelet transform is the feature extraction. As mentioned in the previous section, feature extraction is mostly reducing the number of data which represents the characteristics of a certain state, i.e. the faulty state.

The common property among all the studies that are mentioned in the literature survey section is, in each fault detection problem, there is a need for using both the faulty and healthy states for solving the problem. The main motivation of this thesis is to propose a technique which uses the healthy state data only for identifying a possible fault in a system (here, an electric motor). For accomplishing this, we assume that the healthy state signal can contain features which lead to identifying a potential problem: a possible defect. Under this assumption, we use a special representation of the original time signal by using CWT. The reason for applying CWT is explained in the mathematical interpretation section. Having the special representation of the signal including the features related to a possible fault/defect, these features are extracted by the use of an AANN, since AANNs are capable of identity mapping and used effectively for feature extraction. In data mining perspective, the special representation of the signal using CWT allows us to identify, the *important, previously unknown* information hidden inside. Hence, it constitutes the second target contribution which is a different approach for data mining using signal processing techniques.

1.4 New Algorithm For Signal Based Data Mining Application

In this thesis, under the framework of data mining and signal processing techniques, a new algorithm is introduced for feature extraction and fault diagnosis. In order to accomplish this, two main processes are defined as data pre-processing and neural network application. In data pre-processing stage, signal processing techniques like continuous wavelet transform (CWT) and Fourier transform (power spectral density - PSD) are applied to the raw vibration signal which is obtained from the healthy state of an electric motor. This step allows us to represent the signal more effectively and makes it ready to be used by the second step: the neural network application. In the neural network application, an auto associative neural network (AANN) is trained by the new representation of the vibration signal which is the normalized PSD of the CWT at first scale. After the training, the AANN is recalled again by the PSD of the CWT at first scale, this time with a different set of data which has not been used in the training. This phase is for the identification of a threshold level for novelty detection. At the test phase, the PSD of the raw signal is given to the AANN. According to the error change at the output of the AANN at the test phase, a comparison is done with the defined threshold. The test sets that are exceeding the threshold level are identified as novelty i.e. a faulty state. This novelty represents the potential defects in the electric motor. The same method is applied to two different types of simulation data other than the vibration signal. The results again show that the method is identifying the novelty status for these data as well. In addition to these, a graphical user interface (GUI) is developed for automating the mentioned processes and hence, providing a simpler way to analyze the signals. The outputs related to the GUI are given in the Appendix G.

Since the vibration signal used during the whole process is obtained from the healthy state of the electric motor, it can be said that the proposed algorithm provides an effective and unsupervised fault detection system.

Even though data mining applications exist in literature for fault detection, this study introduces a different (signal based) approach to data mining by applying signal processing techniques and neural network applications in parallel.

1.5 Organization Of The Thesis

The outline of the thesis is listed as follows:

- Introduction
 - Purpose of the thesis, background study, new algorithm in the thesis
- Mathematical Methods
 - Spectral methods, Fourier transform, wavelet transform, artificial neural networks.
- Data Mining
 - Data mining concept, data types, functionalities, techniques.
- Experimental Study for Vibration Measurements and Sensor Validation
- Simulation for the Proposed Data Mining Algorithm and its Mathematical Interpretation
- Signal Based Alternative Approach To The Data Mining Concept
 - Data pre-processing, neural network application, summary of the algorithm.
- Conclusions and Discussions

2. MATHEMATICAL METHODS

In this section, the mathematical methods that are used in the study are introduced in detail.

2.1 Spectral Methods

The determination of the frequency bands that contain energy or power in a sample function of a stationary random signal is called spectral analysis. The mode in which this information is presented is an energy or power spectrum. It is the counterpart of the Fourier spectrum for deterministic signals. The shape of the spectrum and the frequency components contain important information about the phenomena being studied.

2.1.1 Discrete Fourier Transform

For calculating how much power a signal contains at a certain frequency, it is the Fourier Transform that needs to be calculated. This is called as the *spectral information* of that particular signal. Generally, there is no chance to measure the signal's values between $t = -\infty$ to $t = +\infty$. Only for some certain duration, the signal's value can be known. In addition to this, since there is no way of analogue measurement & interpretation of the signal by using a digital processor, the values of the signal can only be known at certain times $t = nTs$, where Ts is the sampling period. By using these samples, the Fourier Transform of the signal can be calculated. [49]

Even if the value of the signal is known at a certain interval, it is not enough to calculate the signal's overall Fourier Transform. In addition to this knowledge, there should be an assumption which is valid for the rest of the time. Generally, the most useful way for this is to assume that the signal has zero value outside this time interval. By this assumption, the Fourier transform can be calculated.

The Fourier transform of a function, $y(t)$, is

$$Y(f) = F(y(t))(f) \equiv \int_{-\infty}^{\infty} e^{-2\pi jft} y(t) dt \quad (2.1)$$

It is found that given a function, $y(t)$, which is zero outside the region $t \in [0, T]$, its Fourier transform can be expressed as:

$$Y(f) = \int_0^T e^{-2\pi jft} y(t) dt \quad (2.2)$$

Supposing that there are only N samples of the function taken at $t = k(T/N)$, $k=0, \dots, N-1$. Then the integral can be estimated by:

$$Y(f) \approx \sum_{k=0}^{N-1} e^{-2\pi jfkT/N} y(kT/N)(T/N) \quad (2.3)$$

If the frequencies of interest are $f = m/T$, $n=0, \dots, N-1$, then it is found that

$$Y(m/T) \approx (T/N) \sum_{k=0}^{N-1} e^{-2\pi jmk/N} y_k, \quad y_k = y(kT/N) \quad (2.4)$$

The Discrete Fourier transform (DFT) of the sequence y_k is defined as

$$Y(m) = DFT(\{y_k\})(m) \equiv \sum_{k=0}^{N-1} e^{-2\pi jfkT/N} y_k \quad (2.5)$$

The value of the DFT is (up to the constant of proportionality T/N) an approximation of the value of the Fourier transform at the frequency m/T [49].

2.1.2 Wavelet transform

Wavelet transform is basically defined as the inner product of the signal $f(t)$ and the basis function which is derived by scaling and shifting of the main wavelet function $\psi(t)$. This way, wavelet transform helps to represent the properties of the signal in both time and frequency domains. The basis functions are short-duration, high-frequency and long-duration, low-frequency functions. While calculating the wavelet transform, the signal can be better represented by keeping the duration short at its high frequencies and keeping the duration long at its low frequencies [50].

$\Psi(\Omega)$; being the Fourier transform of $\psi(t)$ and if $\psi(t) \in L^2(R)$

$$C_\psi = \int_{-\infty}^{\infty} |\Psi(\Omega)|^2 |\Omega|^{-1} d\Omega < \infty \quad (2.6)$$

If the inequity above is valid, $\psi(t)$ is called a wavelet function.

If $\psi(t) \leftrightarrow \Psi(\Omega)$ constitute a Fourier transform pair, then

$$\frac{1}{\sqrt{a}} \psi\left(\frac{t}{a}\right) \leftrightarrow \sqrt{a} \Psi(a\Omega) \quad (2.7)$$

where $a > 0$ is a continuous variable.

The dilates and the translates of the mother wavelet are known as wavelets.

In Figure 2.1 and Figure 2.2. different types of wavelet are shown. In the thesis, Symlet 4 is used as the mother wavelet in CWT since it's an orthogonal function.

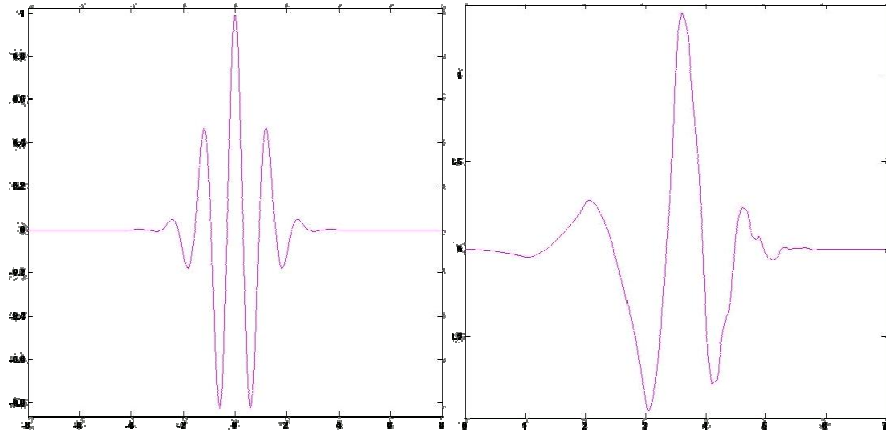


Figure 2.1 : Morlet and Daubechies 4 Wavelets.

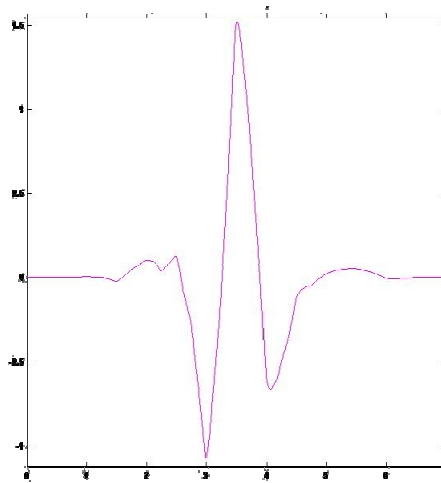


Figure 2.2 : Symlet 4 Wavelet.

In the time-frequency domain, narrowness in one domain necessarily implies a wide spread in the other [49].

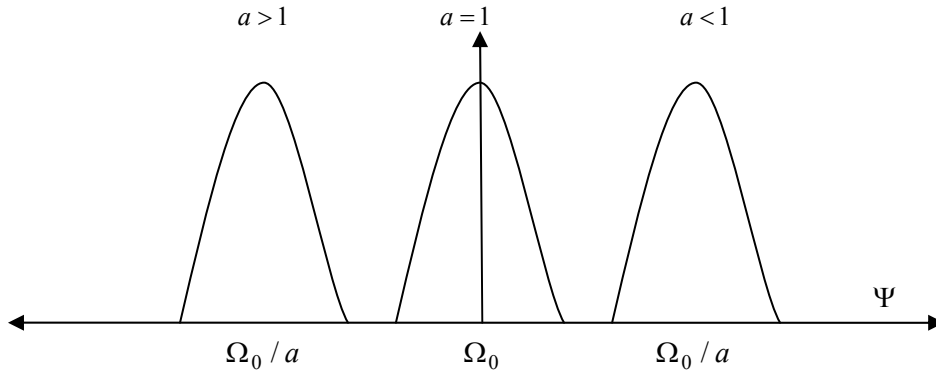


Figure 2.3 : Wavelet band-pass filters on a logarithmic frequency scale.

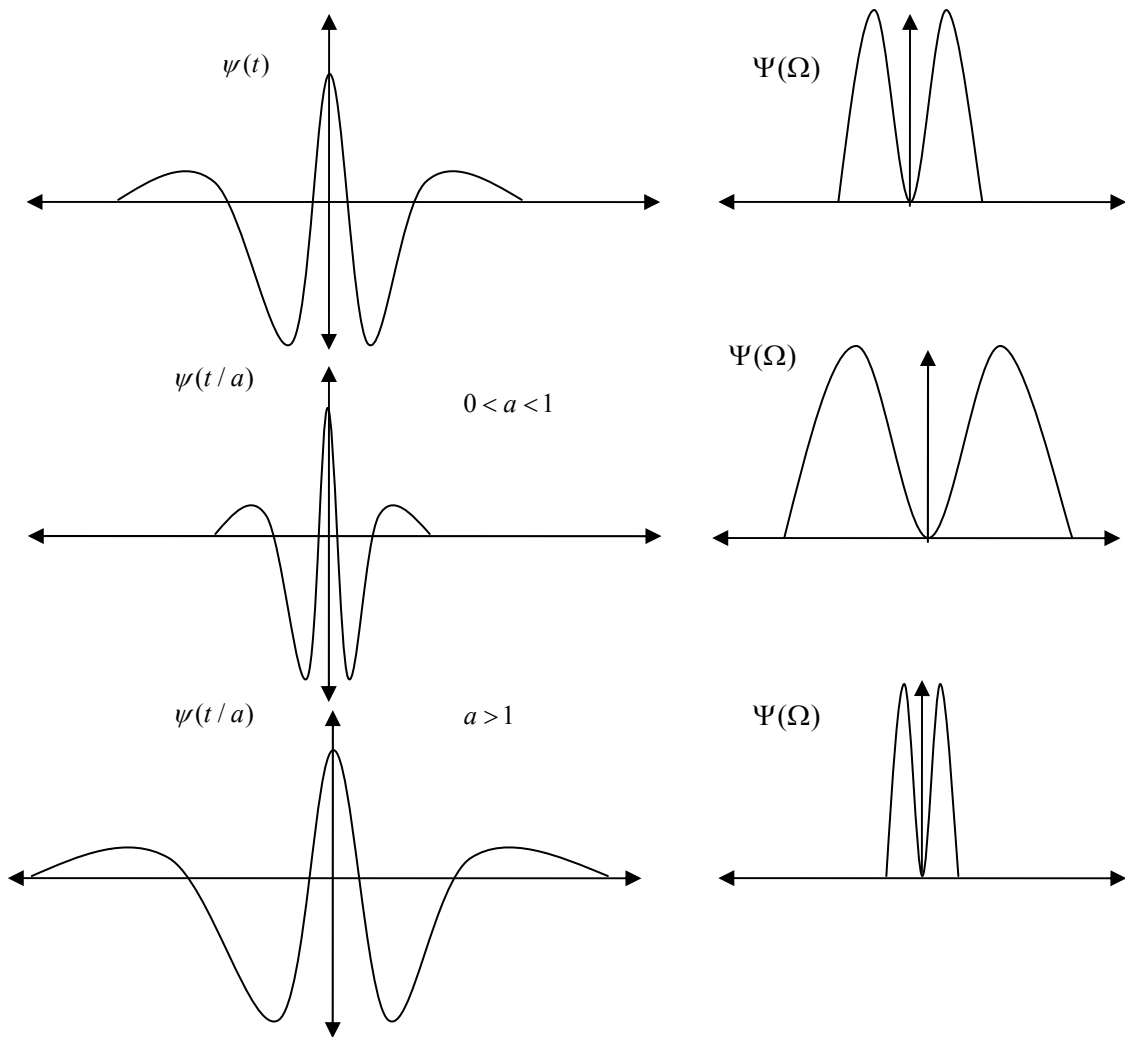


Figure 2.4 : Typical wavelet family in time and frequency domains.

The wavelet family is thus defined by scale and shift parameters a , b as:

$$\psi_{ab} = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (2.8)$$

As the values of scaling parameter a grow, the basis function corresponding to the mother wavelet gets wider. This is useful for low frequency purposes. On the other hand if the values of parameter a , the basis function becomes narrower and this is useful for high frequency. The effect of the scaling parameter on the wavelet band pass filters is shown in Figure 2.3. According to the change in scaling parameter a , the wavelet function $\psi(t)$ dilates or contracts in time, causing the corresponding contraction or dilation in the frequency domain. This can be seen in Figure 2.4. Consequently, it can be said that the wavelet transform provides a flexible time-frequency resolution. In Figure 2.5 the resolution cells can be seen in the time frequency plane.

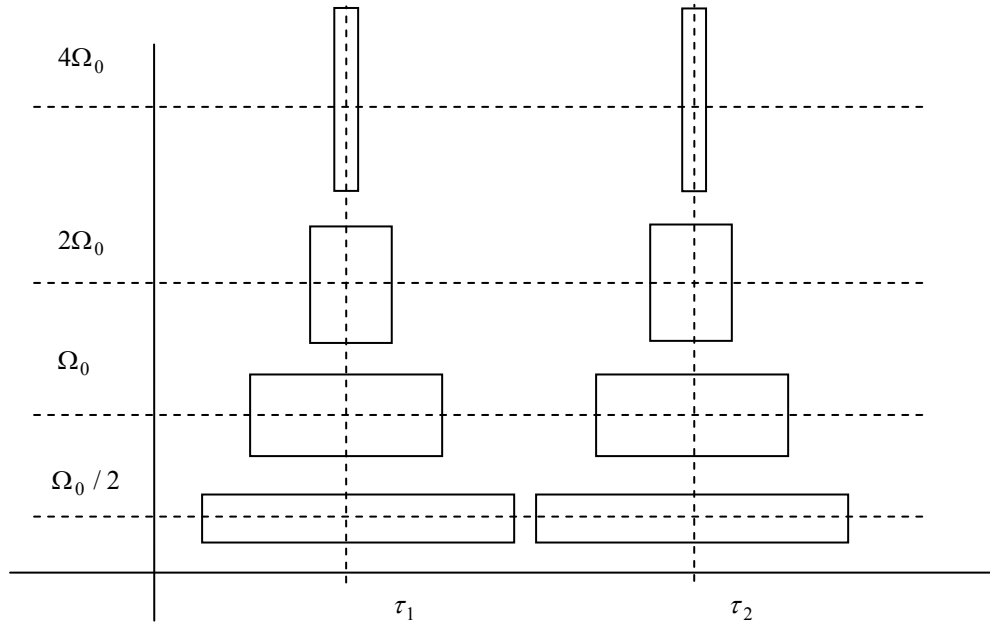


Figure 2.5 : Time-frequency plane showing resolution cells for wavelet transform.

In the next section, continuous wavelet transform is defined and an *admissibility* condition is developed on the wavelet needed to ensure the invertibility of the transform. The discrete wavelet transform (DWT) is generated by sampling the wavelet parameters (a, b) on a grid. The signal can be reconstructed from its transform if the coarseness of the sampling grid is suitable. A fine grid mesh would permit easy reconstruction, but with evident redundancy, i.e., *oversampling*. A too-coarse grid could result in loss of information. The concept of *frames* is defined to address these issues [50].

2.1.3 Continuous wavelet transform

The continuous wavelet transform (CWT) is defined by Equation 2.9 in terms of dilations and translations of a prototype or *mother* function $\psi(t)$. In time and Fourier transform domains, the wavelet is:

$$\psi_{ab}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \leftrightarrow \Psi_{ab}(\Omega) = \sqrt{a} \Psi(a\Omega) e^{-jb\Omega} \quad (2.9)$$

The CWT maps a function $f(t)$ onto time-scale space by

$$W_f(a, b) = \int_{-\infty}^{\infty} \psi_{ab}(t) f^*(t) dt = \langle \psi_{ab}(t), f \rangle \quad (2.10)$$

$W_f(a, b)$: Wavelet transform of the signal $f(t)$

$f(t)$: Signal to be transformed

$\psi(t)$: Wavelet function

The transform is invertible if and only if the *resolution of identity* holds and is given by the superposition:

$$f(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_0^{\infty} \frac{da db}{a^2} W_f(a, b) \psi_{ab}(t) \quad (2.11)$$

Where

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\Psi(\Omega)|^2}{\Omega} d\Omega \quad (2.12)$$

provided a real $\psi(t)$ satisfies the *admissibility* condition. The wavelet is called *admissible* if $C_\psi < \infty$.

$\psi(t)$ analysis function should be short and oscillatory: i.e. it must have zero average and decay quickly at both ends. This restriction ensures that the integral in (2.10) is finite [51].

Thus, $\psi(t)$ behaves as the impulse response of a band-pass filter that decays at least as fast as $|t|^{1-\epsilon}$. In practice, $\psi(t)$ should decay much faster to provide good time-localization.

While we can speak of the Fourier transform of a function in Fourier analysis, wavelet transform can not be directly defined [52]. The transform is done according to the mother wavelet. For a function $f(t)$, there can be two different wavelet transforms for two different wavelets. For this reason, the choice of the wavelet is important. In most of the literature, mother wavelets are orthogonal or dyadic.

Similar to the Fourier transform, the wavelet transform is a linear filtering process. If a signal $f(t)$ is passed through a filter whose transfer function is $a^{1/2}\Psi^*(a\Omega)$, its output at time t , will be $W_f(a, b)$.

2.2 Artificial Neural Networks

Here, the general concept of artificial neural networks is given and auto associative neural networks are described since they are used in this study.

2.2.1 Definition

An Artificial Neural Network (ANN), often just called a “neural network” (NN), is a mathematical model or computational model based on biological neural networks, in other words, is an emulation of biological neural system. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase [53].

A biological neuron and its components is shown in Figure 2.6.

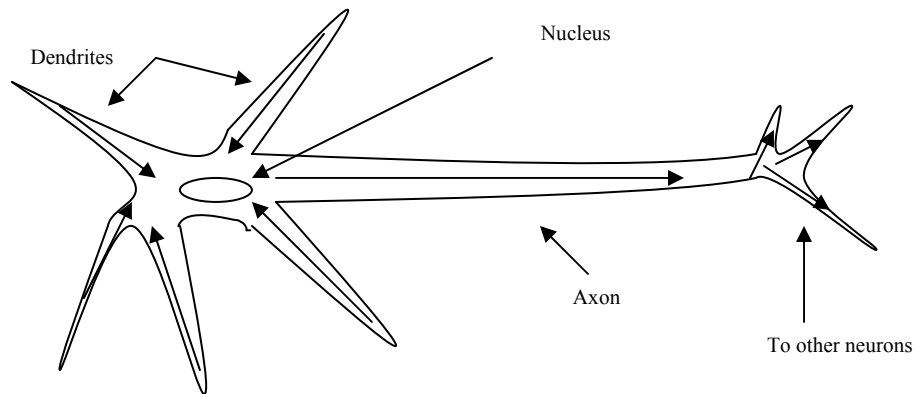


Figure 2.6 : Structure of a biological neuron.

An artificial neuron can be seen in Figure 2.7.

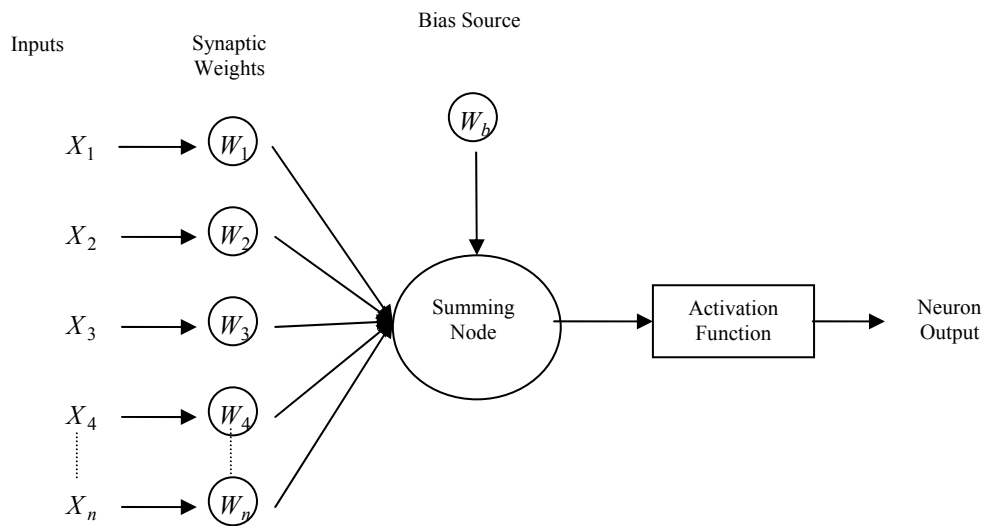


Figure 2.7 : Structure of an artificial neuron.

2.2.2 Main concepts

The Artificial neural networks are consisted of artificial neuron layers that are connected to each other in different topologies. Artificial neurons are very similar to biological neurons. While biological neurons have dendrites for getting information from other neurons, the artificial neuron has inputs that are corresponding to these dendrites. The biological neuron uses its nucleus for making the decision inside itself; the artificial neuron gets the inputs, multiplies with the correspondent weights, adds these results together to obtain a sum and gives an output according to the activation function (AF) it uses. The activation functions' output can be either 0 or 1 or a value in between -1 and 1. In Figure 2.8 and 2.9, the threshold function and sigmoid function are shown respectively as examples for activation functions.

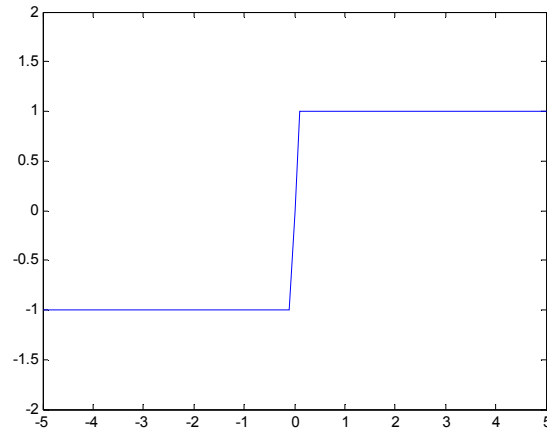


Figure 2.8 : Threshold function.

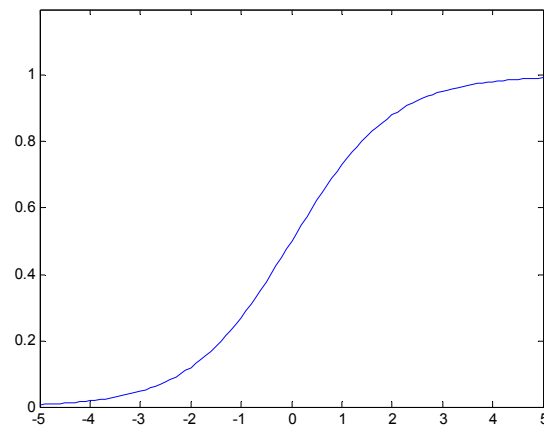


Figure 2.9 : Sigmoid function.

The networks use learning methods for adjusting the weights of the neurons. The types of these methods are: supervised, unsupervised and reinforcement methods [61]. In supervised learning, the network is provided with inputs and the matching outputs against these inputs. In unsupervised training, there are no matching outputs against the given inputs. While supervised learning is like classification, the unsupervised learning can be treated as clustering. Unlike these two methods, the reinforcement learning is a mixed method in which the network gets feedback from the environment and adjusts the parameters according to the feedback of the environment.

In addition to the learning types, the other important aspect is the learning algorithm used during the training of the networks. There are several learning algorithms like, gradient descent method, resilient backpropagation [59] or Levenberg-Marquardt algorithm [60]. Error-correction learning algorithms [61] adjust the weights of the neurons in order to minimize the error function. The definition of sum of squared error –SSE, or mean square error-MSE are introduced in the equations 2.13 and 2.14.

$$SSE = \sum_{p=1}^P \sum_{q=1}^Q [D(q) - a(q)]^2 \quad (2.13)$$

where, P=patterns, Q=output neurons

$$MSE = \frac{1}{Q} [SSE] \quad (2.14)$$

2.2.3 Types of neural networks

In 1958, Rosenblatt [62] introduced the simplest form of a neural network, the perceptron, used for the classification of patterns said to be linearly separable. Single-layer perceptrons (SLP) are formed by having more than one neuron in the layer, giving the chance of classification for more than two classes. Multi-layer perceptrons (MLP) represent a generalization of the single-layer perceptron. MLPs typically consist of an input layer, one or more hidden layers and an output layer [63]. There are three distinctive characteristics of a MLP. First, the model of each neuron in the network includes a nonlinear activation function. One commonly used form of nonlinearity is the logistic function, which is used in the neural network application of this thesis as well. Logistic function is defined in Equation 2.15.

$$y_j = \frac{1}{1 + e^{-v_j}} \quad (2.15)$$

Where, v_j is the induced local field (i.e., the weighted sum of all synaptic inputs plus the bias) of neuron j , and y_j is the output of the neuron. The presence of nonlinearities is important because the network could be justified by a single layer perceptron. The second characteristic of the MLP is that it includes one or more hidden layers in the topology. The presence of the hidden layers allows the network to learn complex tasks by extracting progressively more meaningful features from the input patterns. The last important characteristic of MLPs is it shows high degrees of connectivity. A change in the connectivity of the network requires a change in the population of synaptic connections or their weights [63]

ANNs can also be classified according to the way that the information flows in the network topology. In this perspective, the networks can be either feedforward or feedback (recurrent). In feedforward networks, the information flows in the forward direction; while in the recurrent networks the flow is both forward and backward.

Another classification of the neural networks is according to the tasks that they perform. According to this perspective, the networks can be either pattern-classifiers or pattern associators. Pattern classifiers use input vectors and give a value that classifies that vector. An example would be giving certain sensor values as input and classifying them according to their status (healthy or faulty). While pattern classifier networks classify the input vectors, the pattern-associator networks get vectors as an input, and again, create another vector in the output. Pattern associator networks are divided into two groups: hetero-associative (HA) and auto-associative (AA) networks. Hetero-associative networks use certain types of vectors as the input and create completely different type of vector in the output. An example would be a speech-to text application where the input would be a sound file and the output would be the text corresponding to it. In auto-associative networks, the input and the output vectors are the same. An example of this auto-associative network would be training the network with several networks and testing it with a corrupted vector, which is missing certain data. In this case the auto-associative network would correct the missing data. In the following section, the auto-associative networks are discussed, since the network used in this thesis is auto-associative.

2.2.4 Auto-associative neural network

A special type of feed forward neural network is the auto-associative neural network (AANN). The main function of the AANN is identity mapping (the target vector is identical to the input presented to the network) [54, 55]. This property represents a form of unsupervised training, as no independent target data is provided [87]. The specification of the auto association network is that it contains a bottleneck layer between the input and output layers. This layer provides the data compression and is very suitable for a variety of data screening tasks. AANNs are able to reduce the noise and this pre-processing property allows that sensor-based calculations can be done even when failures and biases exist in the sensors. Two different representations of AANN topologies can be seen in Figure 2.10 and Figure 2.11.

One of the applications of feed forward neural networks is the creation of input-output models with using backpropagation training. Its purpose is to create a mapping between the input and the output. As mentioned previously, AANN provides the same mapping in a specific case where the input vector is same as the output vector. Since the input is identical to the output, the mapping is called an identity mapping. At a first glance, it can seem useless to provide such a mapping. However, it is proven that this mapping is useful in many cases like data screening, noise reduction and replacement of missing values [64-65].

The main concept of using auto-associative networks is that they can not create a perfect mapping between the input and output layers. This property demonstrates the useful property of the network. Because if the network was to learn the identity function exactly, then there would be no transformation between the layers and the network would make no sense. Since the auto-associative networks include a bottleneck inside, they can not learn the identity mapping perfectly. This layer is called as the bottleneck layer because it includes less number of nodes compared to the input, and consequently, the output layers.

Several training algorithms can be used for AANNs. Gradient descent, resilient backpropagation or Levenberg-Marquardt are examples for these algorithms. In this thesis, resilient backpropagation is used because of its fast convergence.

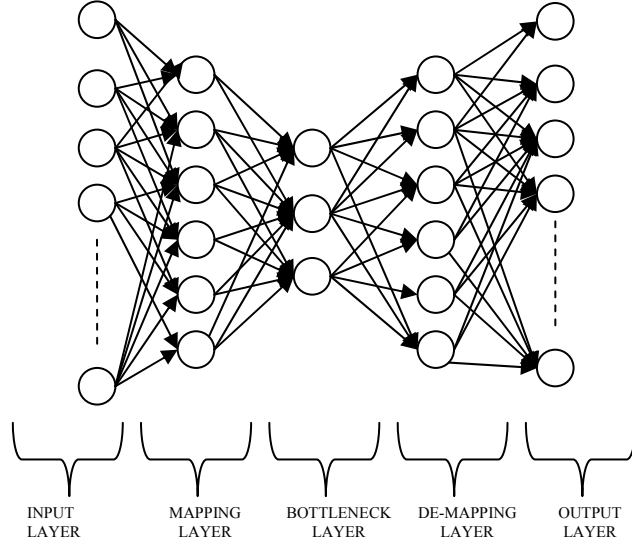


Figure 2.10 : Example AANN topology with three hidden layers.

Assuming that there are n nodes in the input and output layers; the number of nodes present in the bottleneck layer, m , should satisfy $m < n$. In this perspective, the network has to reconstruct the input vector at the output layer under this dimension reduction constraint.

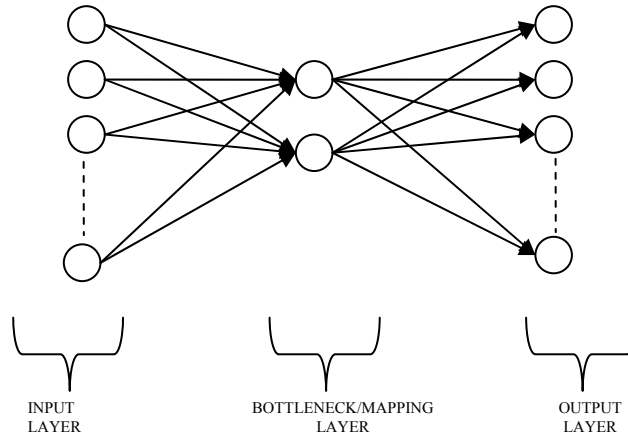


Figure 2.11 : Example AANN topology with single hidden layer

The basic principle of resilient backpropagation is to eliminate the harmful influence of the size of the partial derivative on the weight step. During the weight update the direction is identified by the sign of the derivative. The size of the weight change is determined by a weight-specific ‘update value’ $\Delta_{ij}^{(t)}$ [59].

$$\Delta w_{ij}^{(t)} = \begin{cases} -\Delta_{ij}^{(t)}, if \frac{\partial E^{(t)}}{\partial w_{ij}} > 0 \\ +\Delta_{ij}^{(t)}, if \frac{\partial E^{(t)}}{\partial w_{ij}} < 0 \\ 0, else \end{cases} \quad (2.16)$$

3. DATA MINING

3.1 Data Mining Concept

The amount of data that are present in the files, documents, data warehouses, databases has significantly grown. Hence, the need for representing this huge amount of data in a more efficient way has also grown. In addition to this, identifying the important or valuable knowledge in the data has also become much more important.

The terms data mining and knowledge discovery in databases are defined as the process of obtaining the important, previously unknown and relevant data from a data source like database, data warehouse, a collection of documents etc. Some sources refer to data mining and knowledge discovery in databases as synonyms; data mining is a part of the knowledge discovery process which can be seen in Figure 3.1.

The knowledge discovery process is consisted of a variable number of steps which has the raw data as the input and the useful knowledge as the output. The main sections of the process can be given as:

- **Data cleaning:** This stage is performed to make sure that the data is cleared from various types of noise
- **Data integration:** If there are more than one data sources that need to be evaluated, these sources are merged in the same source
- **Data selection:** Even the data is cleaned and merged (in case of multiple sources), a decision should be done in order to select the data to be analyzed.
- **Data transformation:** The data needs to be transformed before the proposed technique is applied. An unsuitable type of data may not be used in several data mining techniques.
- **Data mining:** This is the main important step to apply the technique for extracting the potentially useful data.

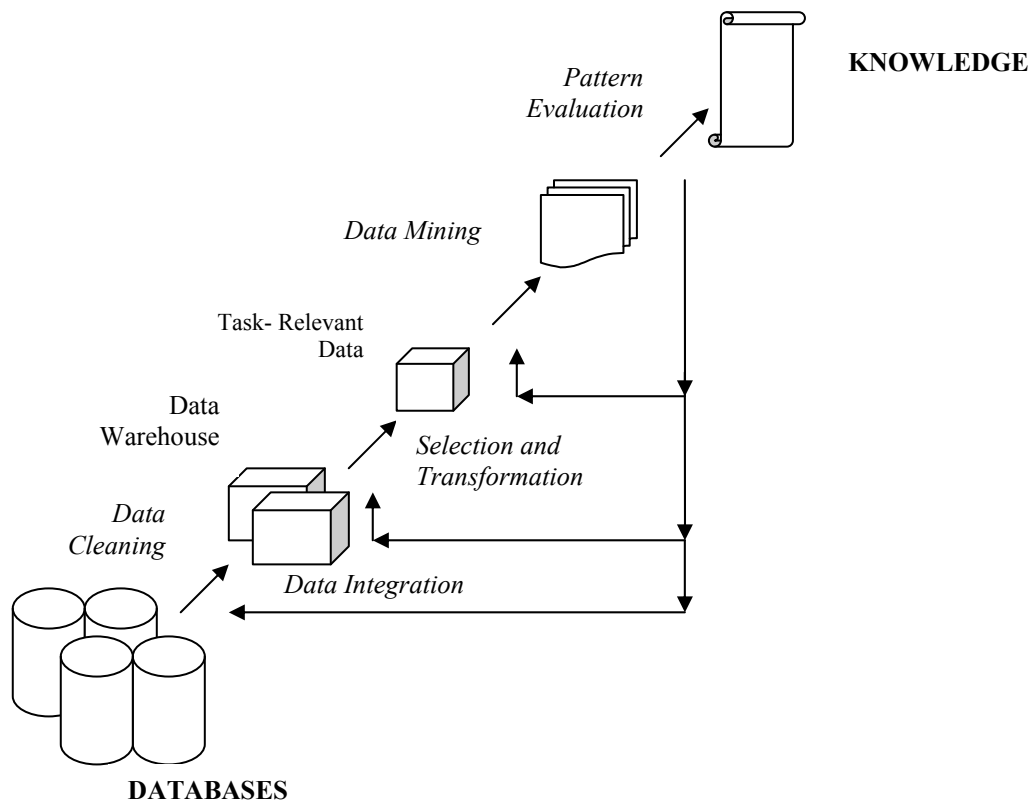


Figure 3.1 : Knowledge discovery process.

- **Pattern evaluation:** After identifying the potentially useful data in the data mining section, here, selection of the interesting patterns are evaluated according to the purpose of the study.

- **Knowledge representation:** This step is used in order to represent the discovered patterns/knowledge. Various methods can be used like charts, tables, etc. Even though it is not a vital step, it needs to be used to realize the findings.

It is not necessary to use the mentioned steps one by one. Some of them can be applied together like data cleaning and data integration to form the pre-processing part or data selection and data transformation can be combined as well. The examples can vary according to the data type being used, the techniques to be applied etc.

The knowledge discovery is not a one-shot process. After having the results after one cycle, changes can be made. The used data can be changed, new data can be merged or the data mining technique can be modified for obtaining different results. [56].

3.2 Data Mining Input Data

Data mining can be applied to different types of data. The short descriptions of the data sources are given below:

- **Flat files:** These are the most common used data sources which include text or binary files. The examples of flat files can be news articles, digital representations of time series data, sensor measurements, etc.

- **Relational Databases:** Relational database can be described as a regular database which is consisted of organized data between several tables. The data included in these different tables are connected to each other by the help of keys.

In relational databases, mainly the structured query language (SQL) is used. SQL can be used to retrieve, update or delete the records in a database.

- **Data Warehouses:** A Data Warehouse is used for collecting similar kind of data from different data sources in a central location.

- **Transactional Databases:** Transactional databases contain transactions which include a timestamp (the time the transaction was done), an identifier and a list of items.

- **Multimedia Databases:** The definition of multimedia includes pictures, audio, video and text. The multimedia databases can contain one or more of these items.

- **Time-Series Databases:** Time-series databases generally include data which change in real-time. The weather temperature change, the stock market data change, the change of the seismic waves, different types of sensor data can be given as examples for time-series data. Usually, the data is constantly flowing into the database by time. This creates a difficulty while analyzing the data in real-time.

Another difficulty would be to sample and represent the data in digital format. Time series data mining is mainly focused on finding the trends in the time-series data, prediction of the coming data structure, finding the trends which may come up in the future, etc. In this thesis, time-series signals (vibration signals obtained from the electrical motor) are used. Figure 3.2 and Figure 3.3 show two examples of time-series data.

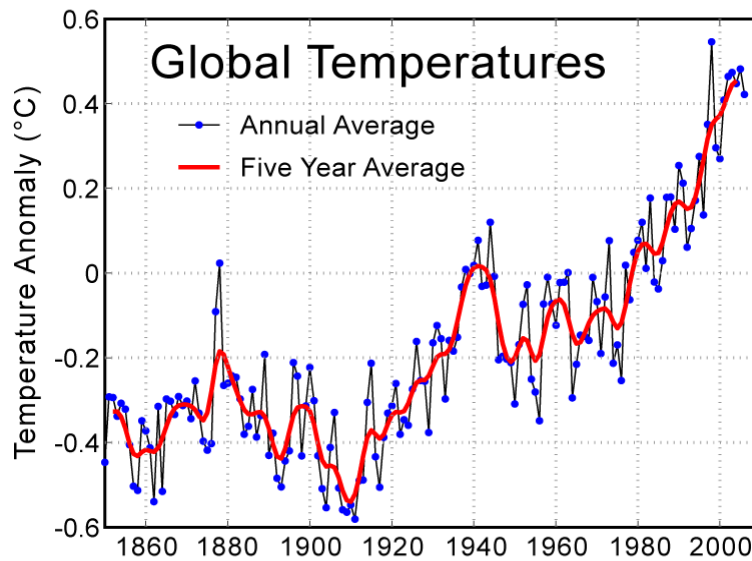


Figure 3.2 : Global temperature change [95].

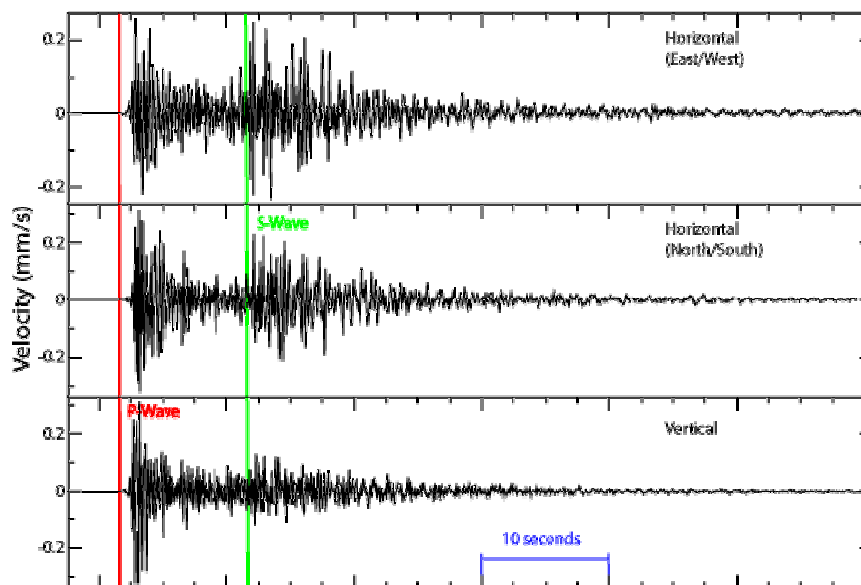


Figure 3.3 : Seismic wave measurement via a sismogram [96].

· **World Wide Web:** World Wide Web is the largest and the most dynamic data source ever. The WWW can be analyzed in three main concepts: its content, structure and the usage of it. Data mining in WWW or sometimes called Web mining mainly deals with these three concepts.

3.3 Data Mining Functionalities

The properties of the data mining techniques define the patterns or hidden information that is to be discovered. Mainly there are two types of output that the data mining functionalities provide. First is related to the patterns that are related to the current data available. The second is the prediction of the unknown patterns that are to be seen in the future.

The following list describes the main data mining functionalities and the output of them:

- **Characterization:** The characterization function of data mining is related with the summary of the objects in a class. By the characterization method, the characteristic rules are created.
- **Discrimination:** Discrimination can be considered as an extension of characterization. In discrimination case, there are two classes that are compared in terms of the general features of objects: the target class and the contrasting class.
- **Association analysis:** Association analysis, also known as association rules, is about finding the items that appear in the same transactions in transactional databases.
- **Classification:** Classification is the process of organizing several items in classes that share the similar properties. Classification is a supervised process. This means that the user initially knows a set of items and the classes that they belong to. By this information, the classification technique is trained. After the training, the items are given as inputs to the classification system for classification.
- **Prediction:** Prediction is the process of estimating or forecasting an event on which there is no certain valid information. In case of time series data, the issue is to predict the upcoming event or pattern.
- **Clustering:** Clustering can be identified as unsupervised classification. The difference between classification and clustering is that in clustering it is not known to which class an item or list of items belong to.

3.4 Data Mining Techniques

In this section, data mining technologies are represented in terms of data retention; whether the data is retained or discarded after it has been mined [57]. The breakdown is shown in Figure 3.4.

In early approaches to data mining, the data set was maintained for future pattern matching. The retention-based techniques only apply to tasks of predictive modeling and forensic analysis, and not knowledge discovery since they do not distill any patterns.

Approaches based on pattern distillation fall into three categories: logical, crosstabulation, and equational. The logical approach deals with both numeric and non-numeric data. Equations require all data to be numeric, while crosstabulations work only on non-numeric data.

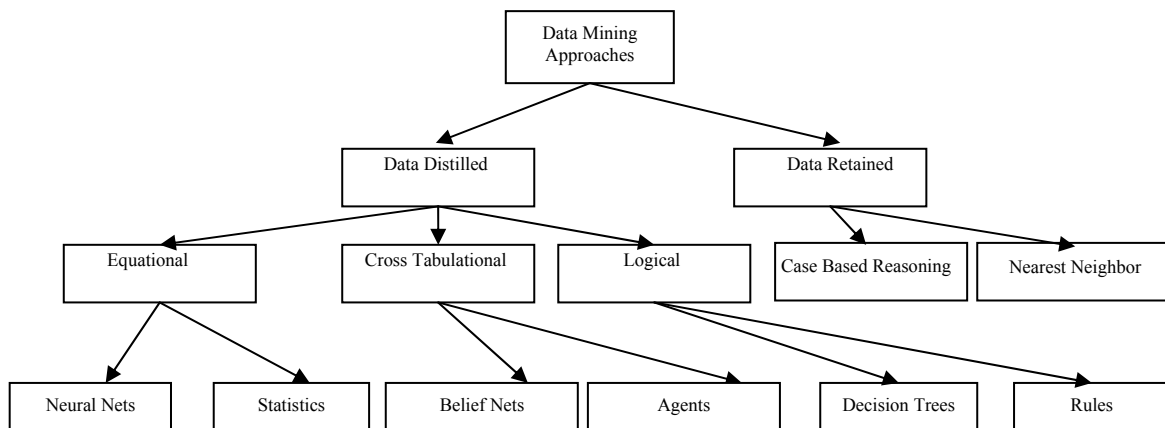


Figure 3.4 : Data mining technologies based on retention of data.

As seen in Figure 3.4., there are several data mining techniques based on the type of data mining functionality (classification, clustering, prediction, regression, etc.). These techniques include, but are not limited to neural networks, regression model, naïve bayes, decision trees, fuzzy logic, etc. [57].

In this study, the proposed methodology for the signal based data mining approach can be shown in Figure 3.5.

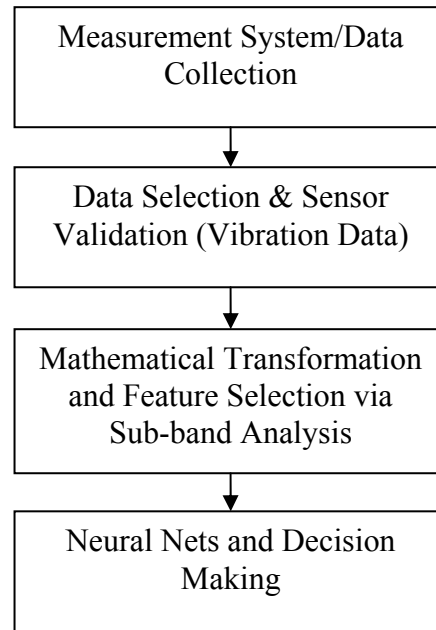


Figure 3.5 : Proposed methodology in the thesis.

4. EXPERIMENTAL STUDY FOR VIBRATION MEASUREMENTS AND SENSOR VALIDATION

In this section, the setup is described which is used for obtaining the experimental data for the study. The setup was created by Prof. B.R. Upadhyaya in the Maintenance and Reliability Center at the University of Tennessee, Knoxville in the USA. The measurement data is used in this study by their permission. The setup includes several measurements of different quantities, like temperature, vibration and current obtained from an electric motor. In this study, vibration signals are used as the data source. Since there is certain need for comparing the feature extraction of the study and the actual faulty state of the electric motor, an artificial aging process is also described.

4.1 Motor Aging Process And The Creation Of The Artificial Bearing Fault

The purpose of the aging process is to expose the electric motor to extensive heat, mechanical and electrical strain.

This experimental study is focused on bearing faults based on current transmission in electric motors.

In order to get the statistical distribution in the aging tests, three 5-HP motors have been tested for bearing fault mode. These 5-HP 4-poled motors have been provided by *Emerson Electric Company*. The motors are triangle connected and their plate data is given in Table 4.1.

The dimensions of these 3-phase motors are given in Table 4.2.

The performance tests of the motor are done on an adjustable platform with 3kW, 1800 rpm dynamometer. The dynamometer is induced with 0-2A and 0-200V and loaded with a 21.63 Ohm resistor. Test diagram can be seen in Figure 4.2. The motor is driven by changing the induction current and voltage. A vibration accelerometer and tachometer is used for the motor load measurements.

Table 4.1: Plate data of the motor

Motor Type	Premium Efficiency Motor
Manufacturer	US Electrical Motors Division of Emerson electric Co. St. Louis, MO
Model Number	7965B
HP	5
Motor speed (rpm)	1750
Phase	3
Body Number	184T
Body Type	TCE
Volt	230 / 460
Amp	13.0 / 6.5
Service Factor	1.15
Hertz	60
NEMA Design	B
Isolation Class	F
Maximum Environment Temperature	40 °C
Power factor	82.5
Efficiency	90.2
Maximum KVA _r	1.4
Short end Bearing Type	6206-2Z-J/C3
Process End Bearing Type	6205-2Z-J/C3
ID #	Z08Z177R190F /

Table 4.2: Dimensions of the motor

Dimension	Value (cm)
A	18.1
B	17.2
C	40.2
D	11.43
2F	14
H	1.03
Max N	2.7
Max O	21.3
Max P	24
U	2.2
Max AA	1.9
SQ Key	0.6

Bearing faults occur on the bearing surfaces because of the defects of the electrical sparks. These surface defects cause the vibration levels rise and create faulty state. The test setup for demonstrating the electrical motor discharge from the shaft to the bearing is shown in Figure 4.1. External 20 V and 30 A is applied to the shaft and the motor is run for 30 minutes without load.

In addition to applying voltage to the shaft for aging the bearing, thermal and chemical aging process is also applied to the electric motor.

Thermal aging is done by heating the motor at a constant temperature of 140 °C.

Each motor aging cycle is defined as follows:

1. 5-HP motor is run in the bearing defect test setup for 30 minutes.
2. The motor is inserted into an industrial oven at 140 °C.
3. The stationary motor is heated at 140 °C for 72 hours. Then, it is cooled for 6 hours.
4. The motor is inserted into water and then again inserted into the heater for another 72 hours of heating. This causes corrosion in the motor parts.
5. The motor is cooled for 6 hours.
6. The motor is tested in the load performance test setup.

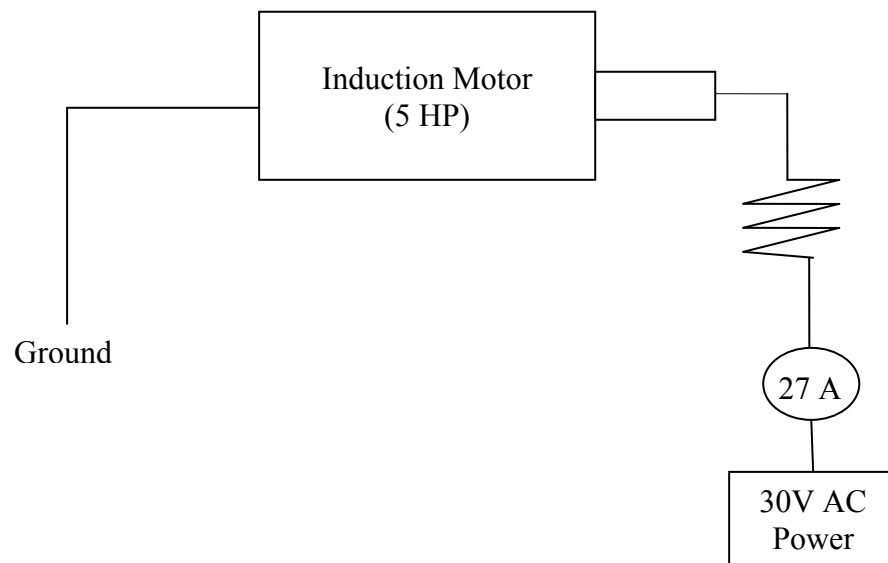
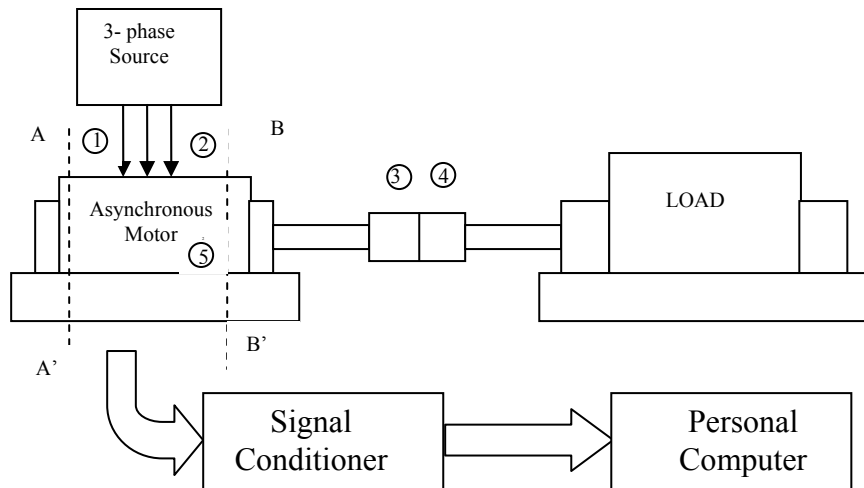
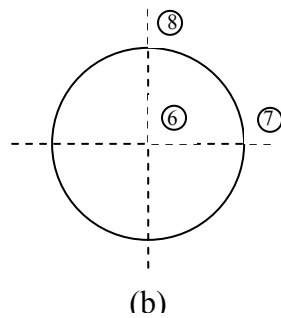


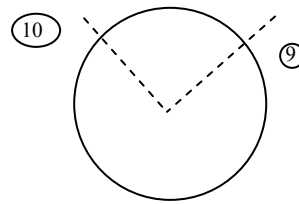
Figure 4.1 : Schematic of the electrical motor bearing EDM setup.



(a)



(b)



(c)

Figure 4.2 : Motor load testing and data acquisition system a) experimental setup configuration; b) Cross section (A-A') at short end; c) cross section (B-B') at pulley end.

The Electric motor is tested with various loads after each aging cycle. Parameters like mechanical (vibration), electrical and thermal are measured for identifying the status of the motor. The data acquisition system records the following measurements:

- Temperature measurements from 12 different places from the motor
- 3-phase motor voltage
- 3-phase motor current
- Load current
- Load voltage
- Motor speed
- Load momentum

- Accelerations in 6 different places of the motor

In this thesis the measurements from the accelerators are used as a data source of which structure will be explained in detail.

The accelerometers that are used in this experimental setup are fixed in the following locations on the motor:

- Process End, 2:00 o'clock location
- Process End, 10:00 o'clock location
- Short end vertical
- Short end horizontal
- Short end axial
- Cover box

These locations can be seen in Figure 4.2(c).

4.2 Motor Load Performance Test And Data Acquisition System

In the previous section, it can clearly be seen that there are 16 measurements (excluding temperature) in the setup. All these measurements are done in 12 kHz sampling frequency for 10 seconds for each channel and for each aging cycle. These measurements are stored in a data matrix which has 120000 columns (12kHz*10 seconds) and 16 rows. The corresponding quantities for these rows are listed as following:

1. V_{xy} (Volt)
2. I_x (Amp)
3. V_{yz} (Volt)
4. I_y (Amp)
5. V_{zx} (Volt)
6. I_z (Amp)
7. V_{Load} (Volt)
8. I_{Load} (Amp)
9. Speed (RPM)
10. Momentum (Nm)
11. Body Vibration (g)
12. Short End Axial Vibration (g)

13. Short End Horizontal Vibration (g)
14. Short End Vertical Vibration (g)
15. Process End 10:00 Clock Location Vibration (g)
16. Process End 2:00 Clock Location Vibration (g)

In this study the data from Process End 2:00 Clock Location vibration is used (Channel 16). This sensor is identical to Process End 10:00 Clock Location (Channel 15). Figures 4.3 and 4.4 show all 16 measurements.

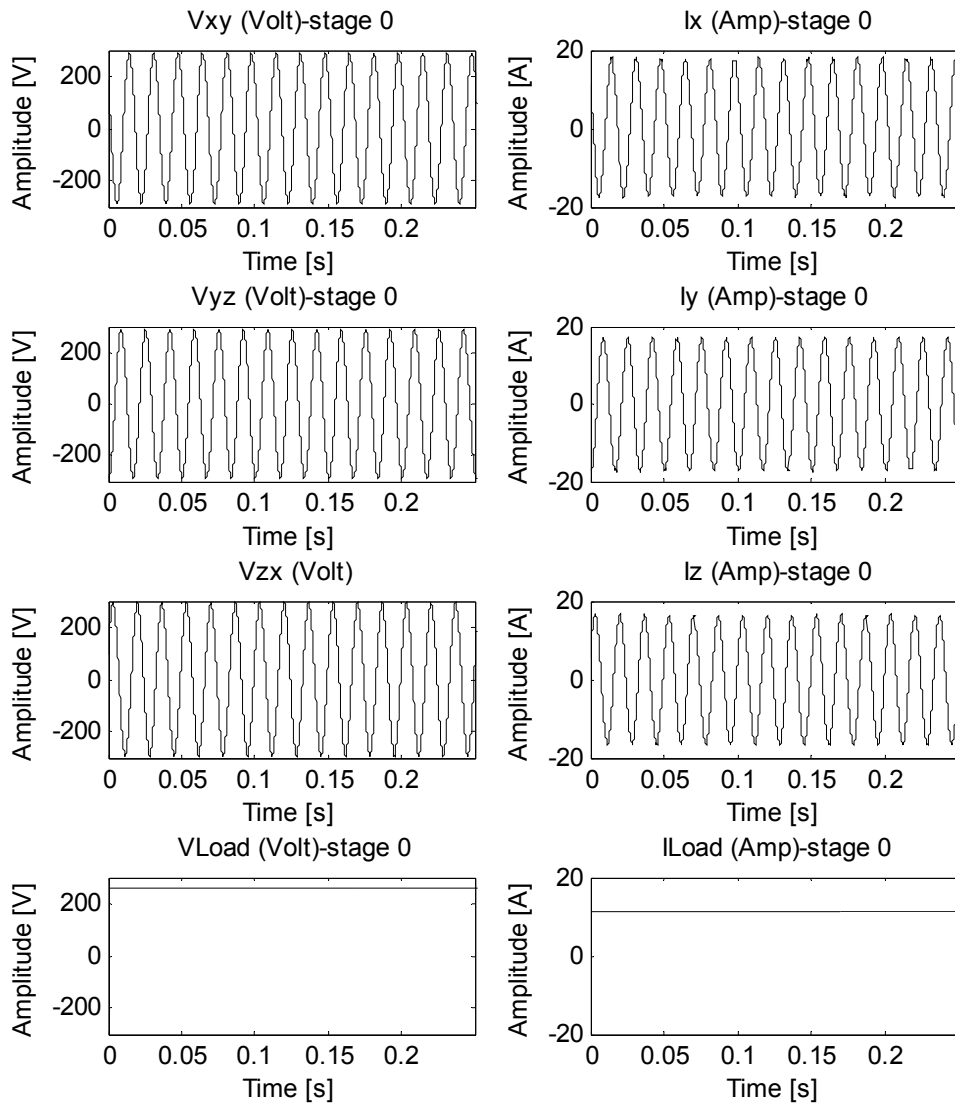


Figure 4.3 : First 8 channels' output.

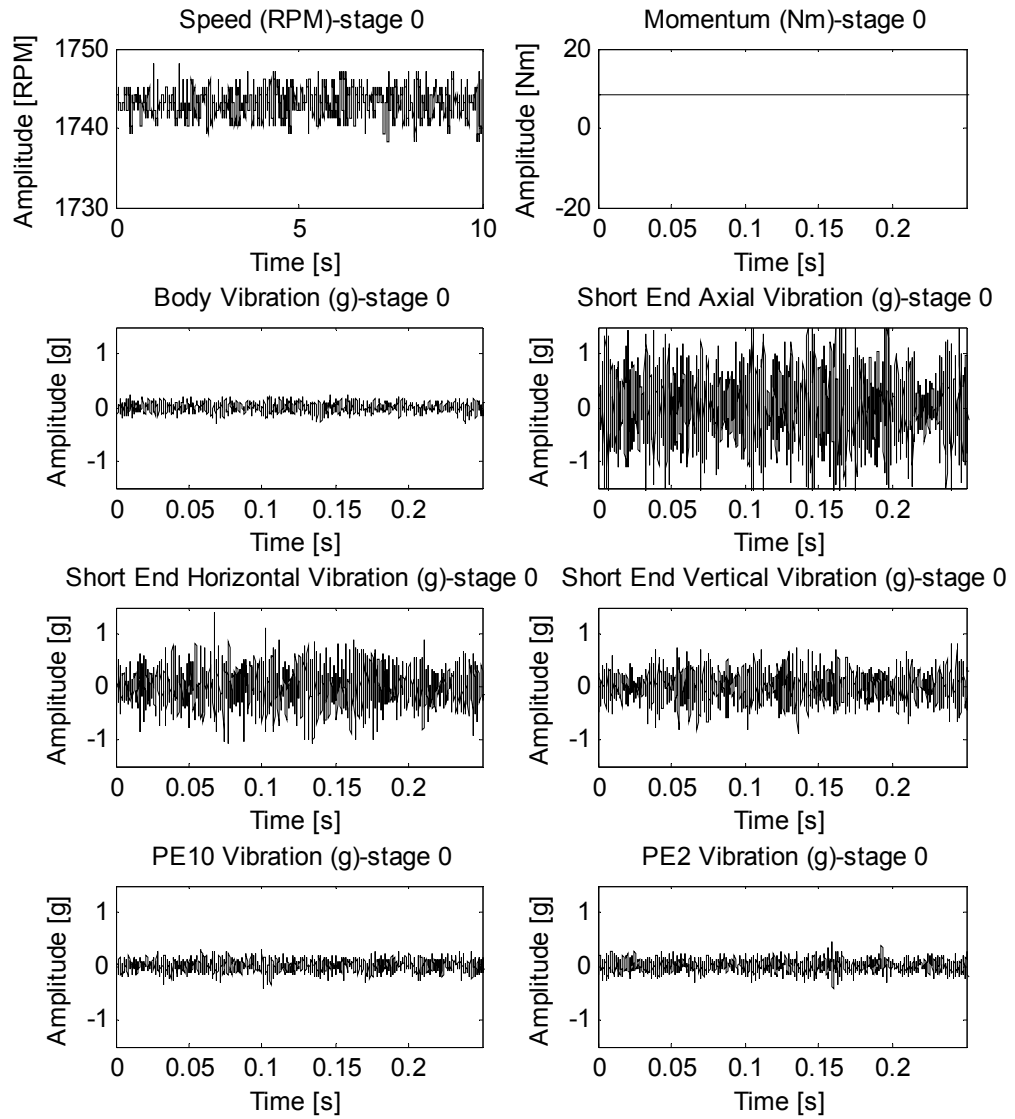


Figure 4.4 : Last 8 channels' output.

In the following figures (4.5 to 4.12), the signals from the two vibration sensors (#10-PE2 and #9-PE10) are shown throughout the aging stages (0-healthy case, 7-faulty case)

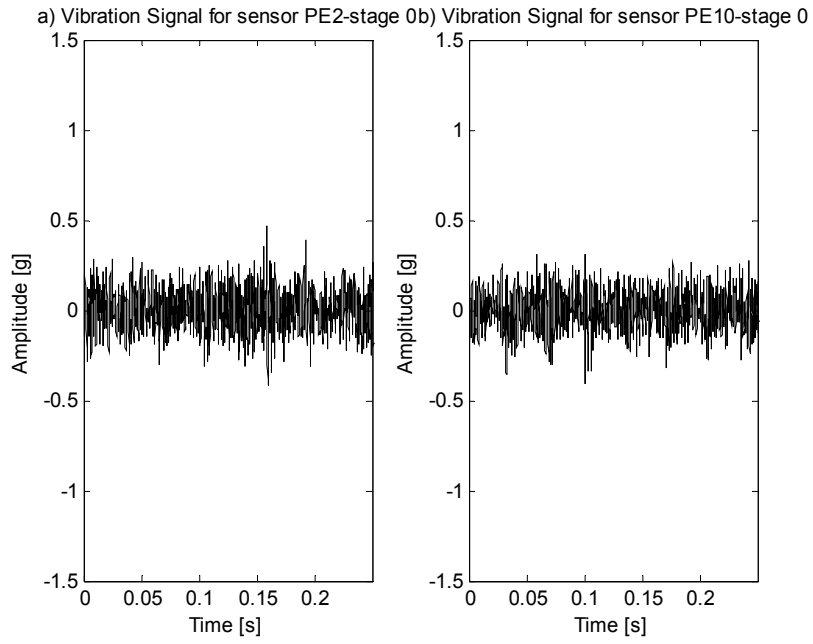


Figure 4.5 : Vibrations of both sensors PE2 and PE10 stage 0.

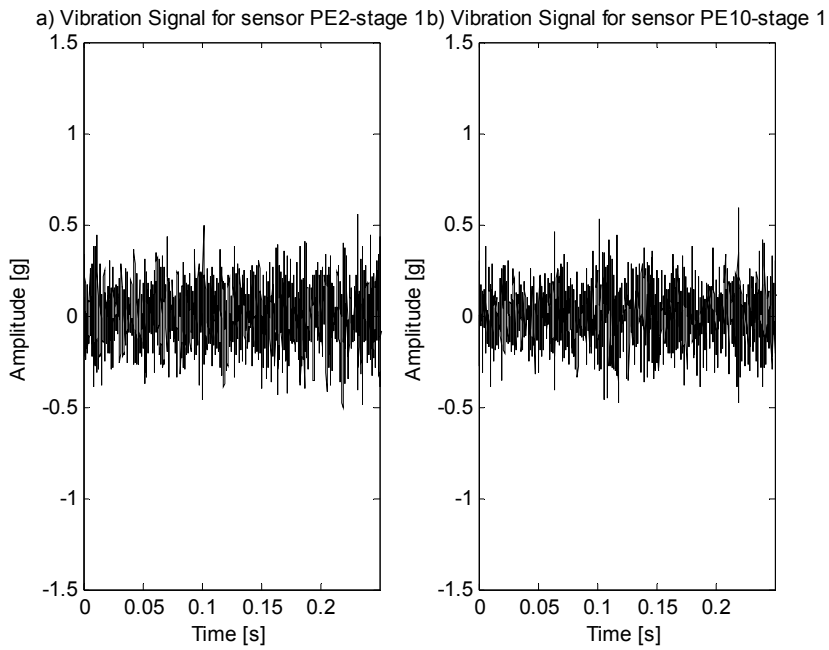


Figure 4.6 : Vibrations of both sensors PE2 and PE10 stage 1.

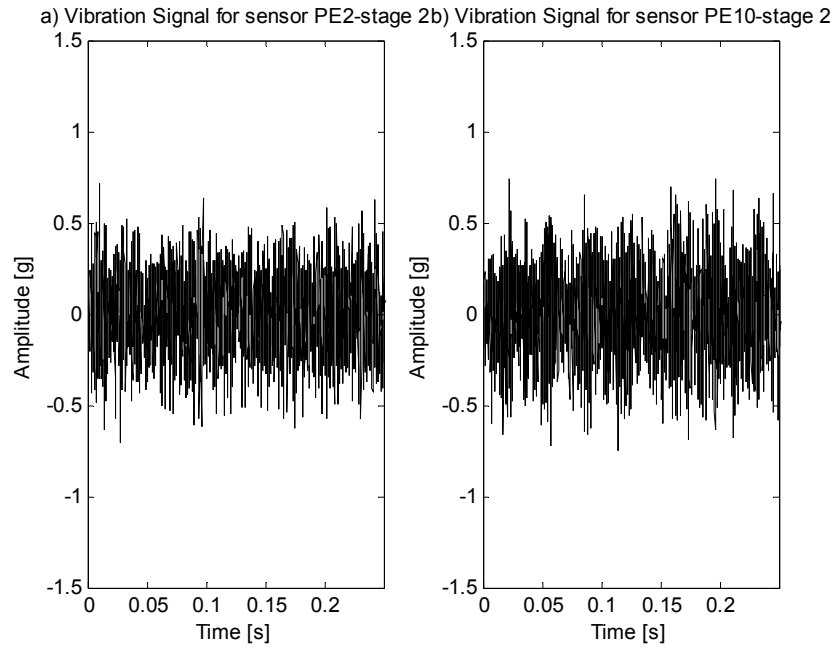


Figure 4.7 : Vibrations of both sensors PE2 and PE10 stage 2.

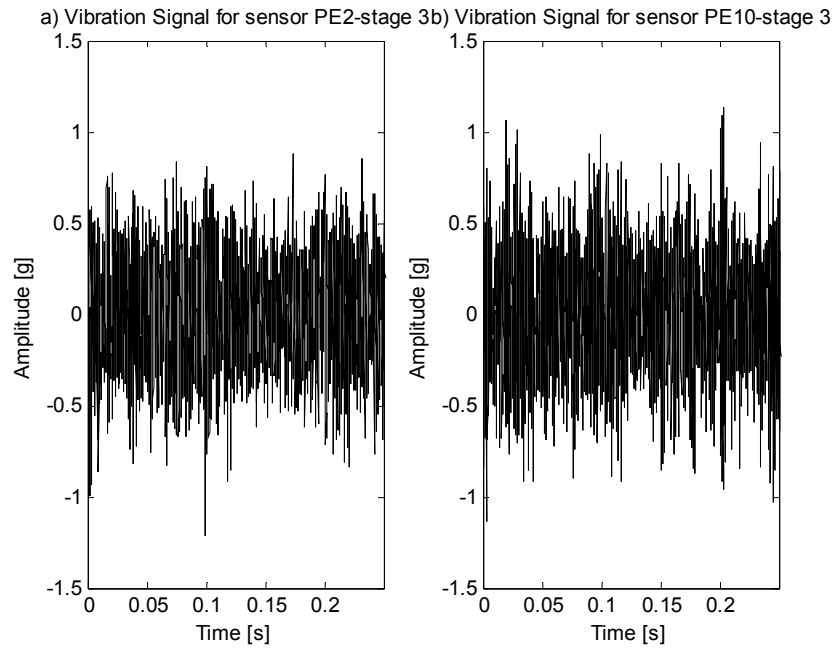


Figure 4.8 : Vibrations of both sensors PE2 and PE10 stage 3.

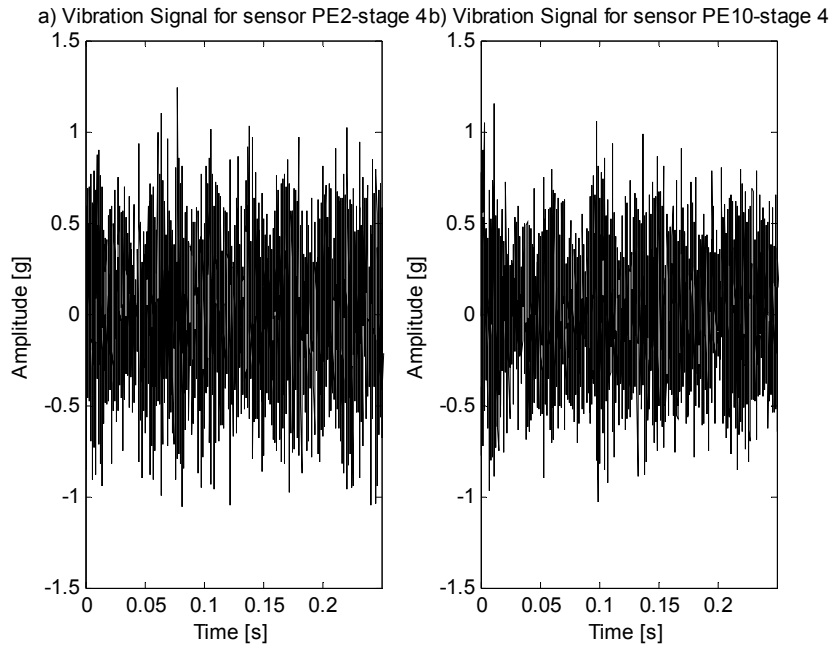


Figure 4.9 : Vibrations of both sensors PE2 and PE10 stage 4.

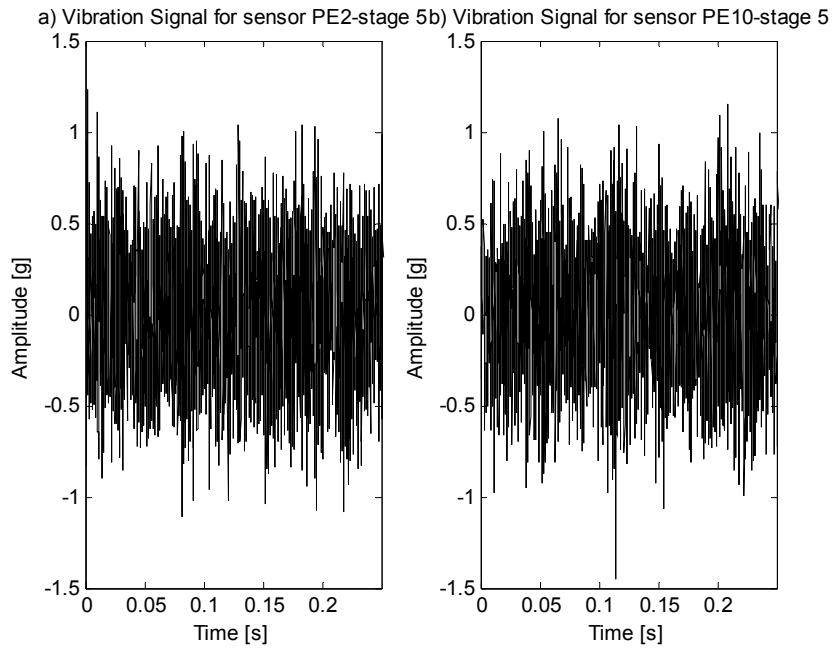


Figure 4.10 : Vibrations of both sensors PE2 and PE10 stage 5.

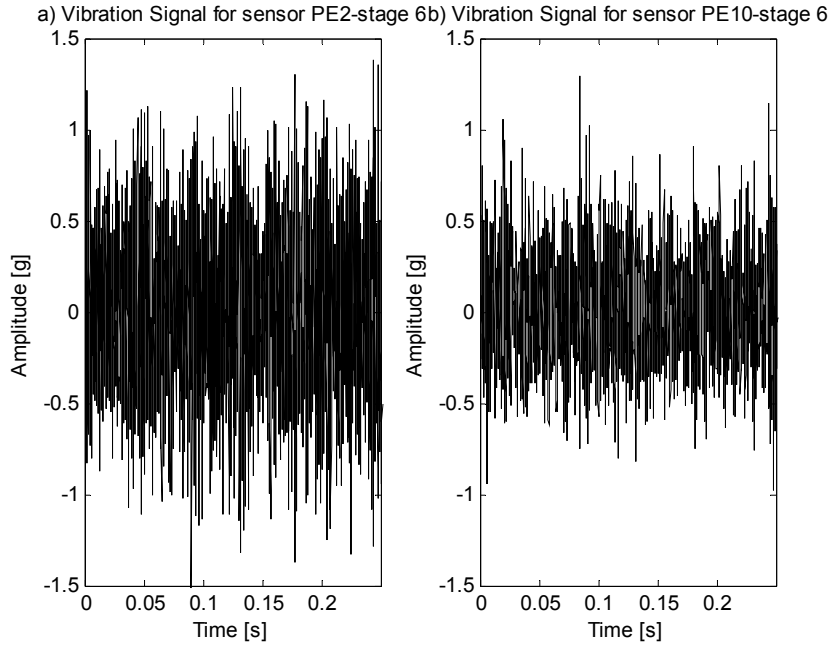


Figure 4.11 : Vibrations of both sensors PE2 and PE10 stage 6.

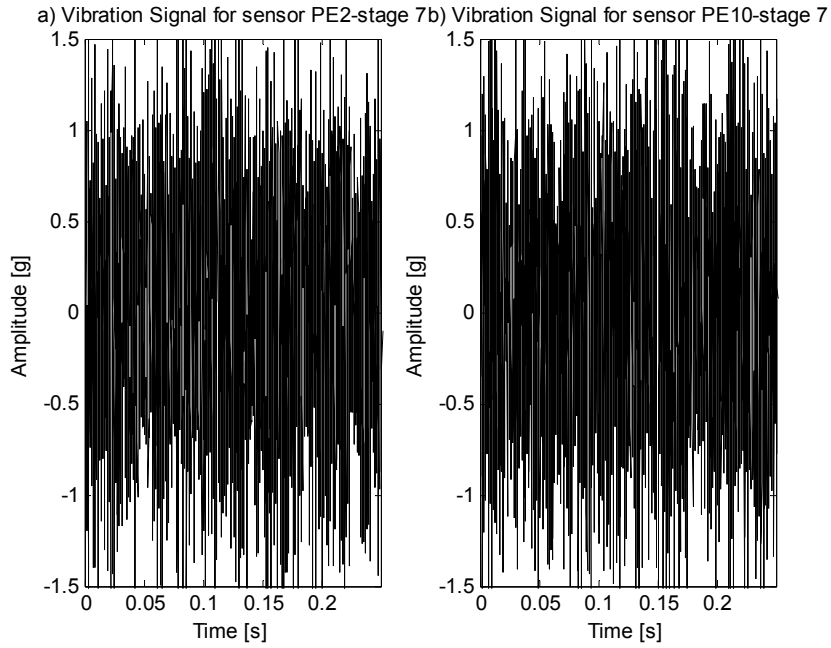


Figure 4.12 : Vibrations of both sensors PE2 and PE10 stage 7.

In these figures it can clearly be seen that the vibration amplitudes are getting bigger by each aging cycle. The spectral densities of these signals will provide to the observer the frequency interval related with this amplitude growth. In the following figures (4.13 to 4.20), the spectral densities of the sensors PE10 and PE2 are given for each aging cycle.

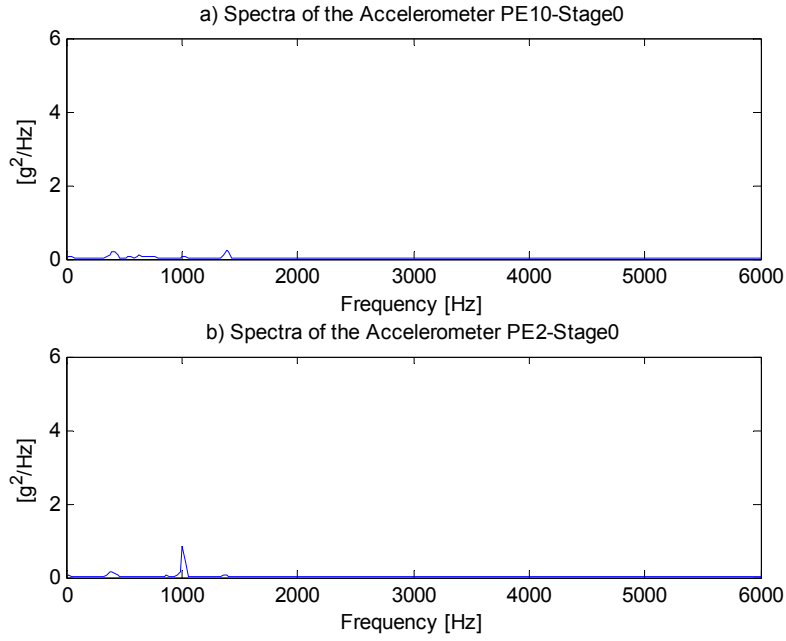


Figure 4.13 : Spectra of the vibration sensors PE2 and PE10, stage 0.

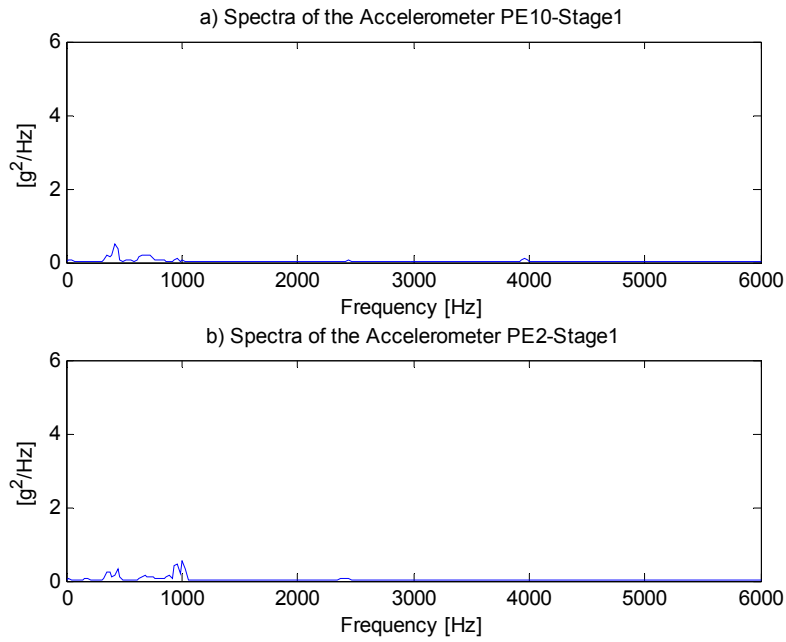


Figure 4.14 : Spectra of the vibration sensors PE2 and PE10, stage 1.

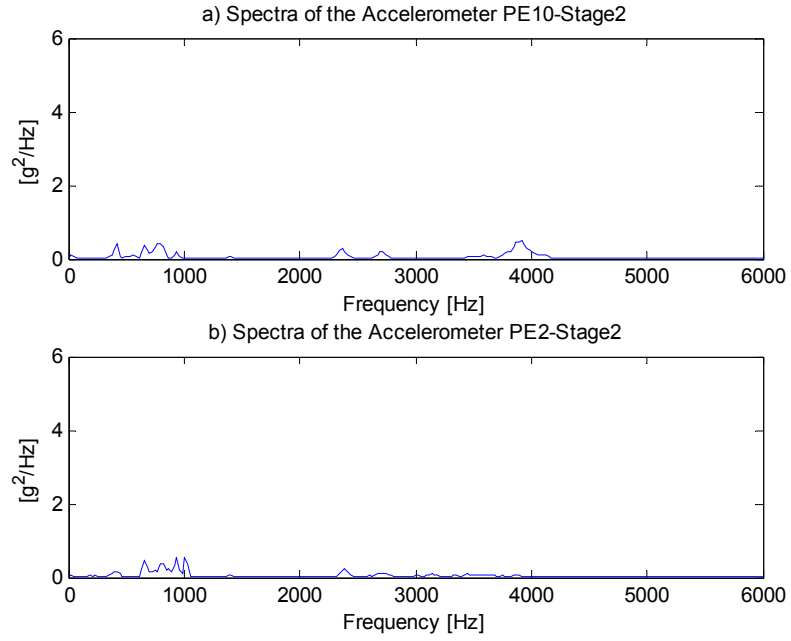


Figure 4.15 : Spectra of the vibration sensors PE2 and PE10, stage 2.

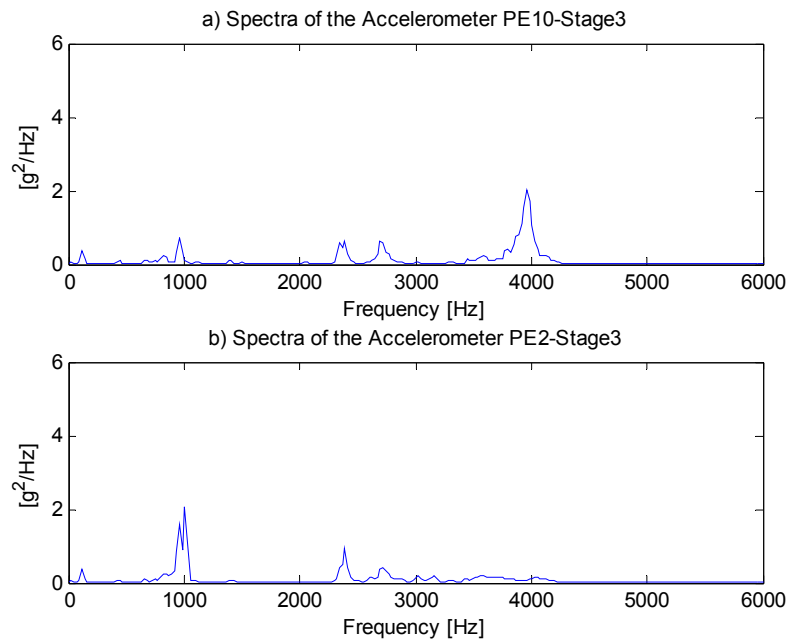


Figure 4.16 : Spectra of the vibration sensors PE2 and PE10, stage 3.

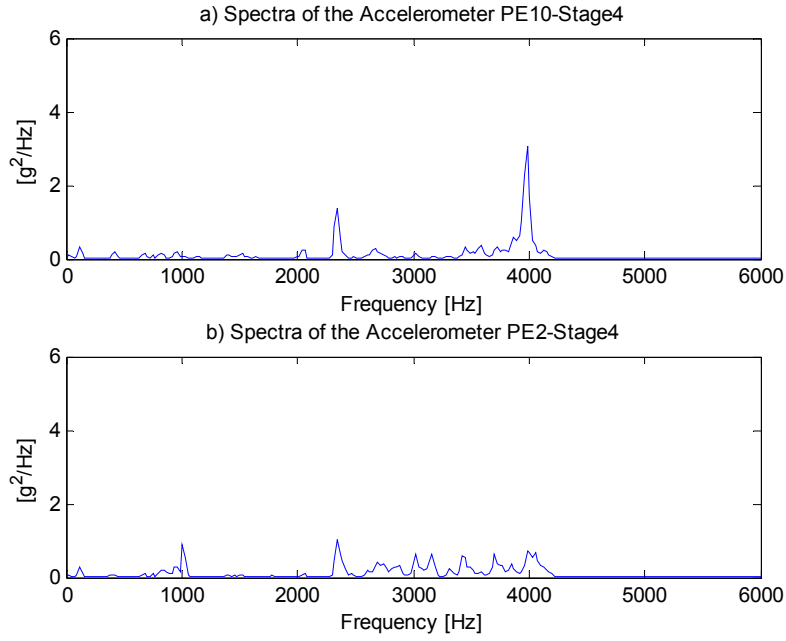


Figure 4.17 : Spectra of the vibration sensors PE2 and PE10, stage 4.

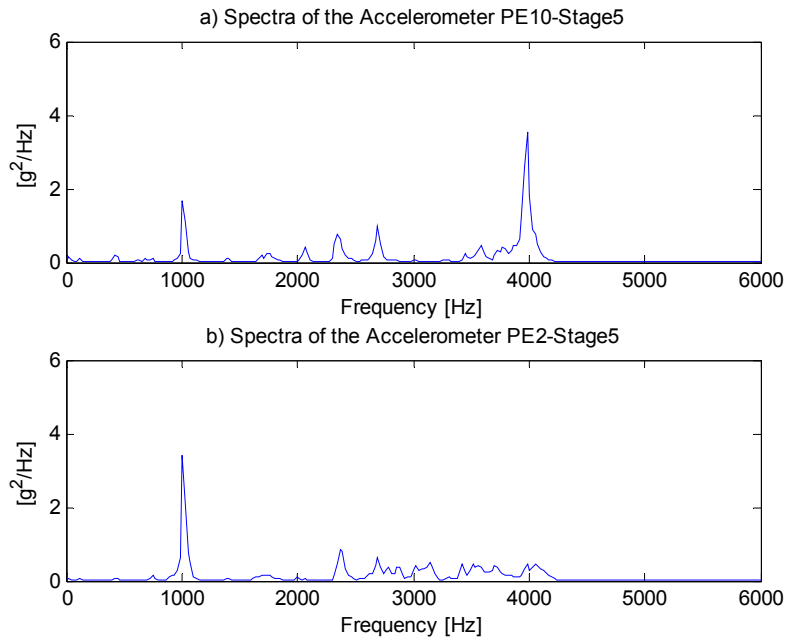


Figure 4.18 : Spectra of the vibration sensors PE2 and PE10, stage 5.

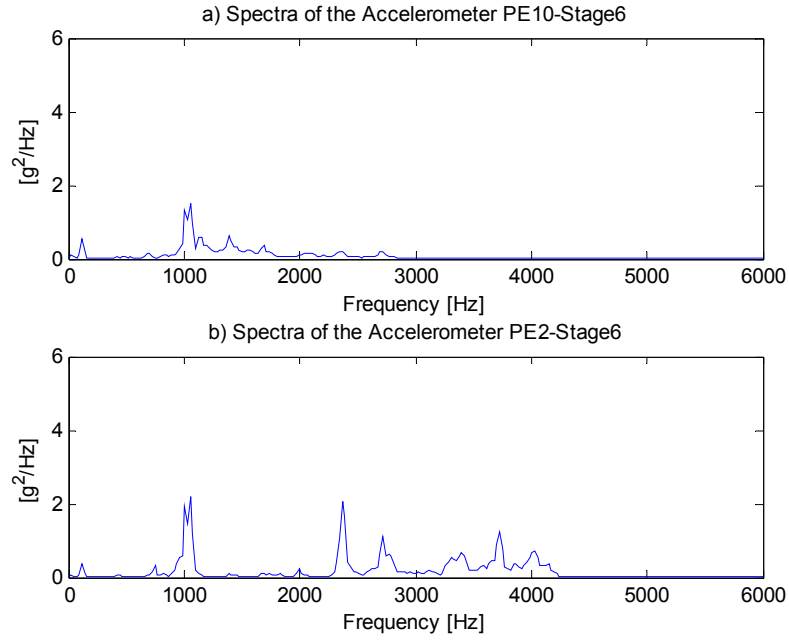


Figure 4.19 : Spectra of the vibration sensors PE2 and PE10, stage 6.

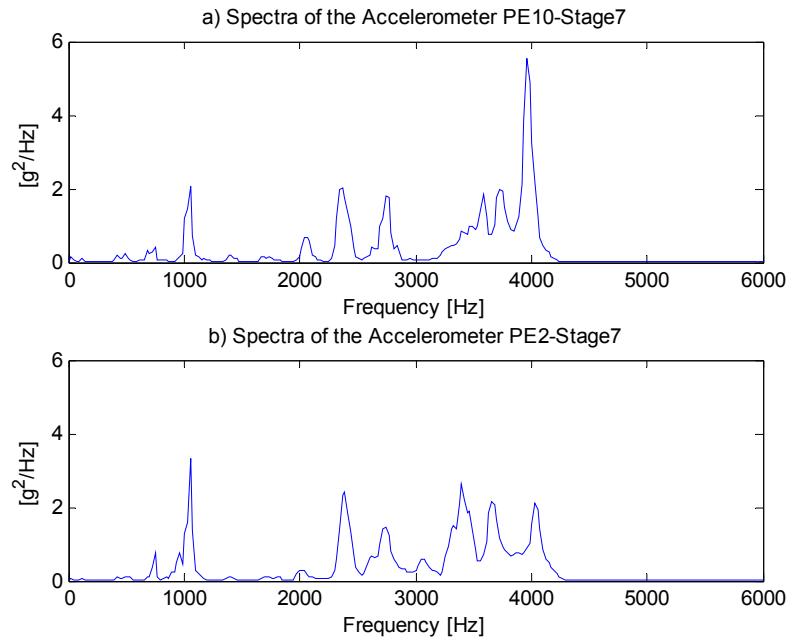


Figure 4.20 : Spectra of the vibration sensors PE2 and PE10, stage 7.

In the next section, sensor validation is done in order to make sure that the data obtained from the sensors are healthy and free from any faults.

4.3 Sensor Validation via Neural Network

In this section, sensor validation is performed by using neural networks. This process will make sure that the data used in the thesis and the related study does not include any sensor errors.

As mentioned in the previous chapter, the vibration sensors used in this study can be seen in Figure 4.1 (c) indicated with numbers 9 and 10. These correspond to the 16th and 15th channel data in the data acquisition system.

In the following figures, power spectral densities of these two sensors (9 and 10) can be seen for both healthy and faulty cases. In the figures, the difference between the faulty and healthy spectra can be seen by the amplitude differences.

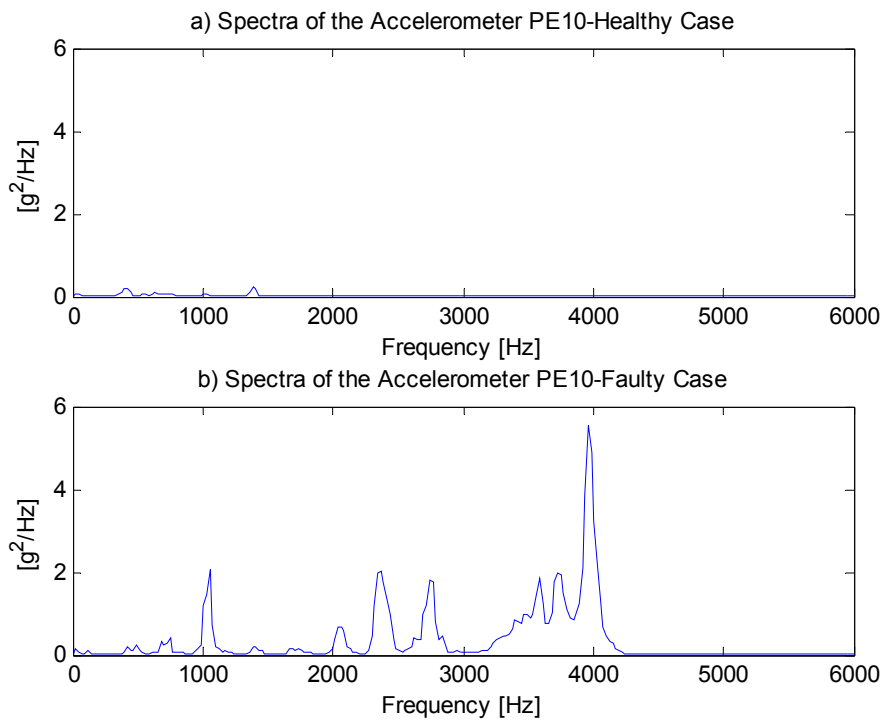


Figure 4.21 : Spectra of the motor vibration data from accelerometer #9.

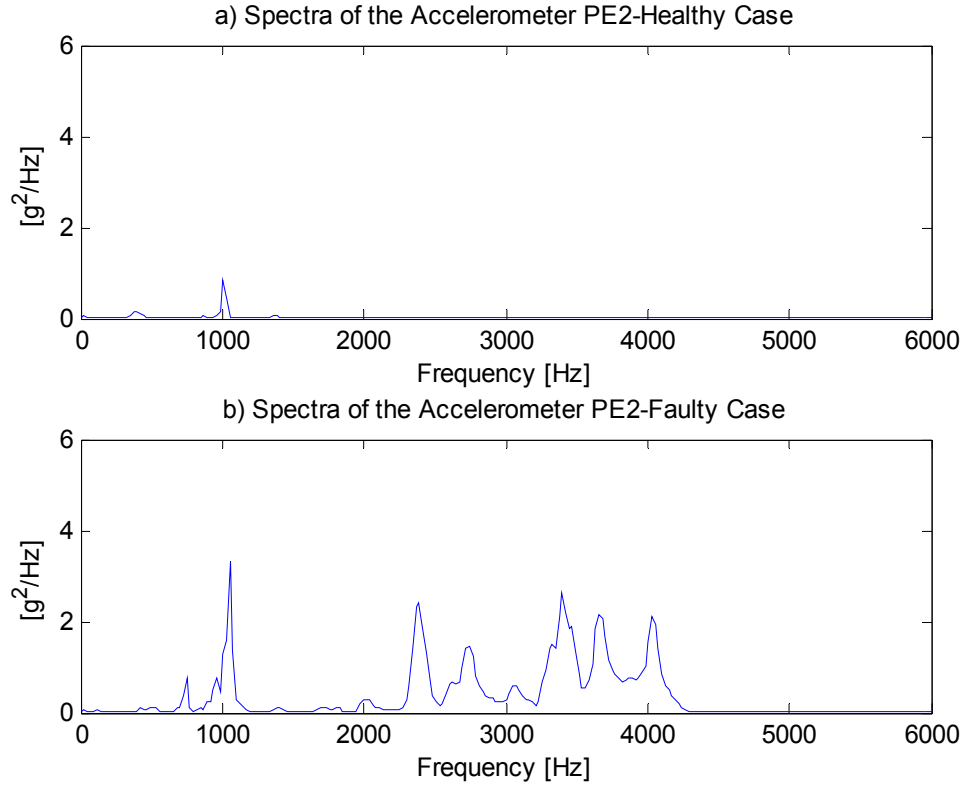


Figure 4.22 : Spectra of the motor vibration data from accelerometer #10.

A neural network application is proposed to find out a sensor failure in data acquisition system. Since two sensors PE10 and PE2 are identical, following neural network structures can be considered as sensor validation model. ($X=PE10$, $Y=PE2$). The neural networks are successful in determining the sensor failure in the accelerometers.

In the following two networks structures (Figure 4.23 and Figure 4.24), resilient backpropagation algorithm is used with two hidden layers which have 5 and 12 neurons respectively. The number of neurons are set by trial and error. Resilient backpropagation algorithm is used because of its speed.

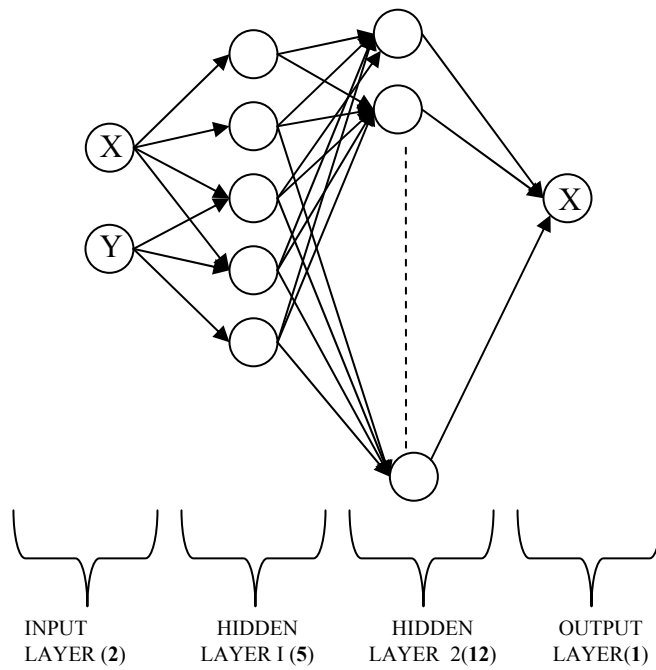


Figure 4.23 : Neural network #1.

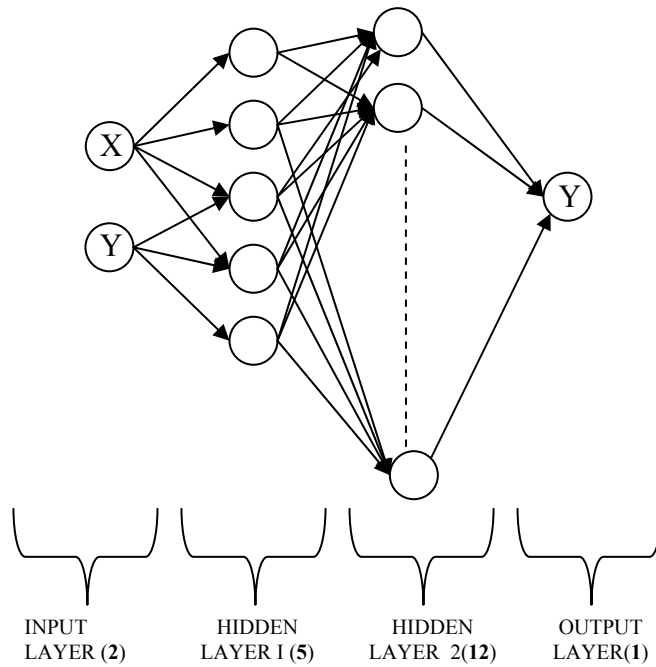


Figure 4.24 : Neural network #2.

NN1 Neural Network is trained with vibration data from accelerometers PE10 and PE2. The target values are considered as PE10. NN2 Neural Network is trained with vibration data from accelerometers PE10 and PE2. The target values are considered as PE2.

As mentioned previously, the data length from the accelerometers is 120000 samples (10 sec) with 12 kHz sampling frequency. 100000 samples are used as training (70%) and validation (%30). The rest of the 20000 samples are used for testing.

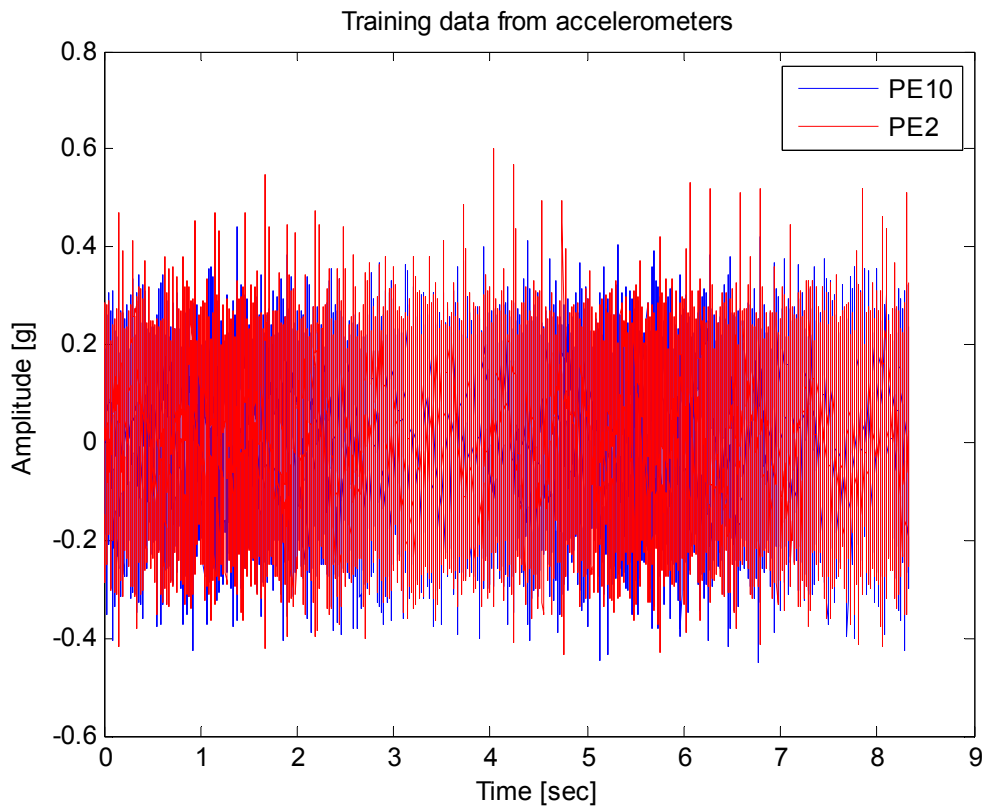


Figure 4.25 : PE10 and PE2 training data.

For simulating a sensor failure, white noise is added to the test data of the accelerometer PE10 (between 0,835 and 1,667 sec). The noise can not easily be seen in time domain as in Figure 4.26. However, it is very clear in the power spectral densities in Figure 4.27.

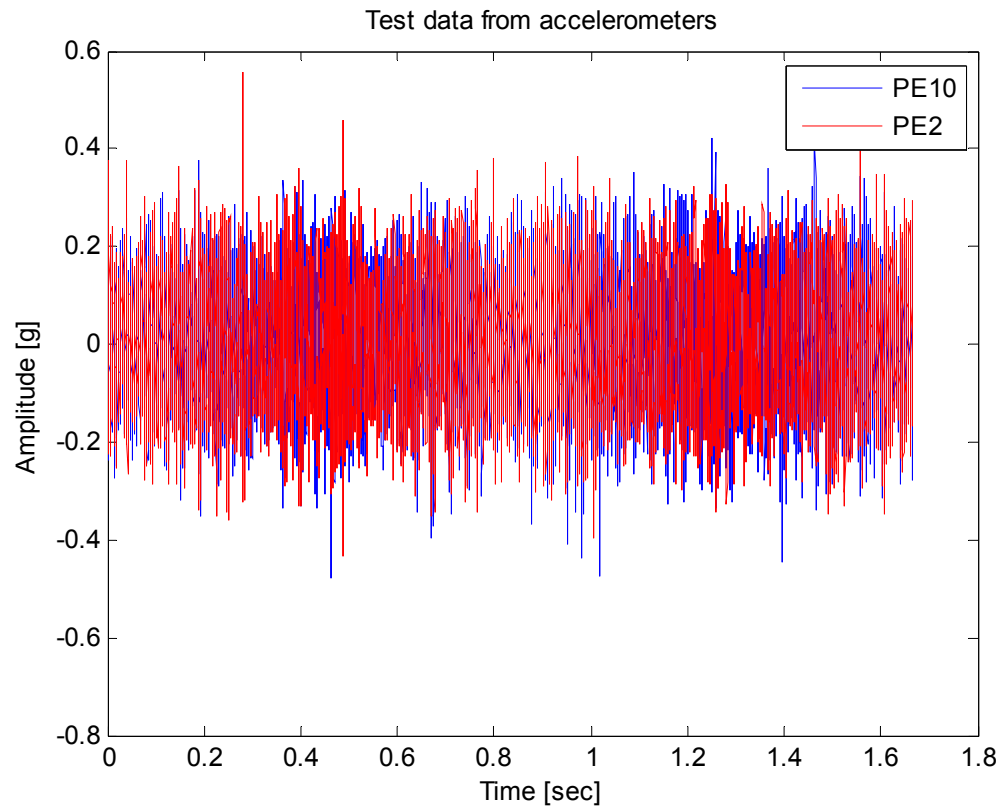


Figure 4.26 : PE10 and PE2 test data.

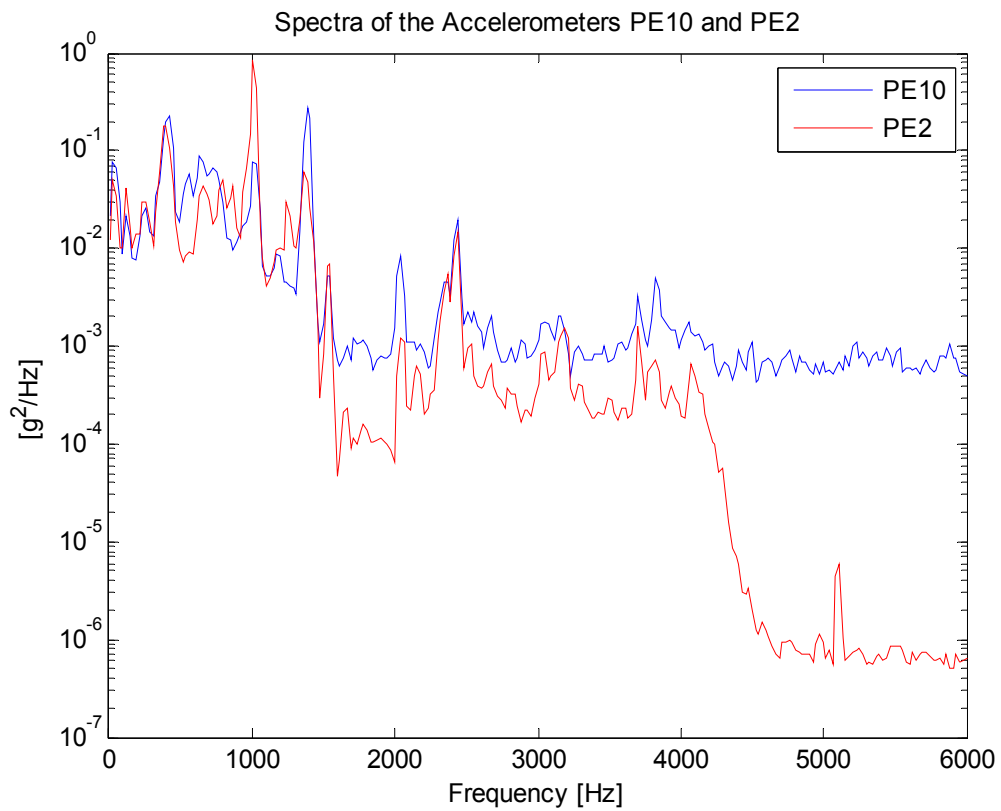


Figure 4.27 : PSD of accelerometers PE10 and PE2 in log scale.

Two networks are simulated with the test data. Test results are shown in Figures 4.28 and 4.29. with the error between targets and network outputs. It is clearly seen that NN1 output error gives a sensor failure in PE10.

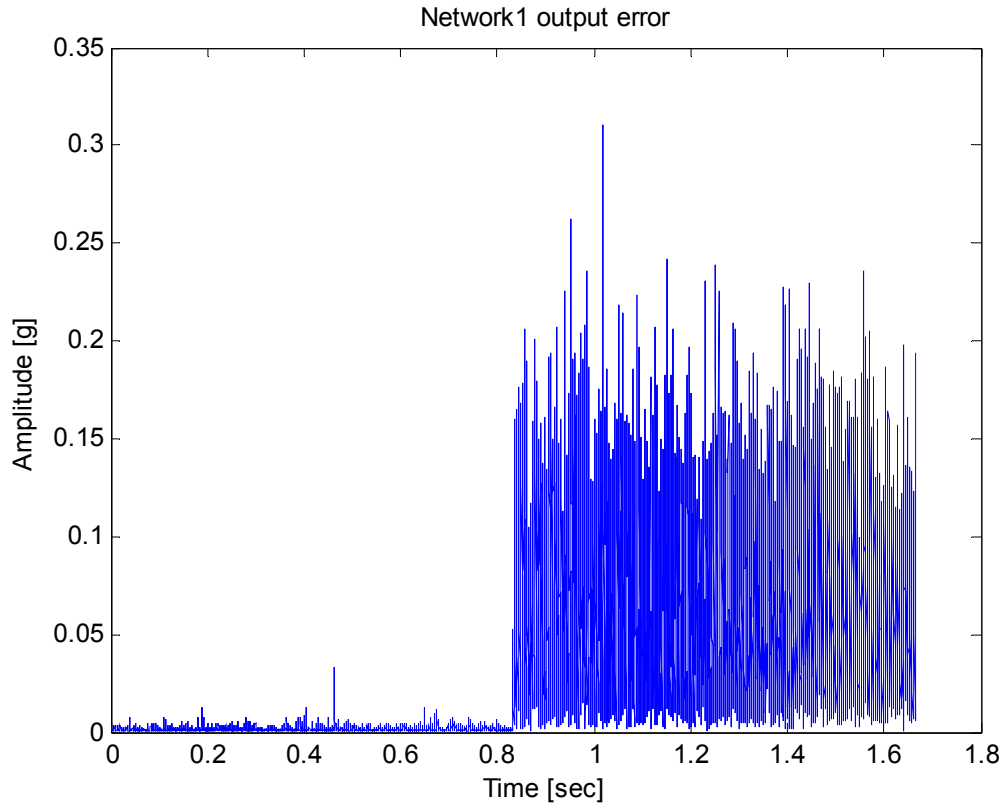


Figure 4.28 : NN1 output error.

These two sensors are identical, for the healthy sensor cases, the error level is around 0 value, however while simulating sensor fault case in PE10, as indicated in the scenario, the big amplitude of the error levels in the output of NN1, shows us the fault in the sensor PE10. Since the information that comes from these two identical sensors is the same, the comparison of the output of these two neural networks show that the data obtained from these two sensors are reliable.

As a result, it is clear that, one of the two identical sensors can be chosen as the data source for the study. For the main application, the PE10 sensor is used.

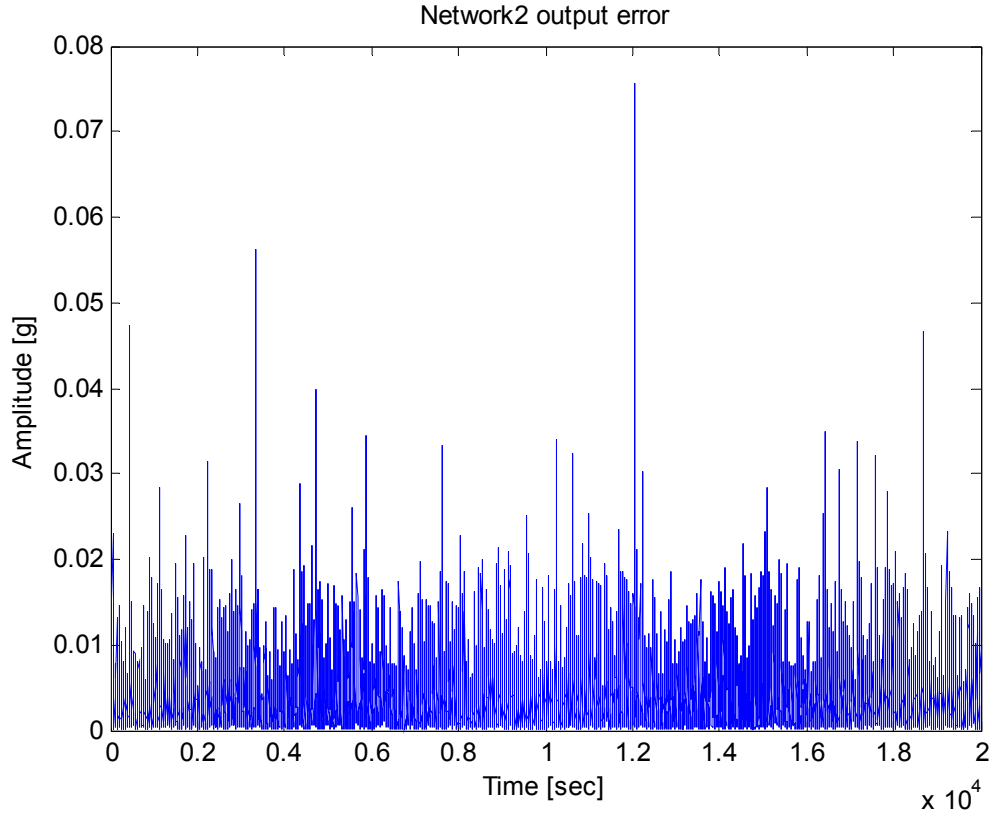


Figure 4.29 : NN2 output error.

4.4 General Characterization Of Bearing Damage

In section 4.3, power spectral densities of the vibration sensors throughout the aging process show that there is a trend of having higher amplitudes in 2-4kHz. This is an apriori information which we do not have in the data for the healthy state of the electric motor. This information is directly linked with the bearing damage characterization in literature [42].

5. SIMULATION FOR THE PROPOSED DATA MINING ALGORITHM AND ITS MATHEMATICAL INTERPRETATION

In this section, the algorithm which is proposed in this thesis is applied to an artificially created signal, which has similar characteristics to the vibration data obtained from the test setup explained in the previous chapter. Furthermore, the mathematical interpretation is given in order to justify the proposed algorithm.

5.1 Simulation Data

In this section, the simulation data is created and the proposed algorithm is being tested on simulation data. The simulation data is similar to the vibration data which is obtained from the healthy state of the electric motor. For obtaining a similar pattern, the simulation data contains the rotation frequency, some major sinusoidal harmonics (1st, 10th, 30th harmonics) and noise in addition to these major components. Figure 5.1 represents this simulation data. The vibration data obtained from the measurement system can be seen in Figure 5.2.

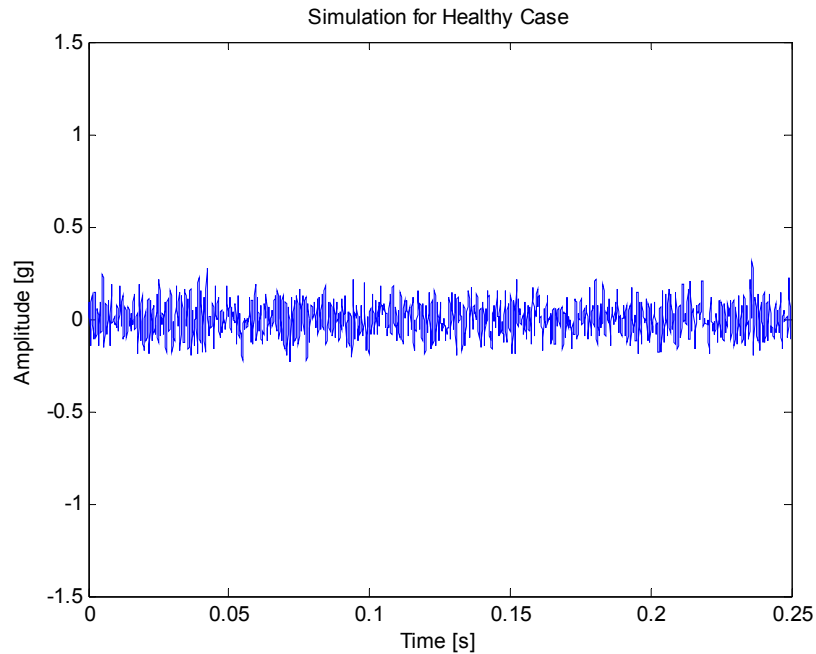


Figure 5.1 : Vibration data simulation.

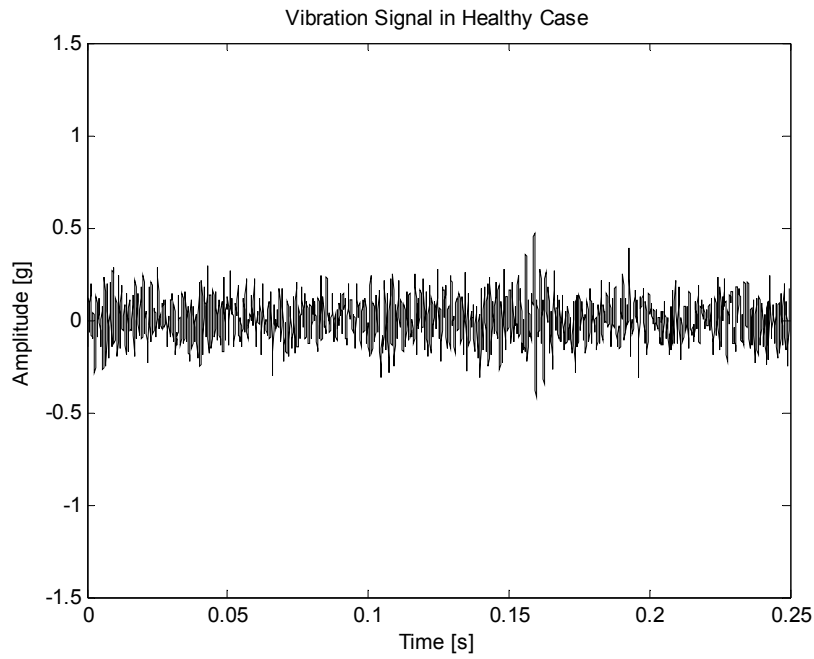


Figure 5.2 : Vibration signal in healthy case (measurement data).

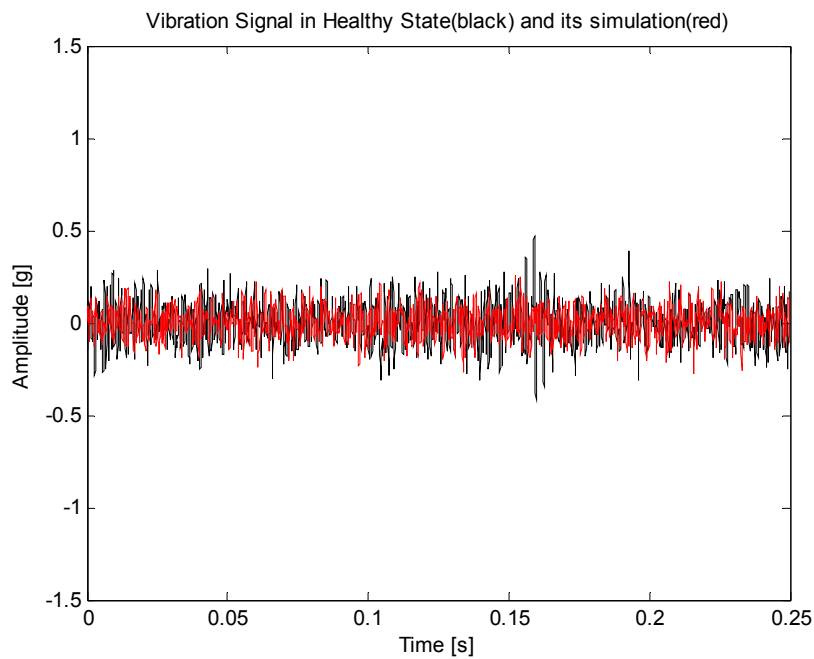


Figure 5.3 : Simulation data and measured vibration signal.

Both simulation and the measured signal can be seen in Figure 5.3.

If power spectral density is applied to this signal, the result can be seen in Figure 5.4.

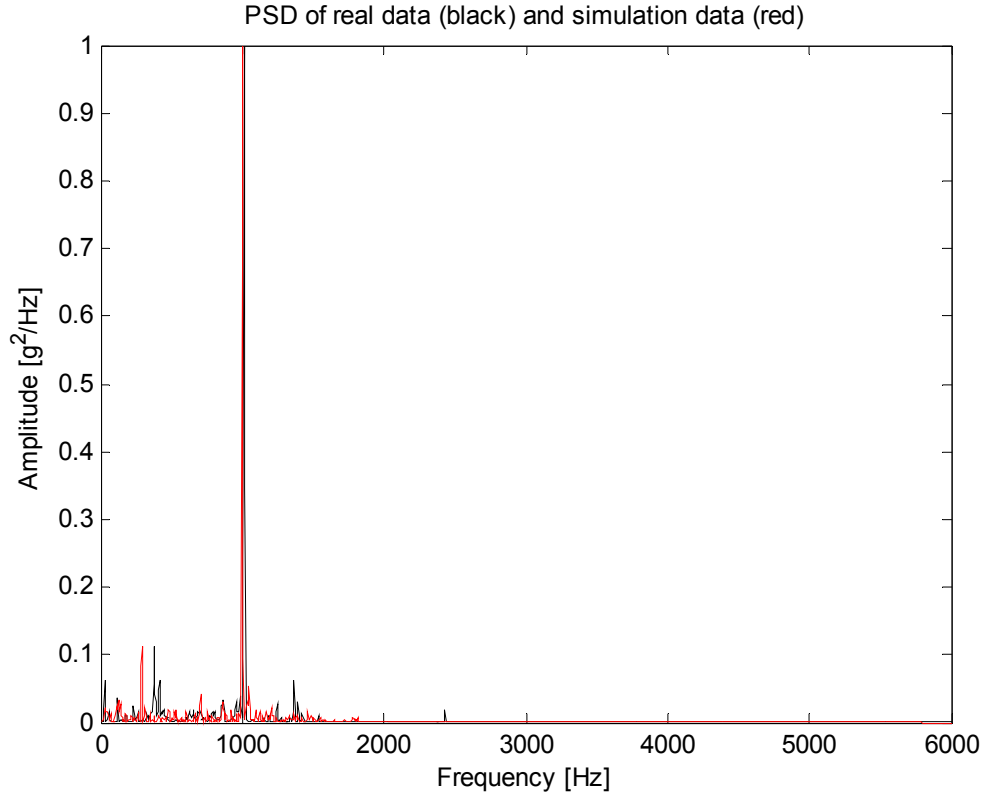


Figure 5.4 : PSD of the real vibration (black) and simulation data (red).

If the statistical values of both the simulation and the real data are compared, it can easily be seen that the simulation is very similar to the real vibration signal. This comparison is done in Table 5.1.

Table 5.1 : Statistical comparison of the signals.

	Mean	St. Dev	Kurtosis	Skewness
Real Data	0.0016	0.1135	29266	0.0591
Simulation	0.00228	0.0836	26995	0.0475

From Figure 5.4 it can easily be told that the power spectral density can not give information about the high-frequency components for the simulation data. However the power spectrum density of the CWT at the first scale extracts the information in the high frequency band (2-4 kHz).

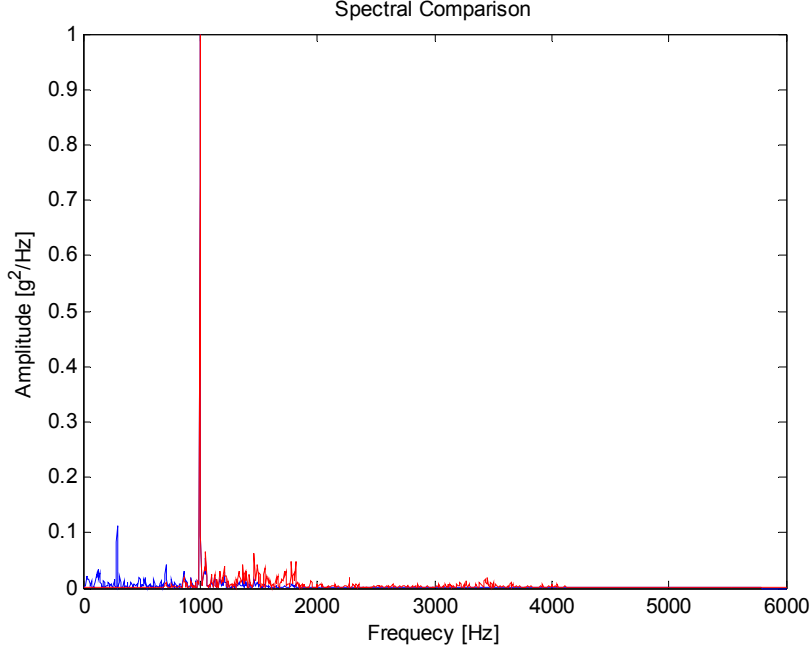


Figure 5.5 : Spectral comparisons: PSD (blue), PSD of CWT (red).

This information corresponds to the potential bearing defects that the algorithm is supposed to extract. The comparison between the PSD (blue) and the PSD of the CWT at first scale (red) can be seen in Figure 5.5.

5.2 Mathematical Interpretation of the CWT as a Correlation

There are two different aspects of this simulation. These are:

- CWT
- CWT as non-orthogonal transform, respectively

For an ordinary CWT approach, mother wavelet function to be used is general, it can be orthogonal or non-orthogonal.

Recalling the formulas 2.9 and 2.10 in Section 2, CWT can be given in the form of an inner product between signal $f(t)$ and the mother wavelet $\psi_{ab}(t)$ as stated below:

$$CWT(a, b) \equiv \langle f(t), \psi_{ab}(t) \rangle \quad (5.1)$$

From Equation 2.9, for $b=0$, the equation can be written as:

$$\psi_{a,0} = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t}{a}\right) \quad (5.2)$$

Hence, Equation 5.1 can be re-written as:

$$CWT(a, b) \equiv \langle f(t), \psi_{ab}(t) \rangle = \langle f(t), \psi_{a,0}(t-b) \rangle \quad (5.3)$$

Using the integral form of the inner product, Equation 5.3. is defined as a cross-correlation function

$$CWT(a, b) = \int_{-\infty}^{\infty} f(t) \psi_{a,0}(t-b) dt \equiv R_{f\psi_{a,0}}(b) \quad (5.4)$$

where $R_{f\psi_{a,0}}(b)$ is a cross-correlation function between the signal $f(t)$ and mother wavelet $\psi_{a,0}$.

From this view, considering the Cauchy-Schwarz inequality, it is re-written as:

$$|CWT(a, b)| = |\langle f(t), \psi_{ab}(t) \rangle| \leq \|f(t)\| \|\psi_{ab}(t)\| \quad (5.5)$$

The square of the Equation 5.5 will be:

$$|CWT(a, b)|^2 \leq \|f(t)\|^2 \cdot \|\psi_{ab}(t)\|^2 \quad (5.6)$$

In Equations 5.5 and 5.6, we can assume the following:

$$\psi_{ab}(t) = \alpha f(t) \quad (5.7)$$

Here, α is a scalar.

The similarity between the signal $f(t)$ and the mother wavelet $\psi_{ab}(t)$ is:

$$\|f(t) - \psi_{ab}(t)\|^2 = \|f(t)\|^2 + \|\psi_{ab}(t)\|^2 - 2\langle f(t), \psi_{ab}(t) \rangle \quad (5.8)$$

or

$$\|f(t) - \psi_{ab}(t)\|^2 = \|f(t)\|^2 + \|\psi_{ab}(t)\|^2 - 2 \operatorname{Re}[CWT(a, b)] \quad (5.9)$$

Here, Re stands for the real part of the complex valued transform CWT.

If the signal $f(t)$ is very close to the mother wavelet $\psi_{ab}(t)$ under the assumption of scalar α , then, the maximum value of $\text{Re}[CWT(a,b)]$ points show strong similarity degree between the signal $f(t)$ and the mother wavelet $\psi_{ab}(t)$. In this manner, Equation 5.9 is provided for each (a,b) pair.

The maximum value of the $\text{Re}[CWT(a,b)]$ can be defined by the maximum value of the cross correlation function $R_{f\psi_{a,0}}(b)$ [97].

$$\max\{\text{Re}[CWT(a,b)]\} = \max\{R_{f\psi_{a,0}}(b)\} \quad (5.10)$$

In this thesis, simulation study is done for the first scale, $a=1$, of the CWT. As seen in Figure 5.6. first scale covers the whole frequency band of the signal.

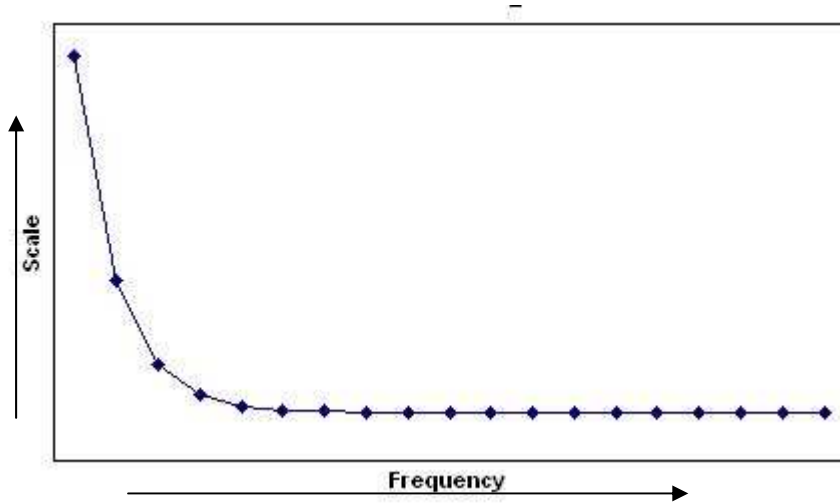


Figure 5.6 : Scale-frequency band relationship.

Hence Equation 5.10 becomes:

$$\max\{\text{Re}[CWT(1,b)]\} = \max\{R_{f\psi_{1,0}}(b)\} \quad (5.11)$$

Where

$$\psi_{10} = \psi_{10}(t-b) \quad (5.12)$$

Wavelet analysis is based on the correlation between the analyzed signal and the function which is analyzing it. Since the first scale contains the widest frequency band, the correlation relationship contains the widest band. In this sense, by the help of the proposed method in the thesis, the weak components in the high frequency band can be seen in the PSD change based on this correlation relationship.

6. SIGNAL BASED ALTERNATIVE APPROACH FOR FAULT DETECTION IN DATA MINING

In this thesis, the data mining techniques are used in order to identify the potential defects in the induction motors. In this section, a new approach is defined for the data mining concept for this purpose. The study is consisted of two main parts:

- Data pre-processing
- Neural network application

The method is first applied on the healthy state vibration data of the electric motor. Then, it is applied on two different sets of simulation data. The first simulation data is similar to the healthy state vibration data reflecting its spectral and statistical characteristics, on the other hand, the second simulation data has no similarities with the healthy state vibration data.

6.1 Data Pre-Processing for Application

The healthy state vibration or the simulation signal is considered as the raw data. According to the data mining methodology, the raw data needs to be processed in order to be used by the data mining techniques. In order to accomplish this, a series of processes are applied. As in the second block of Figure 6.1, the first technique applied to the vibration signal is the CWT. By applying the CWT, a sub-signal representation of the signal is obtained. This sub-signal includes the useful information at the first scale and it is extracted as a feature. Hence, this second block is defined as the feature selection stage. Symlet4 is used as the mother wavelet during the CWT calculation.

First stage in the study is the data preparation. This stage is seen in Figure 6.1.

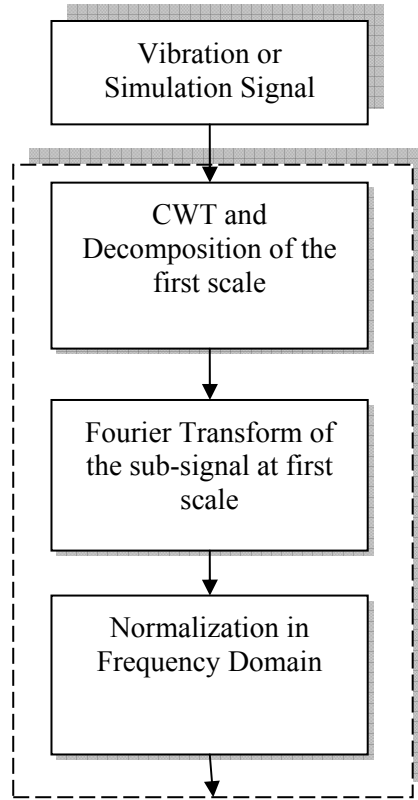


Figure 6.1 : Data pre-processing.

After the feature selection stage (application of the CWT) and selection of the representation in the 1st scale, data conversion is done two stages: the Fourier transform (Power Spectral Density – PSD) and normalization. This conversion helps us represent the data in the frequency domain instead of the time domain. The output of this conversion is seen in Figure 6.12. The abbreviation pcwtc stands for the PSD for the sub-band signal at first scale of the CWT. After this three-step data pre-processing stage, the transformed/converted data is applied to the AANN for training.

6.2 Neural Network Application

The second stage of this study consists of the ANN application. After the data pre-processing stage, a new representation of the signal is obtained. This new data is the frequency domain representation of the raw signal presented with the first scale of the CWT. This data is applied to an auto-associative neural network (AANN) for training. The reason for selecting an AANN is its capability of identity mapping which is described in section 2.2.4. Here, the AANN is used as a fault detector. It

finds out the important details of the signal through the CWT approach and shows big amplitudes at output nodes in terms of the error variation.

6.2.1 Neural network training and recall phase

In order to make the AANN functional, it has to be trained first. As mentioned in the previous section, the input data is the normalized PSD of the sub-band signal at first scale of the CWT. Since the PSD is computed at 512 points, data pattern is represented by 512 points. Due to the nature of the AANNs, the input and output layers consist of same number of nodes. As mentioned in Section 2.2.4, the hidden neurons provide a mapping layer for creating a mapping of the original input space. In this application, two different topologies of AANN are tested. Both topologies can be seen in Figure 6.2 and Figure 6.3. The difference of these two topologies is described in Section 2.2.4. In the application, the comparison is showing that the 5 layered topology is not suitable for our analysis since it faces difficulties during learning, i.e. not reaching the MSE target, having unexpected peaks in the learning curve. Figure 6.4 and Figure 6.5 show the mentioned problems. For this reason, 3 layered topology is preferred in the rest of the study.

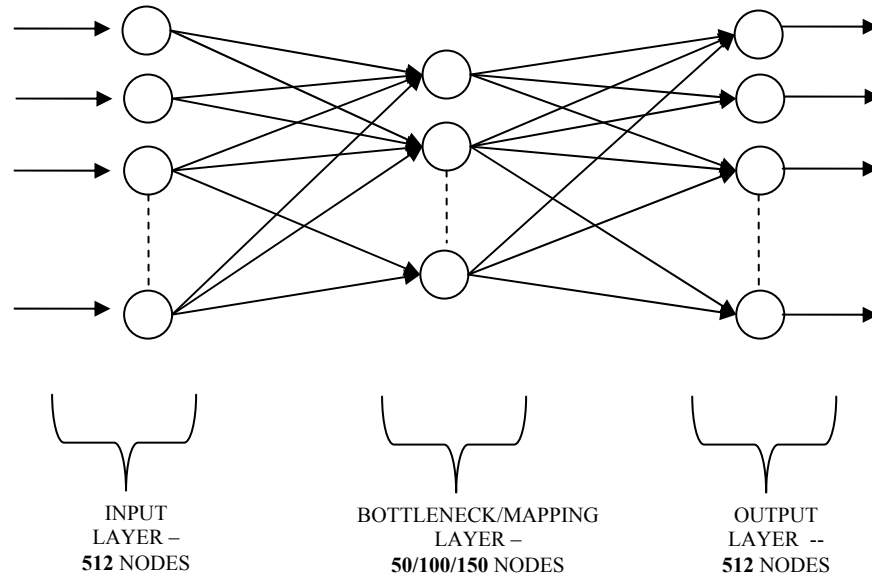


Figure 6.2 : 3-layer AANN topology.

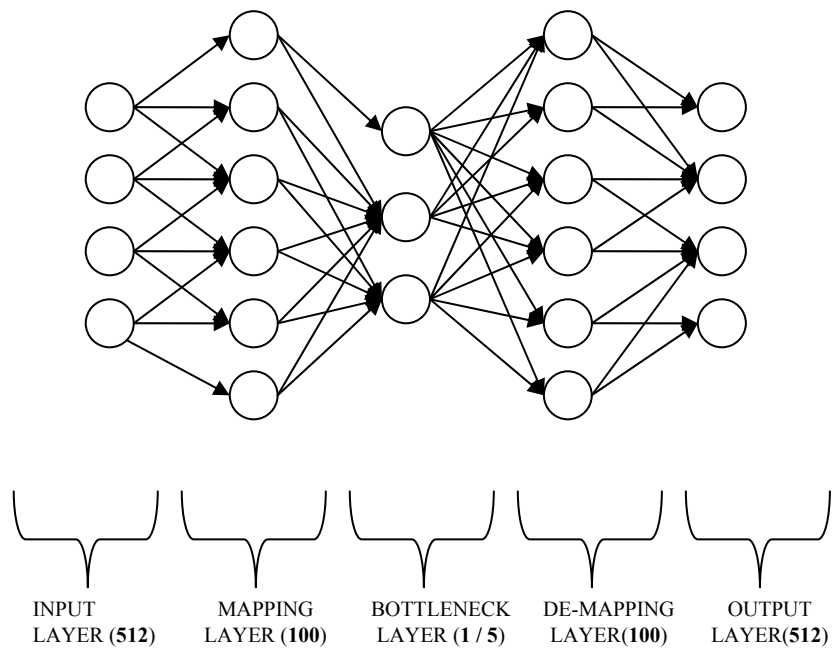


Figure 6.3 : 5-layer AANN topology.

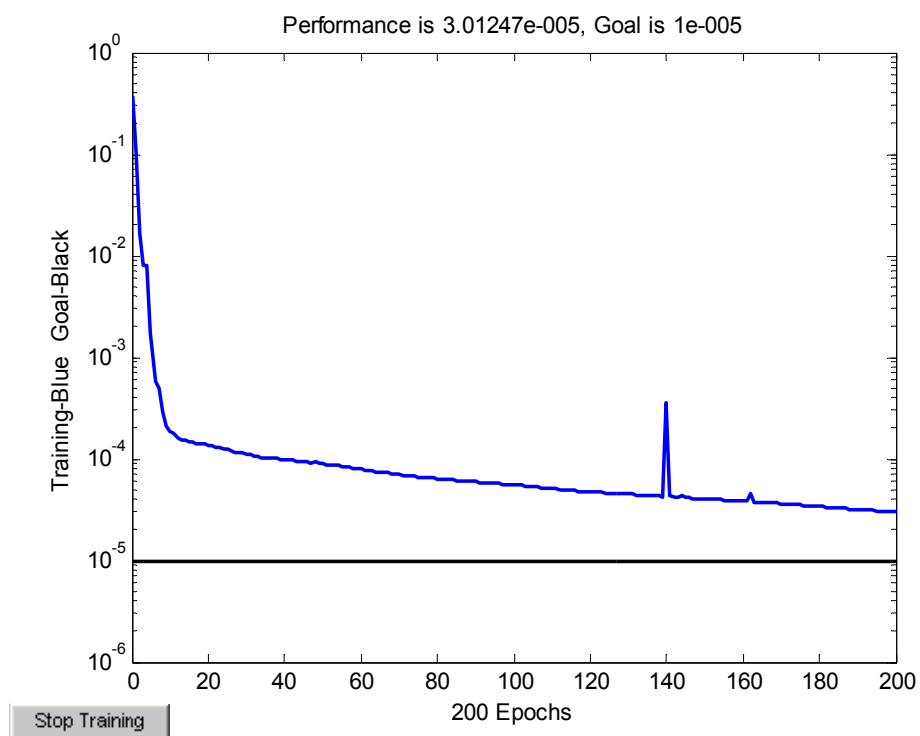


Figure 6.4 : 5-layer AANN convergence.

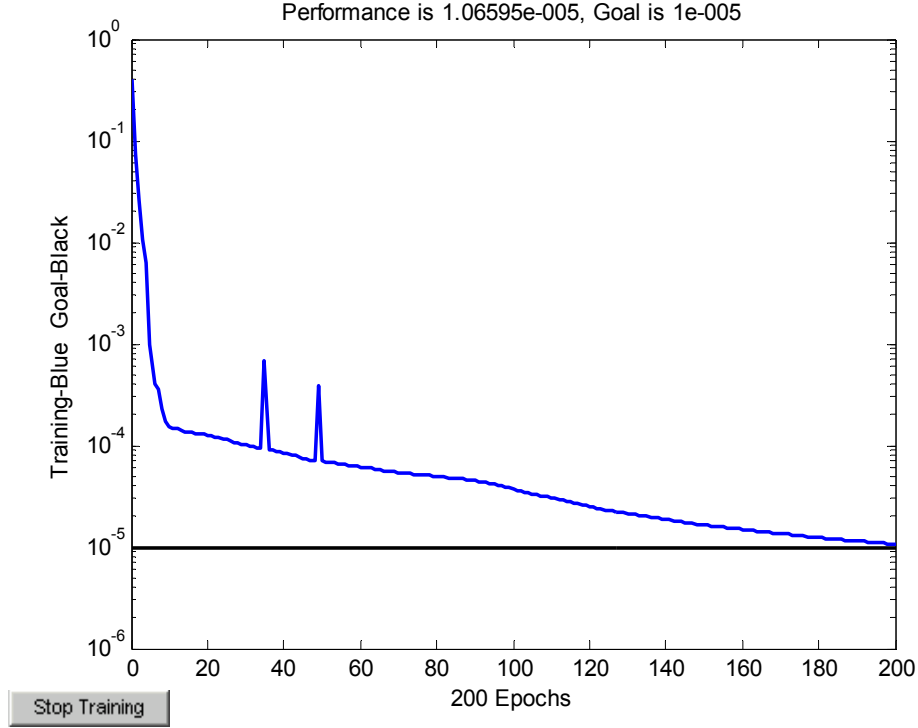


Figure 6.5 : 5-layer AANN learning.

For the learning stage of the neural network, both gradient descent and resilient backpropagation algorithms were used. The results were very similar to each other. However, due to the fact that the training time of gradient descent algorithm is extremely longer, resilient backpropagation algorithm was chosen. The difference of the learning duration for these two algorithms can be seen by comparing Figure 6.6 and Figure 6.7. While GDM reaches only an MSE of 0,27 (goal 0.1) with 10000 epochs, the RBP algorithm reaches the MSE performance goal of 10^{-5} in 50 epochs for the same training set. Sigmoid function is used as the activation function for both hidden and output nodes.

For the training phase, the data is divided into blocks which contain 3000 samples each (0.25 seconds). In total, 5 seconds of data (20 sample blocks) is available for analysis. The training phase of the study is shown in Figure 6.8. The training set is formed by the normalized PSD of the 1st scale representation of the CWT. For each raw data (healthy state vibration signal or one of the simulation data), the total 20 samples are divided into 4 different sets for training and recall: 4/16, 8/12, 12/8 and 16/4. These data sets are applied to three different types of AANN topology. The layer structures of these three topologies are: 512:50:512, 512:100:512 and 512:150:512, where the three numbers are indicating the number of nodes for input,

hidden (bottleneck) and output layers, respectively. In summary, for all three topologies, four different runs are performed, resulting in 12 runs for each signal. In each run, the training set is used to train the network. Then, the recall set is given to the network for calculating the error as the difference between the expected output and the actual output. For each run, the maximum values of the error are calculated and stored as a vector; meaning that after each run, a vector whose size is equal to the number of recall data set is created. In this study, the maximum value of this vector is identified as the threshold of each run and it is used in the test phase of the study.

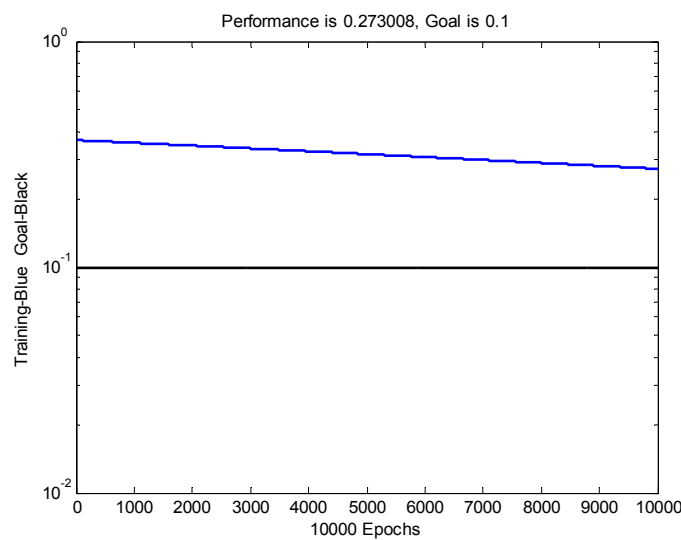


Figure 6.6 : Slow learning with GDM.

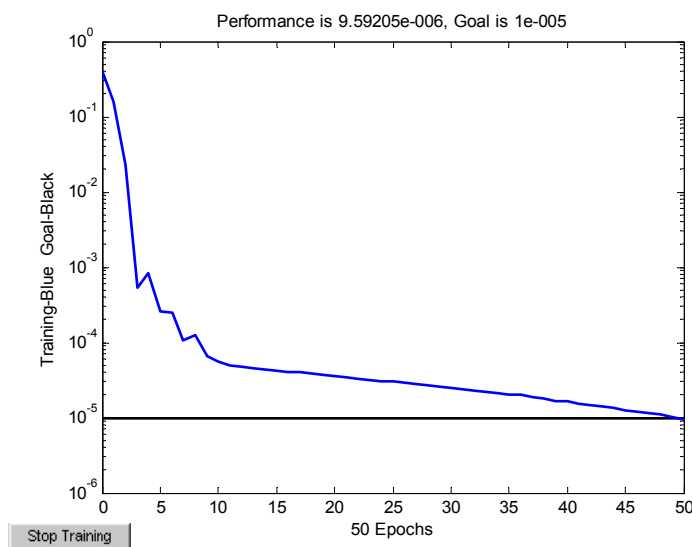


Figure 6.7 : Fast learning with RBP.

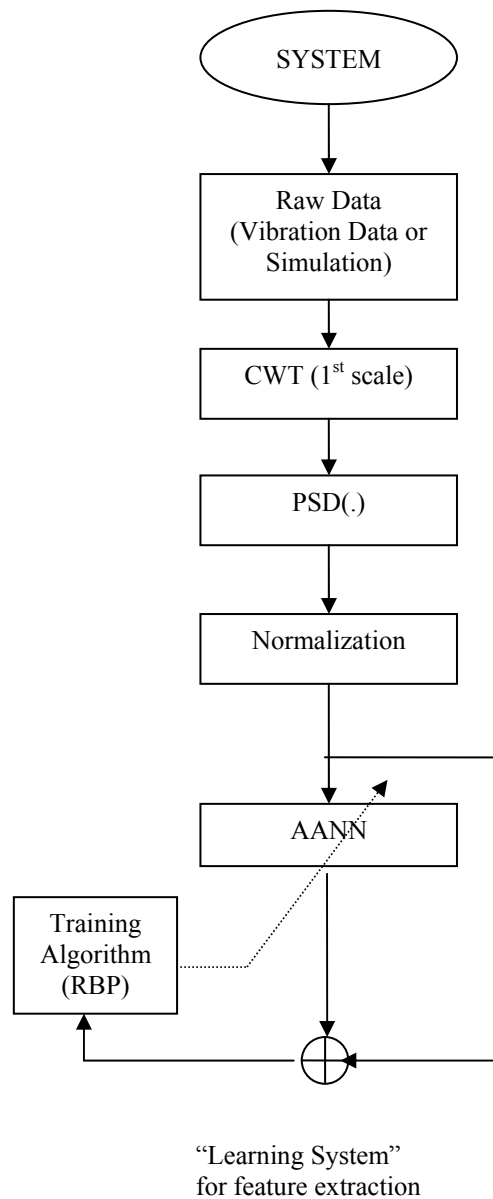


Figure 6.8 : Block diagram of the algorithm and its training.

6.2.2 Neural network test phase

As mentioned in the previous section, the AANN was trained by the PSD of the CWT of the signal. At the test phase, the AANN is tested by applying the normalized PSD of the signal only. The test phase is shown in Figure 6.9.

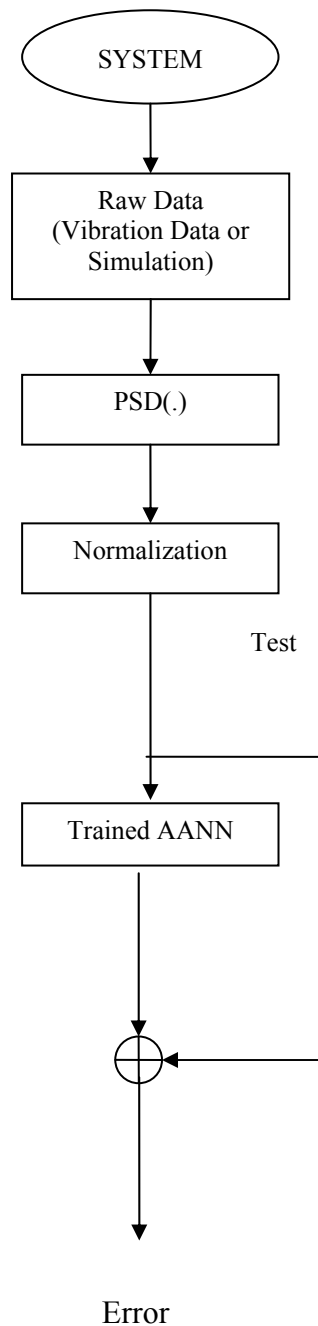


Figure 6.9 : Block diagram of the signal based data mining algorithm and its test.

In the test phase, PSD of all 20 data sets are used; and for each run, the error is calculated as the difference of the actual output of the neural network and the expected output. As mentioned previously, the expected output of the AANNs are identical to the input.

In the training stage of the study, a threshold was identified for each run. This threshold is used as a classifier during the testing so that if the maximum value of the error during the testing stage is bigger than the threshold, it is identified as novelty.

In the following sections, the method is being applied to three different raw data: the vibration signal of the electric motor in the healthy case, the first simulation data which has similarities to the vibration data and finally another simulation data which does not have similarities with the first two data types.

6.2.3 Application on the vibration signal of the electric motor in healthy state

The proposed method is firstly applied on the vibration signal for the healthy state of the electric motor. In Figure 6.10, the signal is shown between 0 and 0.25 seconds. As mentioned previously, the raw data used in the application is 5 seconds in total.

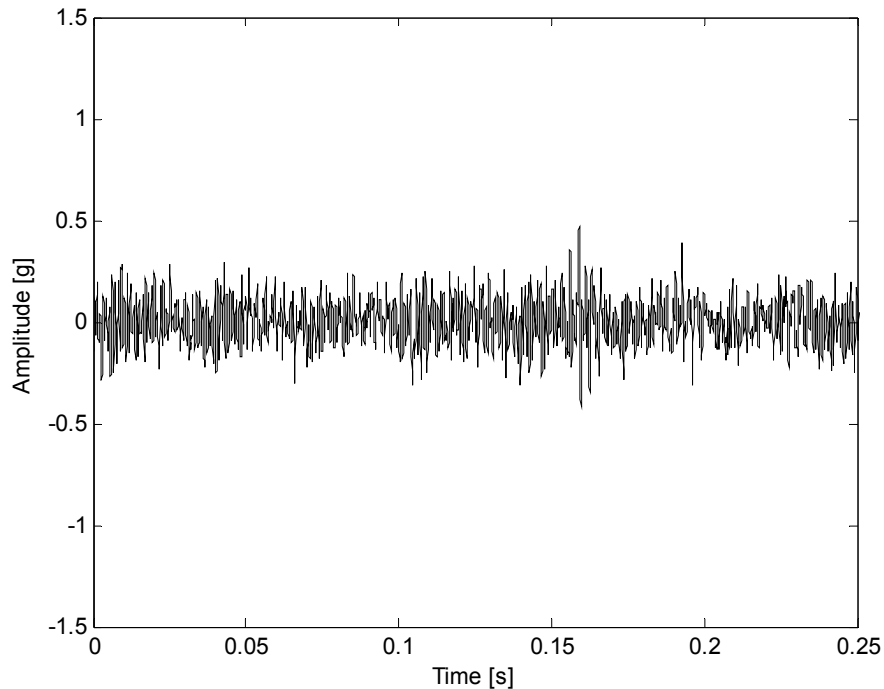


Figure 6.10 : Vibration signal of the electric motor in healthy state.

In Figure 6.11, the CWT of the signal between 0 and 0.25 seconds is shown in different scales. For the same signal between 0 and 0.25 seconds, the PSD of the CWT at the first scale is shown in Figure 6.12.

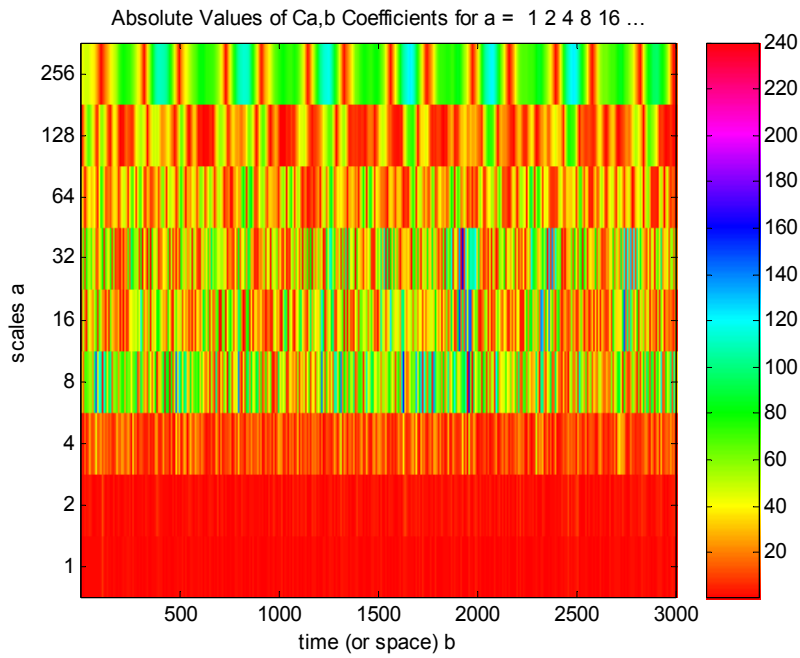


Figure 6.11 : CWT of the vibration signal between 0-0.25 seconds.

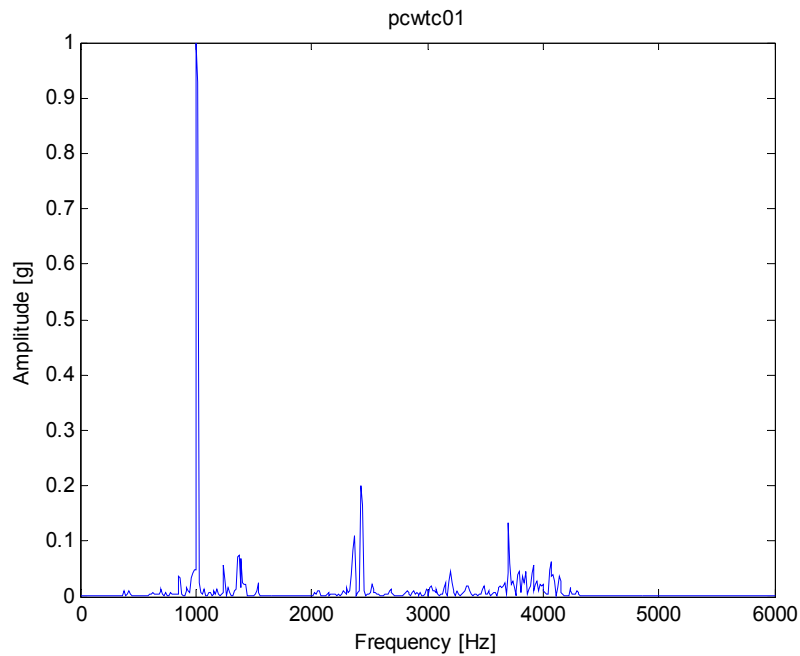


Figure 6.12 : Normalized PSD of the CWT at the first scale.

As mentioned in the training diagram of the algorithm, the normalized PSD of the CWT at the first scale, shortly pcwtc, constitutes the training and the recall sets. The total 20 sets can be seen in Appendix A. The neural network is tested with the normalized PSD of the raw signal. In Figure 6.13, the normalized PSD for the raw signal during 0 and 0.25 seconds is shown. Total 20 sets of the test data is shown in Appendix B.

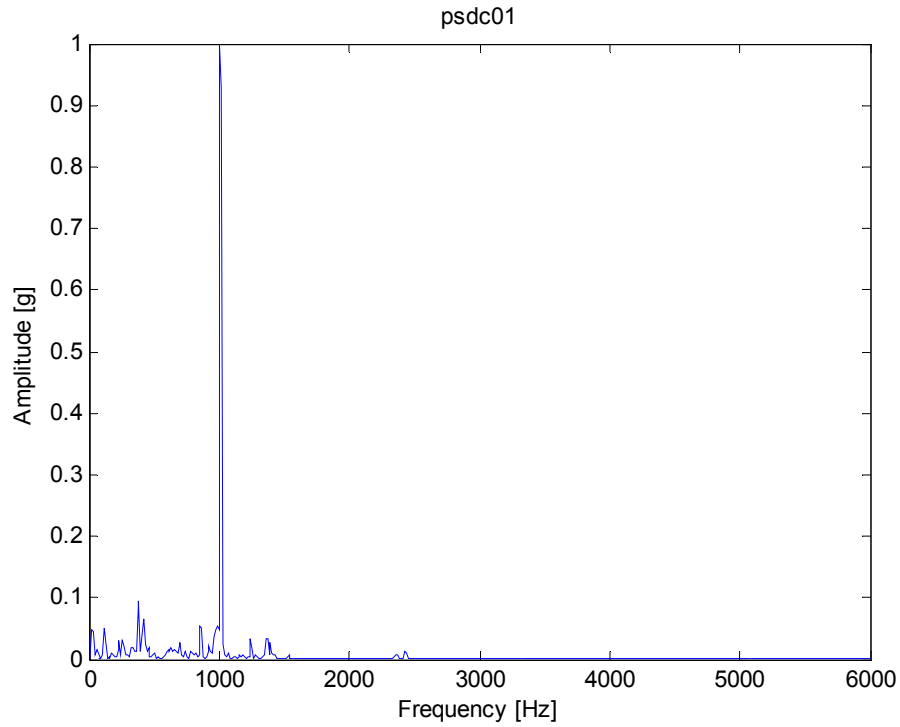


Figure 6.13 : Normalized PSD of the raw signal between 0-0.25 seconds.

For the three network structures, the details and results are listed in Table 6.1. In each run, the number of novelties are calculated by identifying the test sets that are exceeding the recall set maximum error (threshold). The table shows that in the same network structure, as the number of training samples increase, the number of test samples that are exceeding the threshold increase as well (# of novelty detection). This is a result of better training with more data sets. In each network structure, as the number of training set increases, the number of epochs, and consequently, the training duration increases to reach the MSE goal of 10^{-5} . As the number of hidden layer nodes increase, the training duration increases as well. In each network structure, for the training set of 16 samples, all of the 20 test samples are identified as novelties. As we analyze the frequency band where the novelty occurs, it is found that all of these novelties are present in the 2-4 kHz band.

Table 6.1 : AANN training and test results for the vibration signal.

Run #	Training Duration (sec)	Training Set	Recall Set	Network Structure	# of Epochs	MSE	Recall Set Maximum Error	# of Novelty Detection (/20)
1	3,3900	4	16	512:50:512	50	9,59E-06	0,3479	2
2	4,0630	8	12		61	9,91E-06	0,3283	3
3	4,5940	12	8		96	9,99E-06	0,3306	12
4	6,1720	16	4		175	9,87E-06	0,0643	20
5	3,8420	4	16	512:100:512	37	9,96E-06	0,3225	0
6	4,7010	8	12		55	9,79E-06	0,3339	12
7	5,5130	12	8		79	9,99E-06	0,2639	20
8	6,2790	16	4		98	9,97E-06	0,0492	20
9	4,8570	4	16	512:150:512	46	9,93E-06	0,3233	0
10	5,5290	8	12		56	9,38E-06	0,335	4
11	6,2310	12	8		67	9,65E-06	0,3223	7
12	6,9960	16	4		77	9,88E-06	0,0498	20

The error variations for training, recall and test samples can be seen in Figures 6.14, Figure 6.15 and Figure 6.16 respectively. The variations are for run number 8, for which the threshold is minimum.

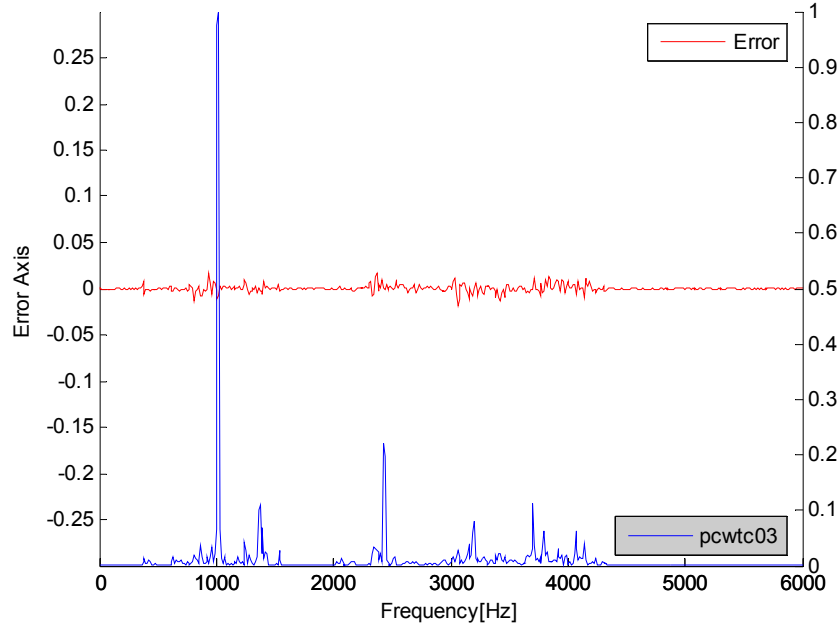


Figure 6.14 : Error variation for the training sample Pcwtc3 – run #8.

In Figure 6.14, it is clearly seen that the network is well trained since it gives very small magnitudes of error for the training set.

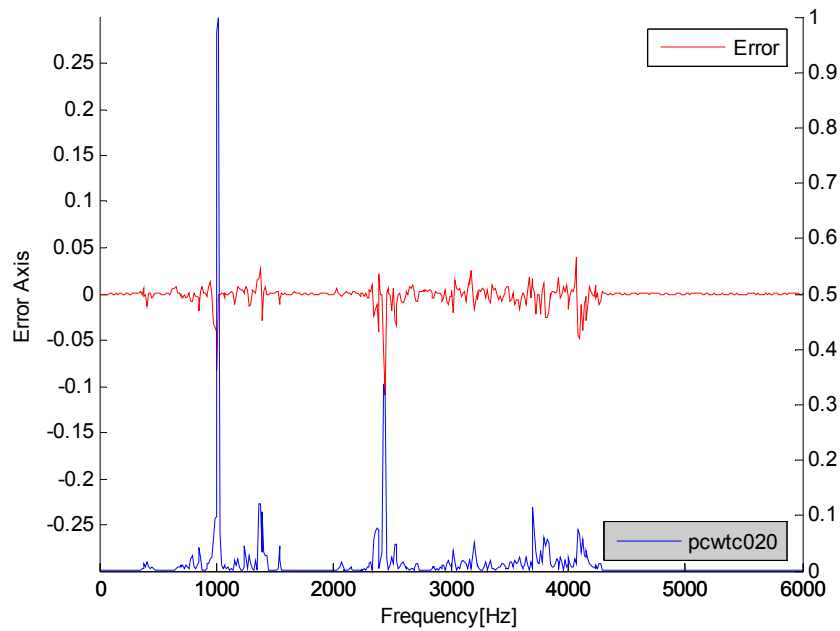


Figure 6.15 : Error variation for the recall sample Pcwtc20 – run #8.

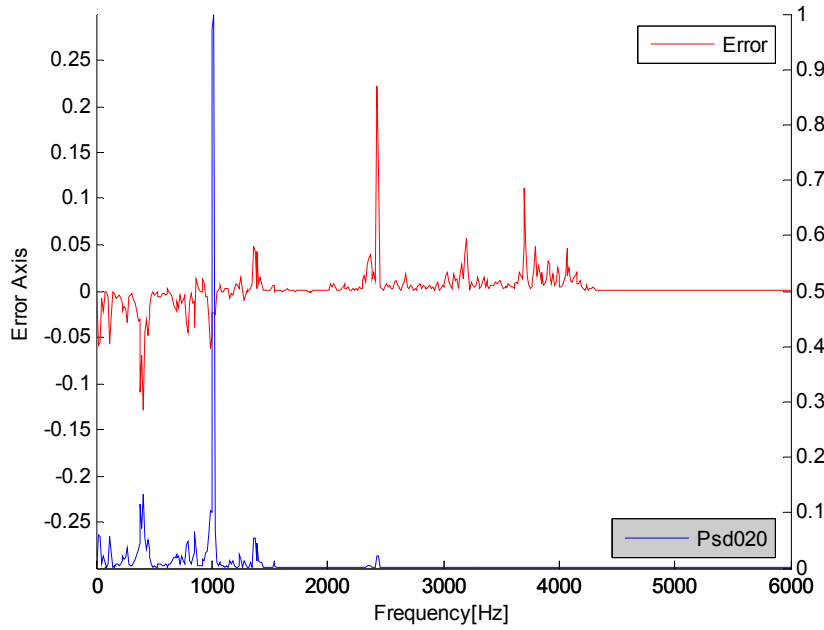


Figure 6.16 : Error variation for the test sample PSD20 – run #8.

In Figure 6.15, the recall sample Pcwtc20 creates error levels around 0.05. In Figure 6.16, the test sample PSD20 creates an output which indicates that the novelty is in the 2-4 kHz frequency band. The learning curve for run number 8 is shown in Figure 6.17, demonstrating a smooth convergence to the MSE goal of 10^{-5} .

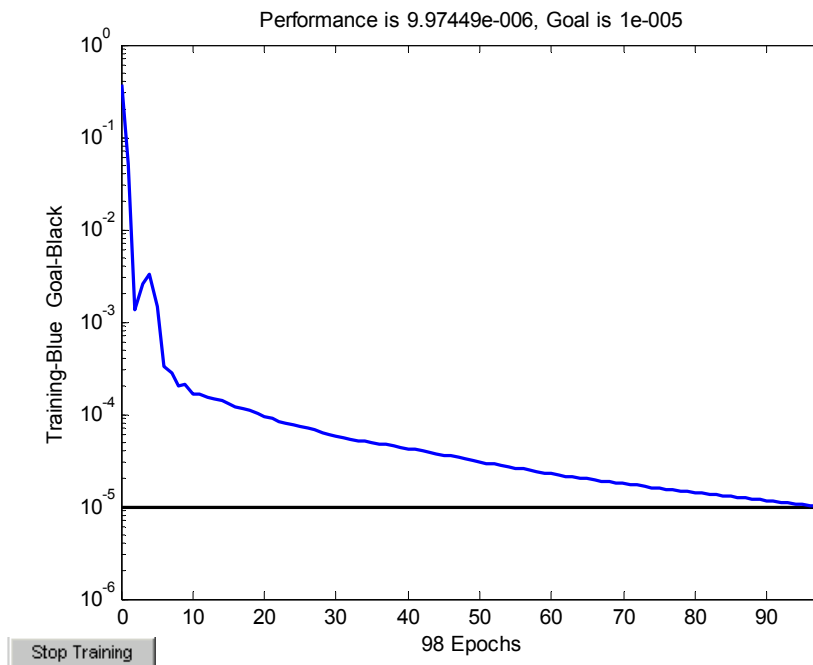


Figure 6.17 : Learning curve with RBP for run #8.

6.2.4 Application on simulation signal - I

The method is secondly applied to a simulation signal which has similar properties to the vibration data which is used in the previous section. For obtaining a similar pattern, the simulation data contains the rotation frequency, some major sinusoidal harmonics (1st, 10th, 30th harmonics) and noise in addition to these major components. Figure 6.18 represents this simulation signal. In Figure 6.19, both the vibration signal for the healthy state of the electric motor and the simulation are shown. The PSDs of these two signals are shown in Figure 6.20, in this figure it is clearly seen that the frequency characteristics are very similar to each other. Hence it indicates that the simulation signal is a good model of the vibration signal. In Figure 6.21, the CWT of the signal between 0 and 0.25 seconds is shown in different scales. For the same signal between 0 and 0.25 seconds, the normalized PSD of the CWT at the first scale is shown in Figure 6.22.

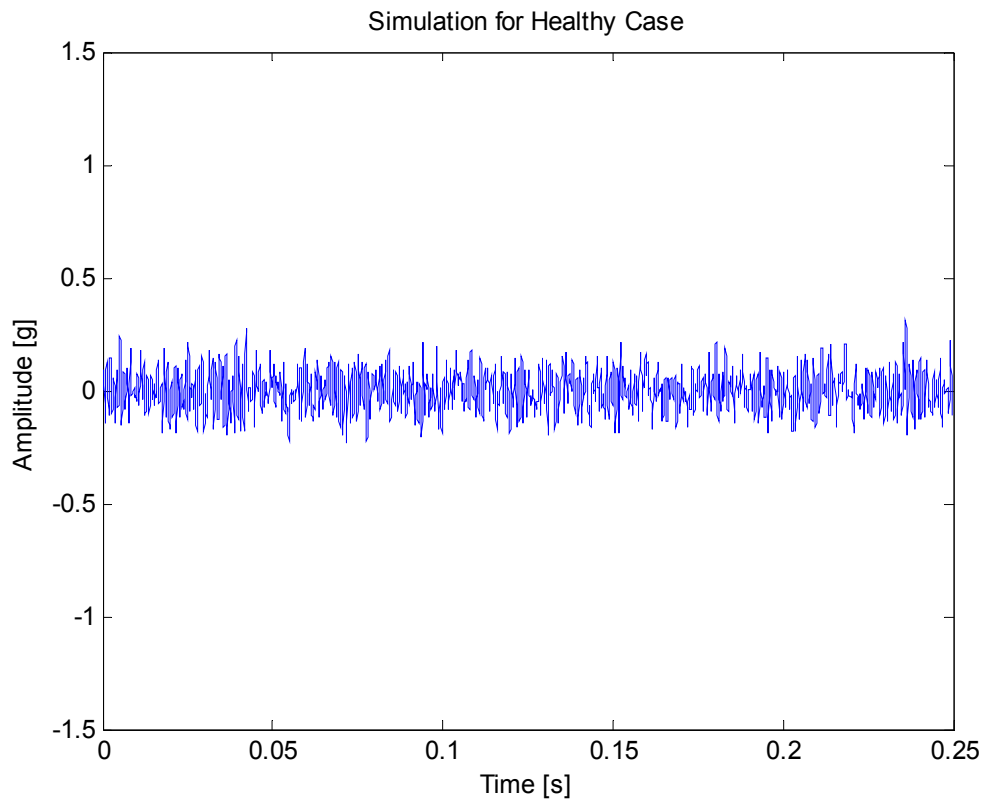


Figure 6.18 : Simulation signal I.

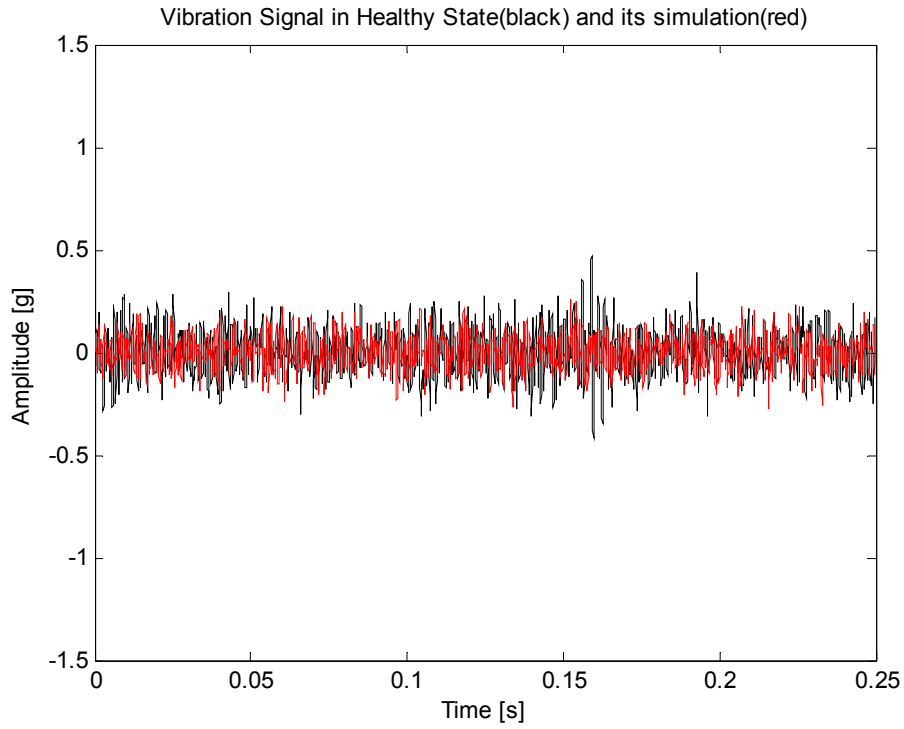


Figure 6.19 : Vibration signal (red) and simulation (black).

The same neural network structures and same number of data sets are used in this data as well.

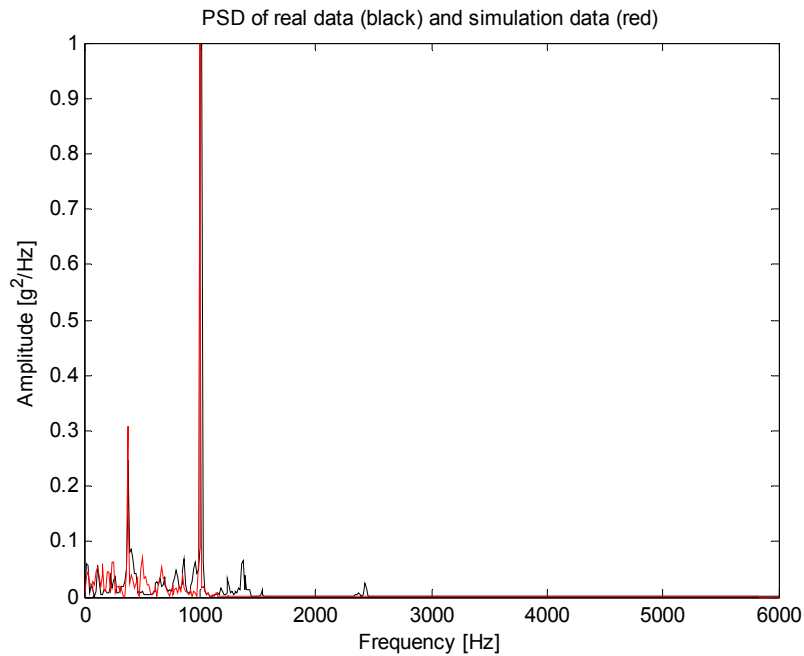


Figure 6.20 : PSD of the vibration signal (black) and simulation I (red).

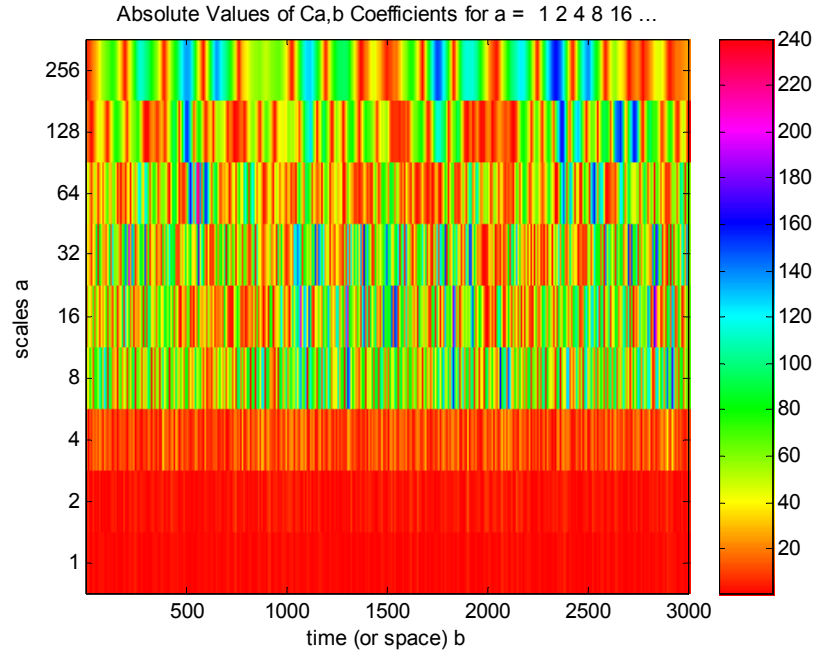


Figure 6.21 : CWT of the simulation signal between 0-0.25 seconds.

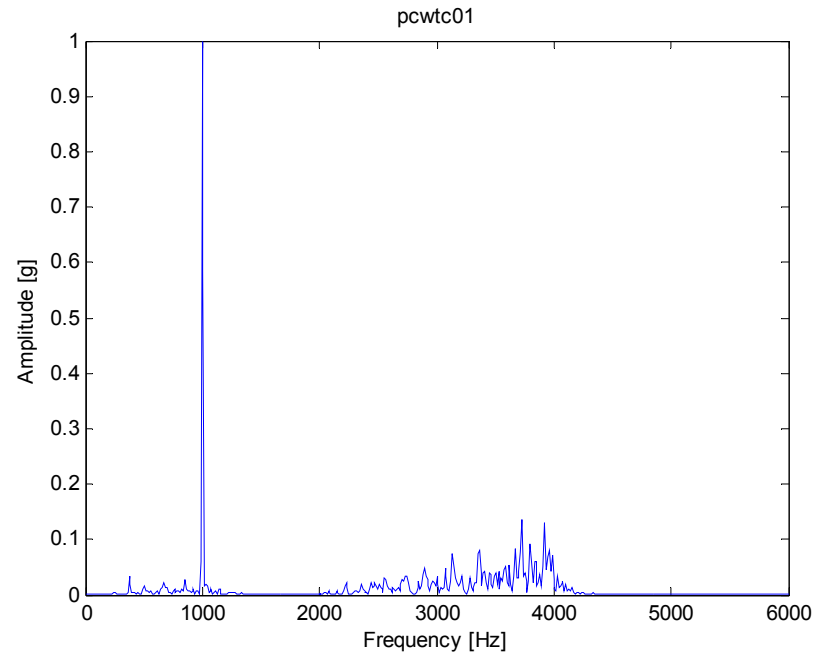


Figure 6.22 : Normalized PSD of the CWT at the first scale.

The training and test data sets for the first simulation data is shown in Appendix C and D respectively.

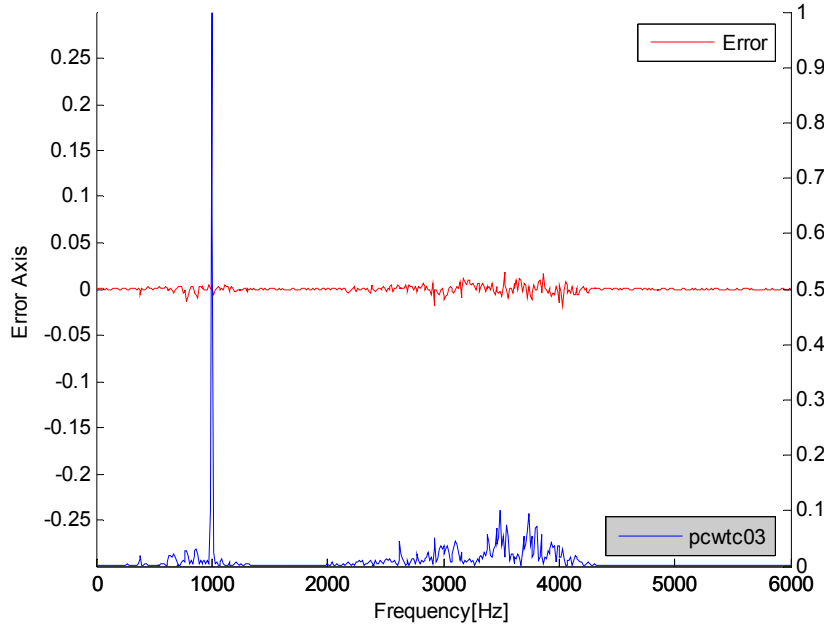


Figure 6.23 : Error variation for the training sample pcwtc3 , run #12.

For the three network structures, the details and results are given in Table 6.2. In each run, the number of novelties are calculated by identifying the test sets that are exceeding the recall set maximum error (threshold). In the table, it is clearly seen that for all number of training sets, all of the test sets are identified as novelties. The training set size is not affecting the detection ratio. The reason is the similar structure of the simulation's training sets. As the number of training data sets increase, the recall set maximum error (threshold) decreases as expected. The table shows that for all of the training sets, the performance goal is reached. The error variations for training, recall and test samples can be seen in Figures 6.23, Figure 6.24 and Figure 6.25 respectively. The variations are for run number 12.

In Figure 6.23, it is clearly seen that the network is well trained since it gives very small magnitudes of error for the training set. In Figure 6.24, the recall sample Pcwtc20 creates small error variations. In Figure 6.25, the test sample PSD20 creates an output which indicates that the novelty is in the 2-4 kHz frequency band. The learning curve for run number 12 is shown in Figure 6.26. The MSE target is reached in 82 epochs.

Table 6.2 : AANN training and test results for the simulation signal I.

Run #	Training Duration (sec)	Training Set	Recall Set	Network Structure	# of Epochs	MSE	Recall Set Maximum Error	# of Novelty Detection (/20)
1	4,1100	4	16	512:50:512	49	9,85E-06	0,2107	20
2	4,2970	8	12		84	9,87E-06	0,1107	20
3	5,2350	12	8		124	9,89E-06	0,129	20
4	6,9380	16	4		199	9,97E-06	0,0498	20
5	4,0000	4	16	512:100:512	43	9,77E-06	0,1332	20
6	4,6720	8	12		59	9,66E-06	0,0658	20
7	5,2190	12	8		75	9,92E-06	0,0552	20
8	6,1250	16	4		96	9,87E-06	0,0472	20
9	4,7340	4	16	512:150:512	42	9,12E-06	0,1013	20
10	5,1880	8	12		50	9,47E-06	0,0938	20
11	6,5780	12	8		74	9,66E-06	0,0599	20
12	7,1880	16	4		82	9,82E-06	0,0497	20

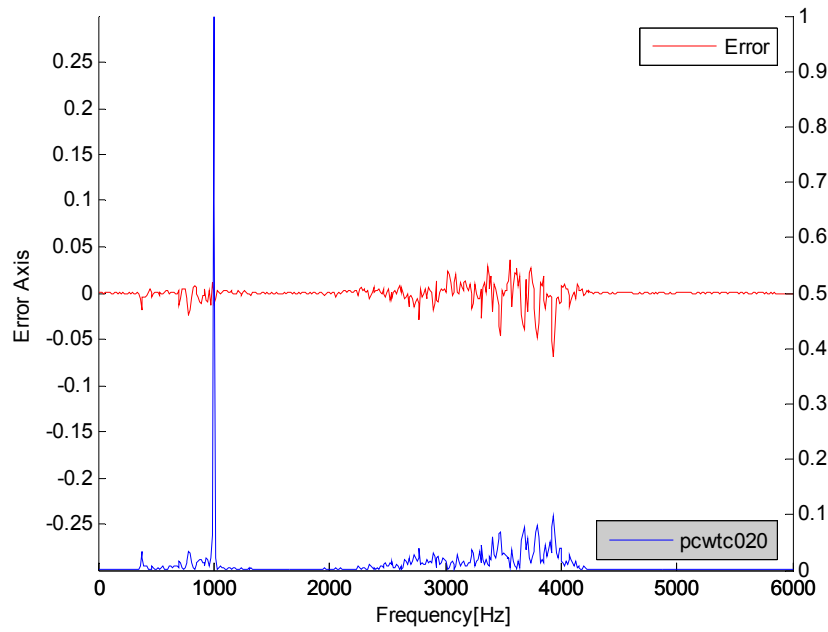


Figure 6.24 : Error variation for the recall sample pcwtc020, run #12.

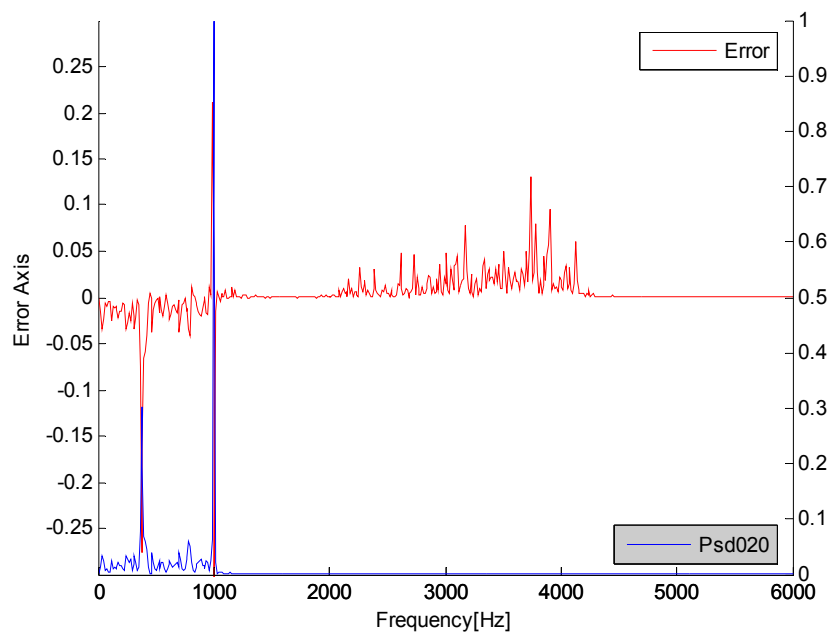


Figure 6.25 : Error variation for the test sample PSD20, run #12.

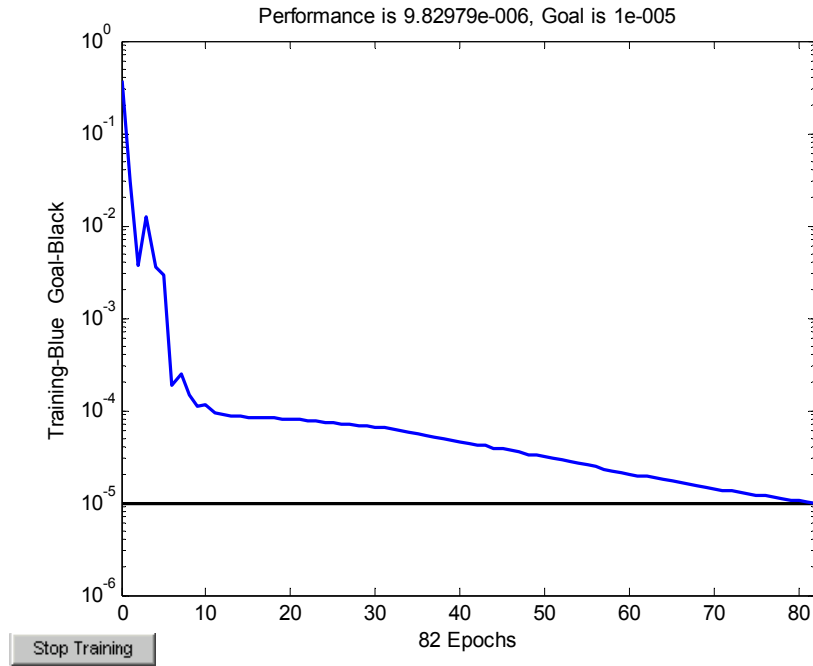
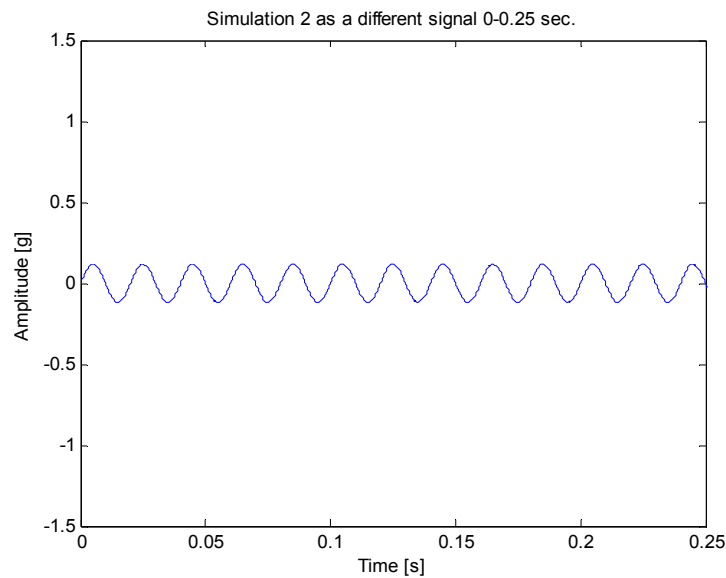


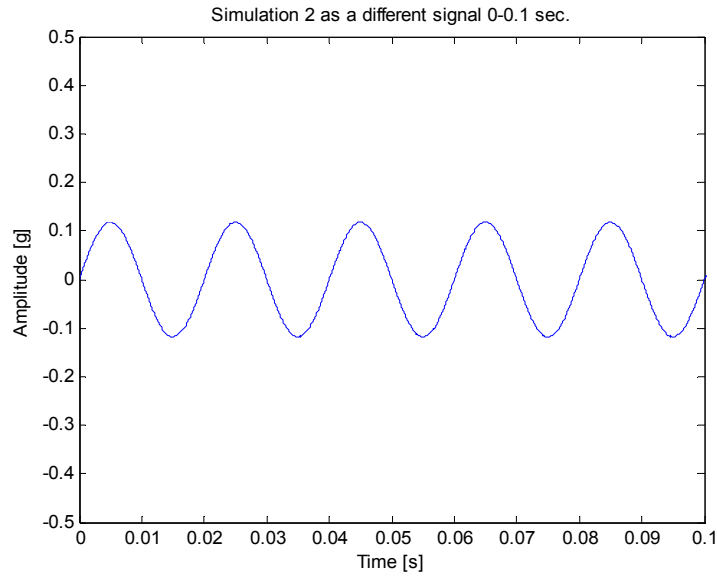
Figure 6.26 : Learning curve with RBP for run #12.

6.2.5 Application on simulation signal - II

The third application of the method is on the second simulation signal. This signal does not have similarities with neither the real vibration signal nor the first simulation signal. It has a major component oscillating at 50 Hz. Additionally, high frequency component is added on this main component with small random amplitudes. The signal can be seen in two different resolutions in Figure 6.27 a & b.



(a)



(b)

Figure 6.27 : Simulation signal II.

In Figure 6.28, the PSD of both the simulation signal II and the real vibration signal is shown. In the figure, it is clear that both signals have totally different characteristics in the frequency domain. The simulation signal only shows a big amplitude on 50 Hz. In Figure 6.29, the CWT of the simulation signal is shown at different scales.

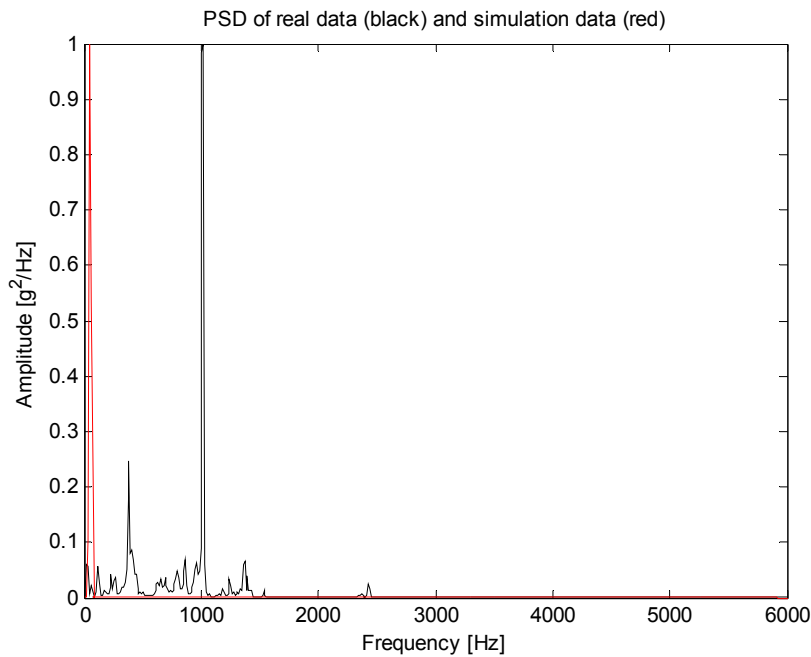


Figure 6.28 : PSD of the vibration signal (black) and simulation II (red).

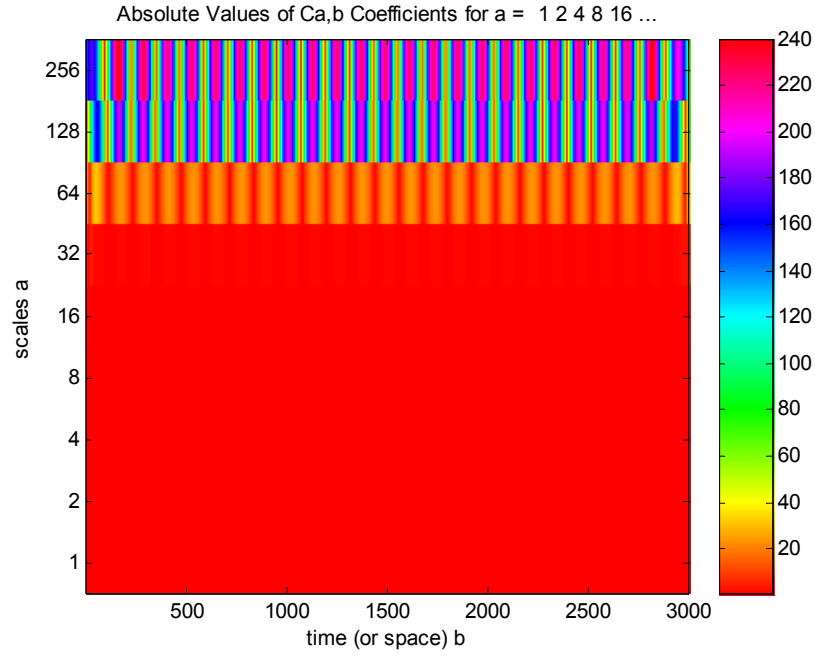


Figure 6.29 : CWT of the simulation signal II between 0-0.25 seconds.

For the same signal between 0 and 0.25 seconds, the normalized PSD of the CWT at the first scale is shown in figure 6.30.

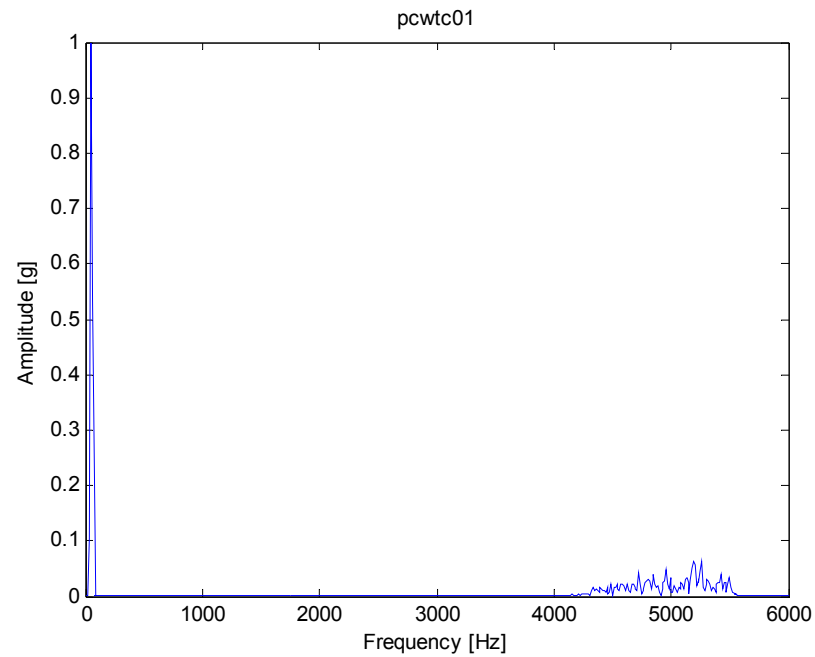


Figure 6.30 : Normalized PSD of the CWT at the first scale.

The training and test data sets for the second simulation data is shown in Appendix E and F respectively.

For the three neural network structures, the results for the second simulation data are given in Table 6.3. In the table, it is clearly seen that for all runs , all of the test sets are identified as novelties which is the same situation as the first simulation data. The novelty detection ratio is independent from the data sets. The reason for this is that the signal includes certain number of frequency components and there are no big differences between the training sets. The table shows that for all of the training sets, the performance goal is reached. The error variations for training, recall and test samples can be seen in Figures 6.31, Figure 6.32 and Figure 6.33 respectively. The variations are for run number 10. Figure 6.31, shows that the network creates very small amplitudes when the training set is used. The recall set in Figure 6.32 also shows small amounts of error. In the test set output, the novelties are occurring in 4-6 kHz band.

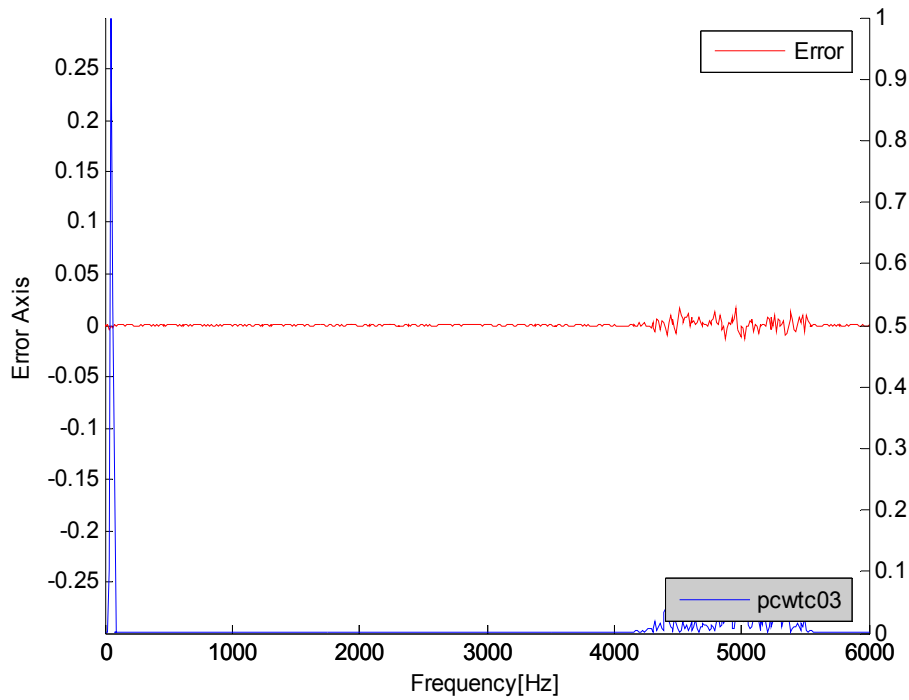


Figure 6.31 : Error variation for the training sample pcwtc3, run#10.

Table 6.3 : AANN training and test results for the simulation signal II.

Run #	Training Duration (sec)	Training Set	Recall Set	Network Structure	# of Epochs	MSE	Recall Set Maximum Error	# of Novelty Detection (/20)
1	3,4380	4	16	512:50:512	45	9,71E-06	0,0367	20
2	3,7970	8	12		66	9,73E-06	0,0439	20
3	4,5620	12	8		98	9,96E-06	0,0542	20
4	4,8910	16	4		112	9,91E-06	0,0333	20
5	3,8280	4	16	512:100:512	39	9,85E-06	0,0382	20
6	4,4840	8	12		56	9,68E-06	0,0397	20
7	5,2500	12	8		73	9,63E-06	0,0441	20
8	5,5150	16	4		79	9,90E-06	0,0351	20
9	4,7810	4	16	512:150:512	44	9,56E-06	0,0373	20
10	5,3750	8	12		54	9,47E-06	0,062	20
11	5,8440	12	8		57	9,73E-06	0,0293	20
12	6,4060	16	4		68	9,81E-06	0,0261	20

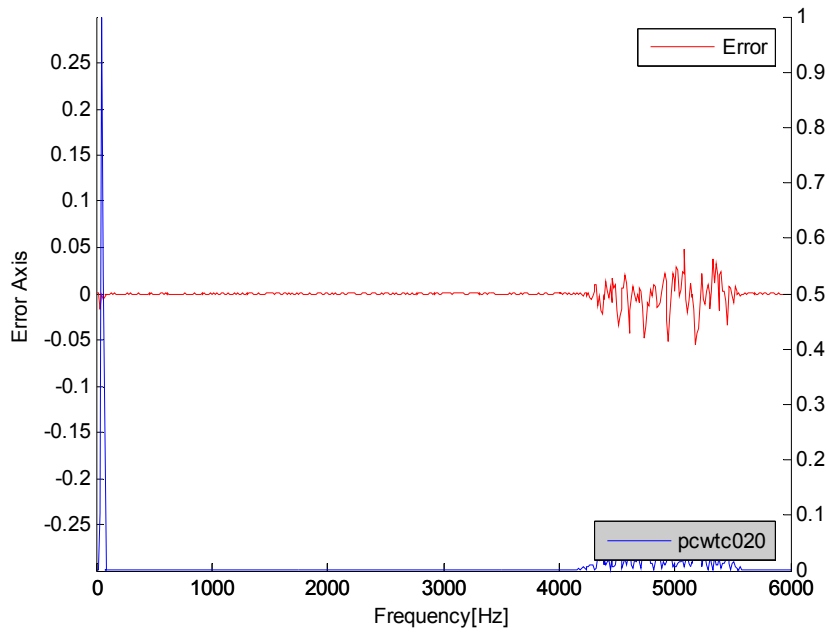


Figure 6.32 : Error variation for the recall sample pcwtc020, run #10.

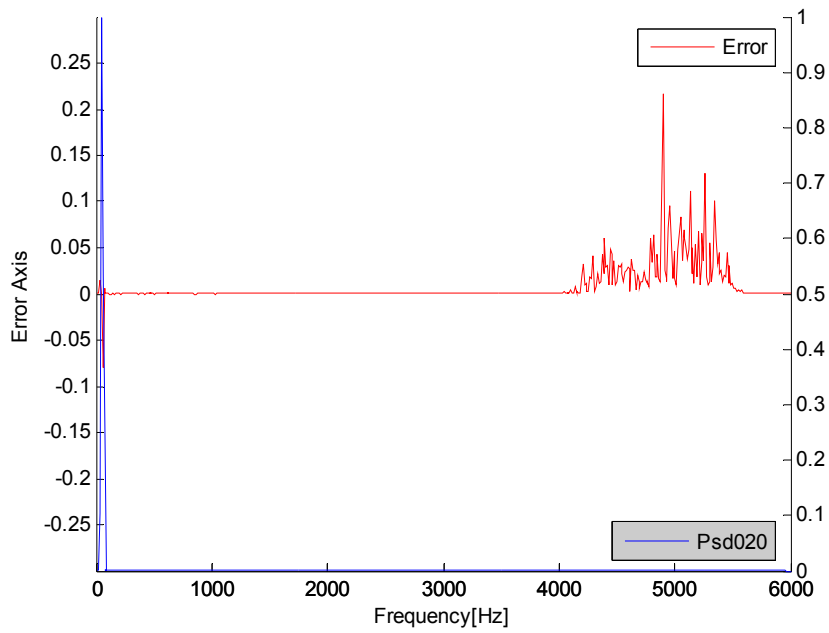


Figure 6.33 : Error variation for the test sample PSD20, run #10.

The learning curve for run number 10 is shown in Figure 6.34. The MSE target is reached in 54 epochs.

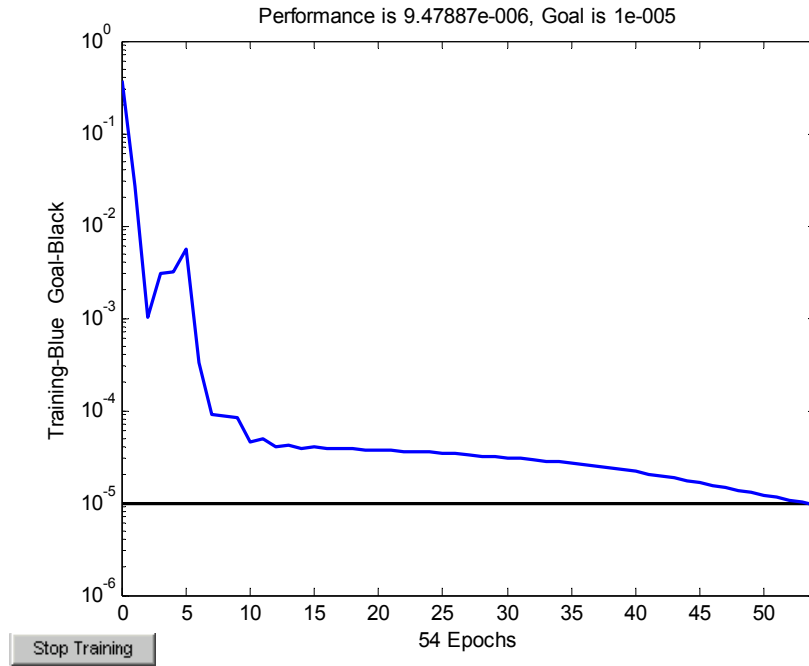


Figure 6.34 : Learning curve with RBP, run #10.

6.3 Summary of the Proposed Algorithm and Evaluation of Results

For the application of this proposed algorithm, fundamental steps can be listed as the following:

1. Vibration measurement or simulation signal is considered as raw data.
2. The raw data and its Continuous Wavelet Transform (CWT) are presented in time-scale domain.
3. First scale of the CWT is extracted.
4. Normalized Power Spectral Density (NPSD) application to the 1st scale of the CWT for frequency domain representation.
5. Training of the AANN with the training set portion of the NPSD of the first scale of the CWT.
6. Recall of the AANN with the recall set portion of the NPSD of the first scale of the CWT.
7. Threshold definition: maximum error variation for the recall set.
8. Testing the AANN with the NPSD of the raw signal.

9. Identification of the novelties as the error sets which exceed the defined threshold.

The first application involved the measured vibration signal of an electric motor in healthy state as the raw data. The novelties are identified in the 2-4 kHz band. High amplitudes in this frequency band correspond to a bearing defect, which is already identified in literature [42-44]. In the second application, a signal which is similar to the vibration signal was used. The identified novelties were again in the 2-4 kHz frequency band, which was expected. In the third application, a totally different simulation signal was used. The novelties were identified in the 4-6 kHz band. For the simulation signals, in order to understand the sensitivity of the method, the ratio of the variances were used for identifying the Signal Noise Ratio (SNR). The SNR threshold was identified by iteration. Having an SNR level of $3.3 \cdot 10^6$ and below, it is possible to identify the novelties with the proposed method.

7. CONCLUSIONS AND DISCUSSIONS

In this study, a new novelty/fault detection method which combines data mining and signal processing techniques is introduced. Time signals are used as raw data. The signals are prepared to be processed by using signal processing techniques like continuous wavelet transform and power spectral density. Applying continuous wavelet transform on the raw signal and having its representation in the first scale allows us to identify certain hidden properties which are not visible in time domain representation or via Fourier Transform. The computation which is done for the first scale of the CWT represents maximum correlation in maximum frequency range. This property is given in the mathematical interpretation of the study. After having the features extracted by the first scale representation of the CWT, PSD is applied for having the frequency domain representation of the signal. The resulting signal is normalized and given as the input to an AANN. The reason for using AANN is to get use of its identity mapping property. After the training of the AANN with different sets, same type of data is used for recalling the network. At the recall stage, a threshold is defined as the maximum value of the error that the recall set produces. By having the threshold identified, the network is then tested with the normalized PSD of the raw signal. The error values are calculated and compared with the defined threshold. According to the method, the error values that are exceeding the threshold are defined as novelties or anomalies.

The method is applied on three different types of data. The first is the vibration data of an electric motor in its healthy state. The method automatically detects novelties in the 2-4 kHz band for this signal. The studies related to bearing defects clearly show that big amplitudes in this frequency band are related with bearing defects [42-44]. Hence, it can easily be said that the method is detecting the potential defects in the electric motor by using its healthy state data only [58].

The second data used for the application is a simulation data which is created so that it has similar frequency characteristics with the healthy state vibration data. The method is again capable of detecting the novelties in the 2-4 kHz frequency band.

The last data used for the application is a simulation data which is totally different than the vibration and the first simulation data in terms of its properties. Here, the method detects novelties in the 4-6 kHz band which is hidden in the simulation data.

For the simulation data, a sensitivity analysis is done by calculating the SNR where the novelty detection starts. This SNR threshold constitutes a sensitivity metric for the method.

The majority of the feature extraction and fault detection algorithms in literature are used for classification or clustering where the faulty state data is given to the artificial intelligence system, i.e. a neural network. Then, unknown state data is tested on the same network to identify whether the input should be classified as a fault or not. In our method, only the healthy state signals of the electric motor is used and via the proposed algorithm, a feature (or hidden information) is extracted which is an indicator of a potential fault. This is a big differentiator and a contribution of our proposed method. In the applications for the simulation data, similar novelties are detected which were hidden inside.

Another contribution of the study is the usage of data mining steps in a different way with signal processing techniques and neural networks. Hence, it introduces a different feature extraction and fault diagnosis method for signals.

As mentioned in the introduction section, there are numerous amounts of data mining applications which are dealing with time series. The main purpose of using signal processing techniques in these applications is reducing the size of the raw data. However, the signal based approach used in this thesis forms a new algorithm and its application provides an effective feature extraction and fault diagnosis tool.

The application in this study uses offline processing. In other words, the signals that are used in the application have certain size and duration. This situation helps during the neural network training and test phases, since both phases need certain amount of time for processing. A new challenge for this study could be to use the algorithm for real-time processing. At that case, a slight modification could help adapting the process. The AANN would again be trained offline with some certain number of samples and then the trained network would be fed with real time data which is sampled at a suitable sampling rate.

Noisy signals could also appear as an obstacle for the proposed algorithm. In that case, regular filtering applications could be used right before the first block of the study (CWT and PSD).

A final suggestion would be to use the overall algorithm with different real signals which are obtained from different systems/domains even though it has been tested on different simulation data.

REFERENCES

- [1] **Zhang, W., Yoshida, T., Tang, X.,** (2008). Text classification based on multi-word with support vector machine, *Knowledge-Based Systems*, 21, 879-886.
- [2] **Kwon O.-W., Lee, J.-H.,** (2003). Text categorization based on k-nearest neighbor approach for Web site classification, *Information Processing and Management*, 39, 25-44.
- [3] **Özel, S.A.,** (2010). A Web Page Classification System Based on a Genetic Algorithm Using Tagged-Terms as Features, *Expert Systems with Applications*, accepted, in press.
- [4] **Hua, Y.-C., Tzeng, G.-H.,** (2003). Elicitation of classification rules by fuzzy data mining, *Engineering Applications of Artificial Intelligence*, 16, 709-716.
- [5] **Lim, C.P., Harrison, R.F., Kennedy, R.L.,** (1997). Application of autonomous neural network systems to medical pattern classification tasks. *Artificial Intelligence in Medicine*, 11, 215-239.
- [6] **Haiwei, P., Li, J., Wei, Z.,** (2007). Incorporating domain knowledge into medical image clustering, *Applied Mathematics and Computation*, 185, 844-856.
- [7] **Nanda, S.R., Mahanty, B., Tiwari, M.K.,** (2010). Clustering Indian stock market data for portfolio management, *Expert Systems with Applications*, 37, 8793-8798.
- [8] **Karamolegkos, P.N., Patrikakis, C.Z., Doulamis, N.D., Vlachas, P.T., Nikolakopoulos, I.G.,** (2009). An evaluation study of clustering algorithms in the scope of user communities assessment, *Computers and Mathematics with Applications*, 58, 1498-1519.
- [9] **Rad, A., Naderi, B., Soltani, M.,** (2011). Clustering and ranking university majors using data mining and AHP algorithms: A case study in Iran, *Expert Systems with Applications*, 38, 755-763.
- [10] **Kim, J.K., Cho, Y.H., Kim, W.J., Kim, J.R., Suh, J.H.,** (2002). A personalized recommendation procedure for Internet shopping support, *Electronic Commerce Research and Applications*, 1, 301-313.
- [11] **Song, H.S, Kim, J.K., Kim, S.H.,** (2001). Mining the change of customer behavior in an internet shopping mall, *Expert Systems with Applications*, 21, 157-168.
- [12] **Sanchez, D., Vila, M.A., Cerda, L., Serrano, J.M.,** (2009). Association rules applied to credit card fraud detection, *Expert Systems with Applications*, 36, 3630-3640.
- [13] **Farvaresh, H., Sepehri, M.M.,** (2010). A data mining framework for detecting subscription fraud in telecommunication, *Engineering Applications of Artificial Intelligence*, accepted, in press.

- [14] **Xu, J.J., Chen, H.,** (2004). Fighting organized crimes: using shortest-path algorithms to identify associations in criminal networks, *Decision Support Systems*, 38, 473-487.
- [15] **Wei, C.-P., Chiu, I.-T.,** (2002). Turning telecommunications call details to churn prediction: a data mining approach, *Expert Systems with Applications*, 23, 103-112.
- [16] **Hung, S.-Y., Yen, D.C., Wang, H.-Y.,** (2006). Applying data mining to telecom churn management, *Expert Systems with Applications*, 31, 515-524.
- [17] **Chen, W.-S., Du, Y.-K.,** (2009). Using neural networks and data mining techniques for the financial distress prediction model, *Expert Systems with Applications*, 36, 4075-4086.
- [18] **Aggarwal, C.C., Han, J., Wang, J., Yu, P.,** (2003). A Framework for Clustering Evolving Data Streams, *Proc. 29th Very Large Data Bases Conf.*
- [19] **Chan, K.-P., Fu, A.W.-C.,** (1999). Efficient Time Series Matching by Wavelets. *Proc. Int'l Conf. Data Eng.*
- [20] **Chan, F.K., Fu, A.W., Yu, C.,** (2003). Haar Wavelets for Efficient Similarity Search of Time-Series: with and without Time Warping. *IEEE Trans. on Knowledge and Data Eng.*, vol. 15, no. 3, 686-705.
- [21] **Lin, J., Keogh, E., Truppel, W.,** (2003). Clustering of Streaming Time Series Is Meaningless. *Proc. Eighth ACM SIGMOD Workshop Research Issues in Data Mining and Knowledge Discovery*, 56-65.
- [22] **Kontaki, M., Papadopoulos, A.N.,** (2004). Efficient Similarity Search in Streaming Time Sequences. *Proc. 16th IEEE Conf. on Scientific and Statistical Database Management*, 63-72.
- [23] **Mörchen, F.,** (2003). Time Series Feature Extraction for Data Mining Using DWT and DFT. Technical Report no. 33, Math and Computer Science Dept., Philipps Univ., Marburg, Germany.
- [24] **Perng, C.-S., Wang, H., Zhang, S.R., Parker, D.S.,** (2000). Landmarks: A New Model for Similarity-Based Pattern Querying in Time Series Databases, *Proc. 16th Int'l Conf. Data Eng.*
- [25] **Yi, B.-K., Faloutsos, C.,** (2000). Fast Time Sequence Indexing for Arbitrary Lp Norms Source. *Proc., 26th Int'l Conf. Very Large Data Bases*, 385-394
- [26] **Vlachos, M., Lin, J., Keogh, E., Gunopulos, D.,** (2003). A Wavelet- Based Anytime Algorithm for K-Means Clustering of Time Series. *Proc. Workshop Clustering High Dimensionality Data and Its Applications, Third SIAM Int'l Conf. Data Mining*, May.
- [27] **Keogh, E.J., Pazzani, M.J.,** (1998). An Enhanced Representation of Time Series Which Allows Fast and Accurate Classification, Clustering and Relevance Feedback. *Proc. Fourth Int. Conf. Knowledge Discovery and Data Mining (KDD 98)*, 239-241.

- [28] **Gunopulos, D., Das, G.,** (2000). Time Series Similarity Measures. *Tutorial Notes Sixth Int'l Conf. Knowledge Discovery and Data Mining*, 243-307.
- [29] **Fink, E., Pratt, K.B., Gandhi, H.S.,** (2003). Indexing of Time Series by Major Minima and Maxima. *Proc. IEEE Int'l Conf. on Systems, Man, and Cybernetics*.
- [30] **Wang, X.-Y., Wang, Z.-O.,** (2002). Stock Market Time Series Data Mining Based on Regularized Neural Network and Rough Set. *Proceedings of the First International Conference on Machine Learning and Cybernetics*, Beijing, 315-318.
- [31] **Last, M., Klein, Y., Kandel, A.,** (2001). Knowledge Discovery in Time Series Databases. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, Vol 31, No 1, February, 160-169.
- [32] **Liu, H., Li, L., Zhu, H.,** (2005). Genetic Neural Network Based Data Mining and Application in Case Analysis of Police Office, *Proceedings of International Symposium on Spatio-temporal Modeling, Spatial Reasoning, Analysis, Data Mining and Data Fusion*, Peking University, China.
- [33] **Chou, P.-H., Li, P.-H., Chen, K.-K., Wu, M.-J.,** (2010). Integrating web mining and neural network for personalized e-commerce automatic service. *Expert Systems with Applications* ,37, 2898-2010
- [34] **Kuo, R.J., Chen, J.A.,** (2004). A decision support system for order selection in electronic commerce based on fuzzy neural network supported by real-coded genetic algorithm. *Expert Systems with Applications* ,26, 141-154
- [35] **Mazurowski, M.A., Habas, P.A., Zurada, J.M., Lo, J.Y., Baker, J.A., Tourassi, G.D.,** (2008). Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. *Neural Networks*, 21, 427-436.
- [36] **Enke, D. ,Thawornwong, D.,** (2005). The use of data mining and neural networks for forecasting stock market returns., *Expert Systems with Applications*, 29, 927-940.
- [37] **Faria, E.L., Albuquerque, M.P., Gonzalez, J.L., Cavalcante, J.T.P., Albuquerque, Ma.P.,** (2009). Predicting the Brazilian stock market through neural networks and adaptive exponential smoothing methods. *Expert Systems with Applications*, 36, 12506-12509.
- [38] **Patterson, D.W., Chan, K.H., Tan, C.M.,** (1993). Time Series Forecasting with neural nets: a comparative study., *Proc. the international conference on neural network applications to signal processing, NNASP*, Singapore, pp 269-274
- [39] **Bengio, S., Fessant, F., Collobert, D.,** (1995). A Connectionist System for Medium-Term Horizon Time Series Prediction., *In Proc. Intl. Workshop Application Neural Networks to Telecoms*, pp308-315
- [40] **Singh, Y., Alok, S.C.,** (2009). Neural Network in Data Mining, *Journal of Theoretical and Applied Information Technology*, Vol 5 No1: 37-42

- [41] **Nandi, S., Toliyat, H., Li, X.,** (2005). Condition monitoring and fault diagnosis of electrical motors – a review, *IEEE Trans. on Energy Conversion* 2005, 20(4): 719-729.
- [42] **Şeker, S., Ayaz, E.,** (2003). Feature extraction related to bearing damage in electric motors by wavelet analysis, *Journal of the Franklin Institute*, 2003; 340(2): 125-134.
- [43] **Şeker, S., Öztürk, A., Ayaz, E., Upadhyaya, B.R., Kayran, A.H.,** (2006). Multi-Resolution Wavelet Analysis for Detection of Potential Defects in Motor Bearings Using Vibration Signals, *IASTED Signal and Systems Conference*, CA San Francisco, November.
- [44] **Ayaz, E., Öztürk, A., Şeker, S.,** (2006). Continuous Wavelet Transform for Bearing Damage Detection in Electric Motors, *The 13th IEEE Mediterranean Electrotechnical Conference*, May 16-19, Benalmádena (Málaga), Spain.
- [45] **Lin, J.,** (2001). Feature extraction of machine sound using wavelet and its application in fault diagnosis, *NDT&E International*, 34, 25-30.
- [46] **Cvetkovic, D., Übeyli, E.D., Cosic, I.,** (2008). Wavelet transform feature extraction from human PPG, ECG, and EEG signal responses to ELF PEMF exposures: A pilot study, *Digital Signal Processing*, 18, 861-874.
- [47] **Çelik, T., Özkahramanlı, H., Demirel, H.,** (2008). Facial feature extraction using complex dual-tree wavelet transform, *Computer Vision and Image Understanding*, 111, 229-246.
- [48] **Şeker, S., Gullulu, S., Ayaz, E.,** (2008). Transfer Function approach based upon wavelet transform for bearing damage detection in electric motors, *IEEE International Symposium on Industrial Electronics*, 749-752.
- [49] **Engelberg, S.,** (2008). *Digital Signal Processing: An Experimental Approach*, Springer-Verlag, London, UK.
- [50] **Akansu, A.N., Haddad, R.A.,** (1992). *Multiresolution Signal Decomposition*, Academic Press, USA.
- [51] **Robertson, D.C., Camps, O.I., Mayer, J.S., Gish, W.B.,** (1996). Wavelets and Electromagnetic Power System Transients, *IEEE Transaction on Power Delivery*, 11, 1050-1068.
- [52] **Weiss, L.G.,** (1994). Wavelets and Wideband Correlation Processing, *IEEE SP Magazine*, 11, 1.
- [53] **Singh, Y., Alok, S.C.,** (2009). Neural Network in Data Mining, *Journal of Theoretical and Applied Information Technology*, Vol 5 No1: 37-42
- [54] **Baldi, P., Hornik, K.,** (1989). Neural Networks and Principal Component Analysis: Learning from Examples Without Local Minima, *Neural Networks*, 2(1), 53-58.
- [55] **Marini, F., Magri, A.L., Bucci, R.,** (2007). Multilayer feed-forward artificial neural networks for class modeling, *Chemometrics and Intelligent Laboratory Systems*, 88, 118-124.

- [56] **Zaiane, O.R.,** (1999). *Principles of Knowledge Discovery in Databases*, University of Alberta, Computing Science Department, 1-15.
- [57] **Rygielski, C., Wang, J.-C., Yen, D.C.,** (2002). Data mining techniques for customer relationship management. *Technology in Society*, 24, 483-502.
- [58] **Güllülü, S., Şeker, S.,** (2011). Signal Based Approach for Data Mining in Fault Detection of Induction Motor. *Scientific Research Essays*, 6(22), 4720-4731
- [59] **Riedmiller, M., Braun, H.,** (1993). A Direct Adaptive Method for Faster Backpropagation Learning: The RPOP Algorithm. *IEEE International Conference on Neural Networks*, 586-591.
- [60] **More, J.J.,** (1978). The Levenberg-Marquardt Algorithm: Implementation and Theory. *Lecture Notes in Mathematics*, 630/1978, 105-116.
- [61] **Haykin, S.,** (1999). Learning Processes. *Neural Networks: A Comprehensive Foundation*, 50-116, Prentice Hall International.
- [62] **Rosenblatt, F.,** (1958). The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, vol.65, 386-408.
- [63] **Haykin, S.,** (1999). Multilayer Perceptrons. *Neural Networks: A Comprehensive Foundation*, 156-255, Prentice Hall International.
- [64] **Kramer, M.A.,** (1992). Autoassociative Neural Networks. *Computers & Chemical Engineering*, 16(4), 313-328.
- [65] **Kramer, M.A.,** (1991). Nonlinear Principal Component Analysis Using Autoassociative Neural networks. *AIChE*, 37, 233-243.
- [66] **Du, Y., Wang, X.,** (2010). Research on Fault Detection of High Voltage Inverter Based on STFT. *Chinese Control and Decision Conference*, 1(5), 2229-2233.
- [67] **Koo, I.S., Kim, W.W.,** (2000). Development of Reactor Coolant Pump Vibration Monitoring and a Diagnostic System in the Nuclear Power Plant. *ISA Transactions*, 39, 309-316.
- [68] **Francois, A., Patrick, F.,** (1995). Improving the Readability of Time-Frequency and Time-Scale Representations by the Reassignment Method. *IEEE Transactions on Signal Processing*, 43, 1068-1089.
- [69] **Wang, W.J., McFadden, P.D.,** (1993). Application of the Wavelet Transform to Gearbox Vibration Analysis. *American Society of Mechanical Engineers – Petroleum Division*, 52, 13-20.
- [70] **Cai, Y., He, Y., Li, A., Zhao, J., Wang, T.,** (2010). Application of Wavelet to Gearbox Vibration Signals for Fault Detection. *2nd International Conference on Advanced Computer Control*, 441-444.
- [71] **Newland, D.E.,** (1994). Wavelet Analysis of Vibration, Part I: theory. *Journal of Vibration and Acoustics*, Transactions of the ASME, 116, 409-416.

- [72] **Newland, D.E.**, (1994) Wavelet Analysis of Vibration, Part 2: Wavelet Maps. *Journal of Vibration and Acoustics*, Transactions of the ASME, 116, 417-425.
- [73] **Newland, D.E.**, (1994). Some Properties of Discrete Wavelet Maps. *Probabilistic Engineering Mechanics*, 9, 59-69.
- [74] **Newland, D.E.**, (1995). Progress in the Application of Wavelet Theory to Vibration Analysis. *American Society of Mechanical Engineers, Design Engineering Division*, 84, 1313-1322.
- [75] **Wang, W.Q., Ismail, F., Golnaragh, M.F.**, (2001). Assessment of Gear Damage Monitoring Techniques Using Vibration Measurements. *Mechanical Systems and Signal Processing*, 15, 905-922.
- [76] **Sung, C.K., Tai, H.M., Chen, C.W.**, (2000). Locating Defects of a Gear System by the Technique of Wavelet Transform. *Mechanism and Machine Theory*, 15, 1169-1182.
- [77] **Suh, J.H., Kumara, S.R.T., Mysore, S.P.**, (1999). Machinery Fault Diagnostics and Prognosis: Application of Advanced Signal Processing Techniques. *CIRP Annals—Manufacturing Technology*, 48, 317-320.
- [78] **Adewusi, S.A., Al-Bedoor, B.O.**, (2001). Wavelet Analysis of Vibration Signals of an Overhang Rotor with a Propagating Transverse Crack. *Journal of Sound and Vibration*, 246, 777-793.
- [79] **Lotfollahi-Yaghin, M.A., Hesari, M.A.**, (2008). Using Wavelet Analysis in Crack Detection at the Arch Concrete Dam. *World Applied Sciences Journal*, 3(4), 691-704.
- [80] **Chen, B.H., Wang, X.Z., Yang, S.H., McGreavy, C.**, (1999). Application of Wavelets and Neural Networks to Diagnostic System Development, 1, Feature Extraction. *Computers and Chemical Engineering*, 23, 899-906.
- [81] **Yen, G.G., Lin, K.-C.**, (1999). Conditional Health Monitoring Using Vibration Signatures. *Proceedings of the IEEE Conference on Decision and Control*, 5, 4493-4498.
- [82] **Momoh, J.A., Dias, L.G.**, (1996). Solar Dynamic Power System Fault Diagnostics. *NASA Conference Publication*, 10189, 19.
- [83] **Momoh, J.A., Oliver, W.E.J., Dolce, J.L.**, (1995). Comparison of Feature Extractors on DC Power System Faults for Improving ANN Fault Diagnostics Accuracy. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 4, 3615-3623.
- [84] **Akbaryan, F., Bishnoi, P.R.**, (2001). Fault Diagnostics of Single-Variate Systems Using a Wavelet-based Pattern Recognition Technique. *Industrial and Engineering Chemistry Research*, 40, 3612-3622.
- [85] **Aminian, F., Aminian, M.**, (2001). Fault Diagnostics of Analog Circuits Using Bayesian Neural Networks with Wavelet Transform as Preprocessor. *Journal of Electronic Testing: Theory and Applications (JETTA)*, 17, 29-36.

- [86] **Kerschen, G., Golinval, J.-C.,** (2004). Feature Extraction Using Auto-Associative Neural Networks. *Smart Materials and Structures*, 13, 211-219.
- [87] **Sanz, J., Perera, R., Huerta, C.,** (2007). Fault Diagnosis of Rotating Machinery Based on Auto-Associative Neural Networks and Wavelet Transforms. *Journal of Sound and Vibration*, 302, 981-999.
- [88] **Apte, C., Weiss, S., Grout, G.,** (1993). Predicting Defects in Disk Drive Manufacturing: A Case Study in High Dimensional Classification. *IEEE Annual Computer Science Conference on Artificial Intelligence in Application*, 212-218.
- [89] **Grout, G., Tamminga, S.,** (1990). RAES – Improving the DASD Rework Process. *Proceedings of the IBM ITL Conference on Expert Systems*, 331-337.
- [90] **Maki, H., Teranishi, Y.,** (2001). Development of Automated Data Mining System for Quality Control in Manufacturing. *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, 93-100.
- [91] **Liao, T.W., Li, D.M., Li, Y.M.,** (1999). Detection of Welding Flaws From Radiographic Images With Fuzzy Clustering Methods. *Fuzzy Sets Systems*, 108, 145-158.
- [92] **Liao, T.W., Wang, G., Triantaphyllou, E., Chang, P.C.,** (2001). A Data Mining Study of Weld Quality Models Constructed With MLP Neural Networks From Stratified Sample Data. *Industrial Engineering Research Conference*, 6.
- [93] **Bangura, J.F., Povinelli, R.J., Demerdash, N.A.O., Brown, R.H.,** (2003). Diagnostics of Eccentricities and Bar/End-Ring Connector Breakages in Polyphase Induction Motors Through a Combination of Time-Series Data Mining and Time-Stepping Coupled FE-State-Space Techniques. *IEEE Transactions on Industry Applications*, 39(4), 1005-1013.
- [94] **Povinelli, R.J., Bangura, J.F., Demerdash, N.A.O., Brown, R.H. ,** (2002). Diagnostics of Bar/End-Ring Connector Breakage Faults in Polyphase Induction Motors Through a Novel Dual Track of Time-Series Data Mining and Time-Stepping Coupled FE-State Space Modeling. *IEEE Transactions on Energy Conversion*, 17(1), 39-46.
- [95] Global Temperature Change. *Newscientist*. Retrieved from <http://www.newscientist.com/article/dn11639-climate-myths-the-cooling-after-1940-shows-co2-does-not-cause-warming.html> (31.08.2010)
- [96] Measuring Earthquakes, *Earth Science in Maine*, Retrieved from <http://earthscienceinmaine.wikispaces.com/7.3+Measuring+and+Predicting+Earthquakes> (31.08.2010)
- [97] **Rao, R.M., Bopardikar, A.S.,** (1998). *Wavelet Transforms: Introduction to Theory and Applications*. Addison Wesley Longman, 8-11.

APPENDICES

APPENDIX A: Training Set of the Vibration Signal (PSD of CWT)

APPENDIX B : Test Set of the Vibration Signal (PSD)

APPENDIX C : Training Set of the Simulation Signal - I (PSD of CWT)

APPENDIX D : Test Set of the Simulation Signal - I (PSD)

APPENDIX E : Training Set of the Simulation Signal - II (PSD of CWT)

APPENDIX F : Test Set of the Simulation Signal - II (PSD)

APPENDIX G : Graphical User Interface (GUI) Outputs

APPENDIX A: Training Set of the Vibration Signal (PSD of CWT)

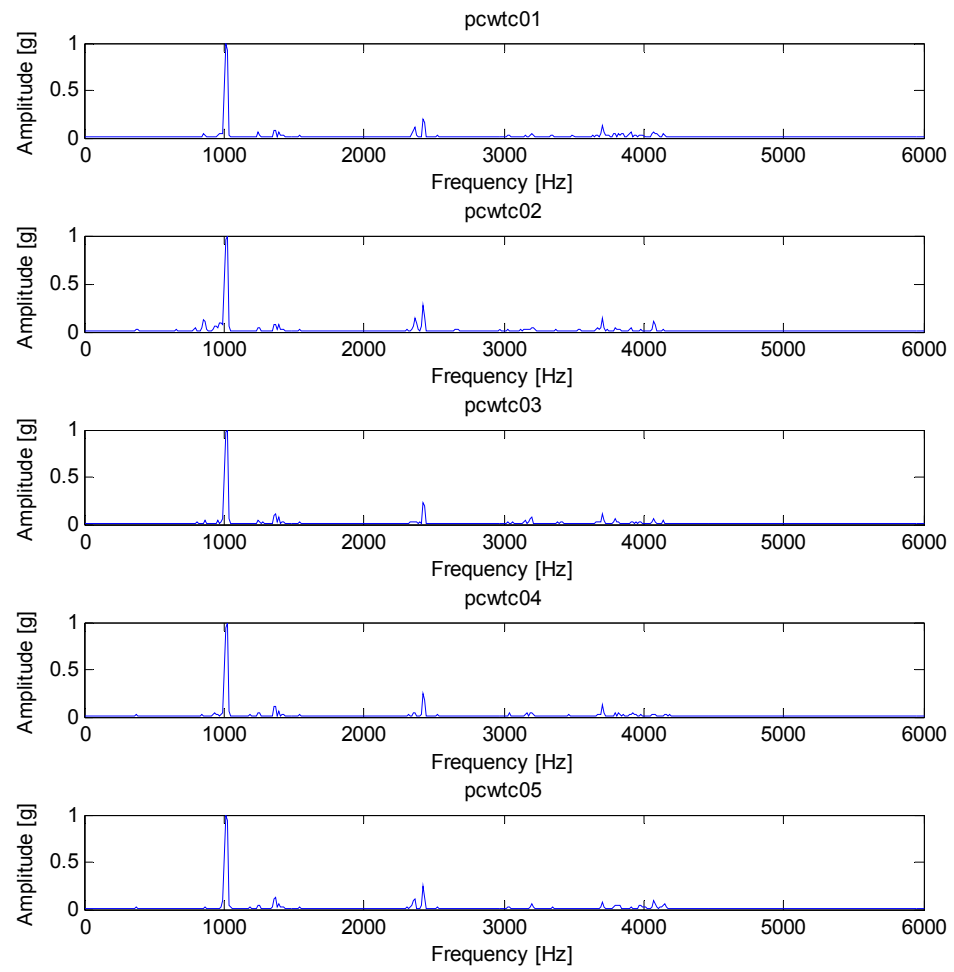


Figure A.1 : Pcw1-Pcw5

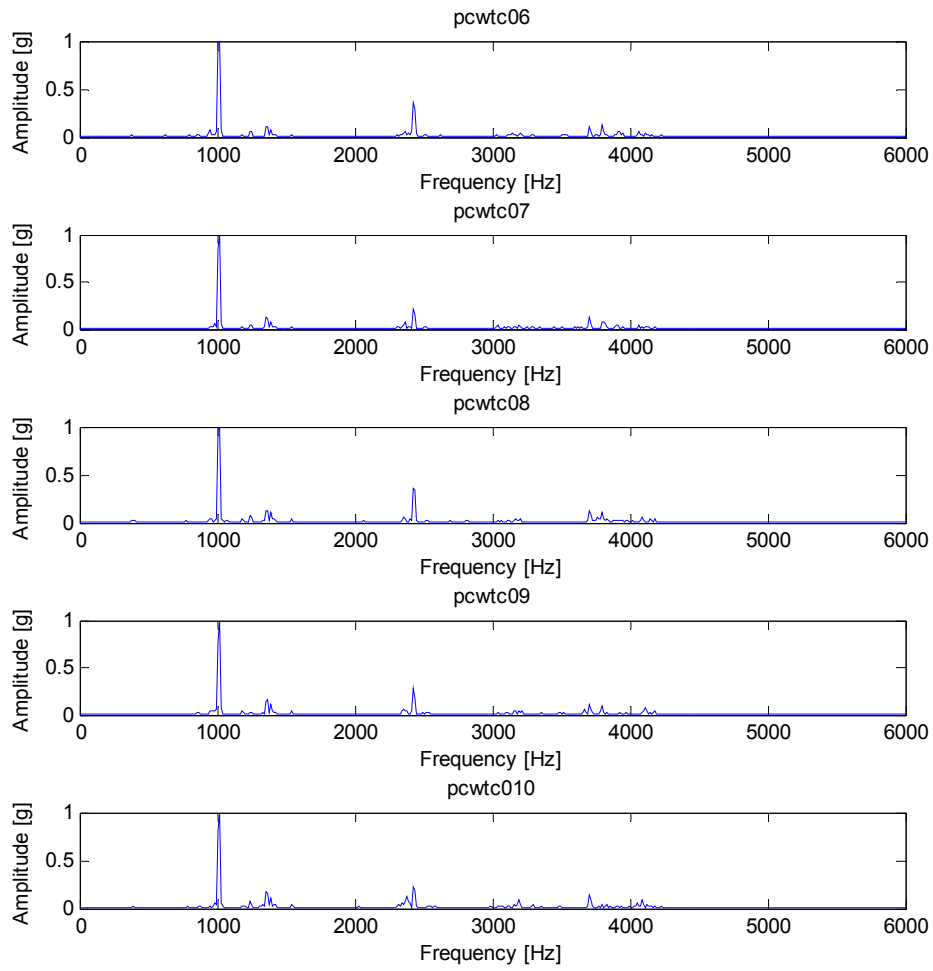


Figure A.2 : Pcwt6-Pcwt10

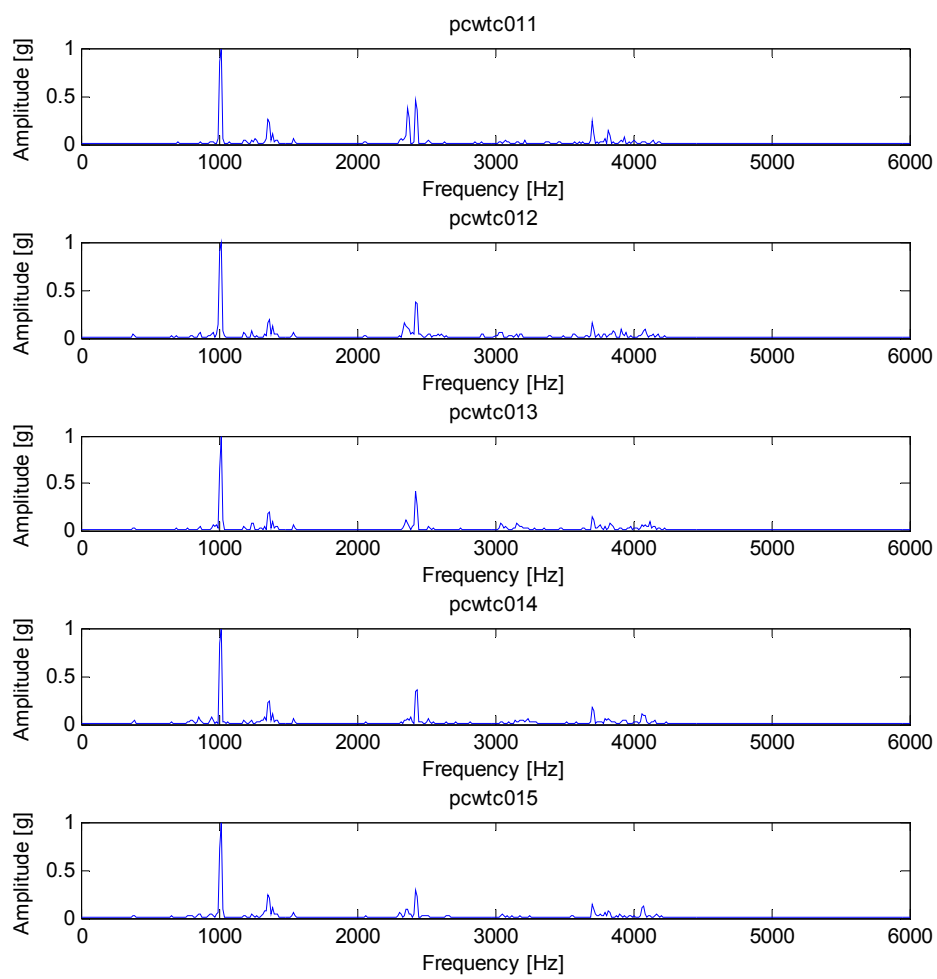


Figure A.3 : Pcwt11-Pcwt15

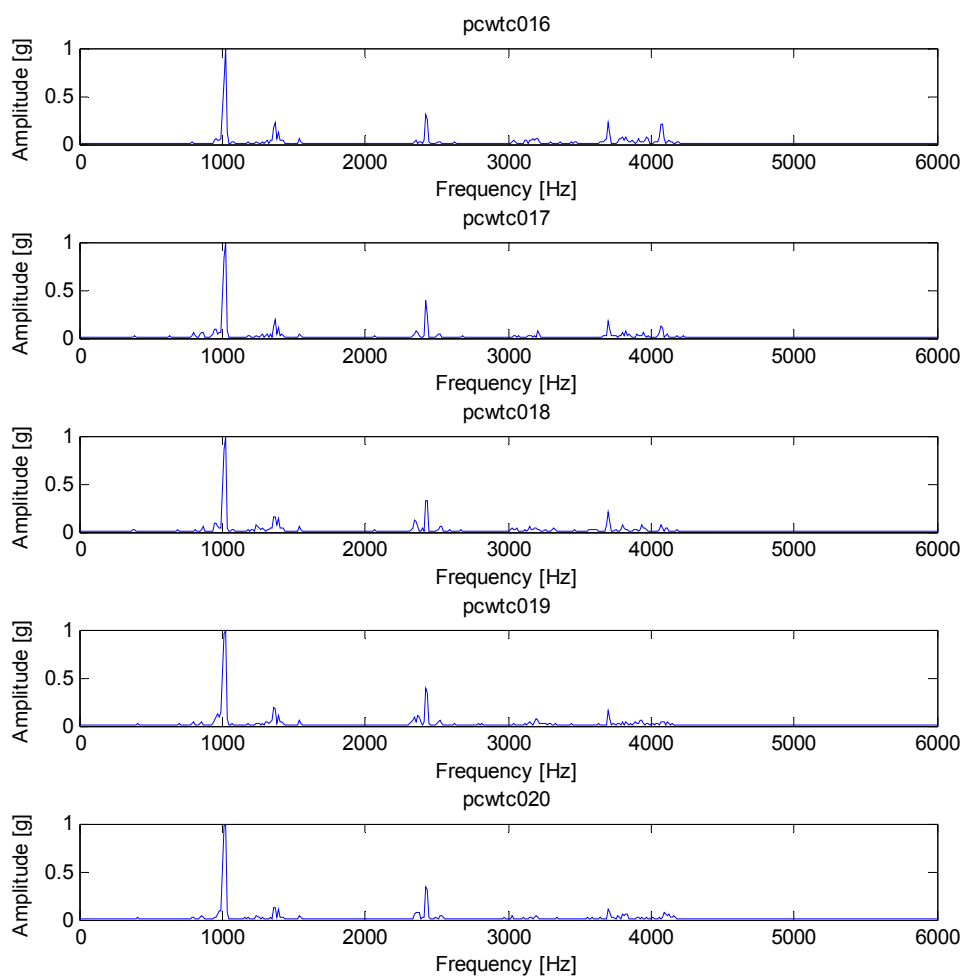


Figure A.4 : Pcwt16-Pcwt20

APPENDIX B: Test Set of the Vibration Signal (PSD)

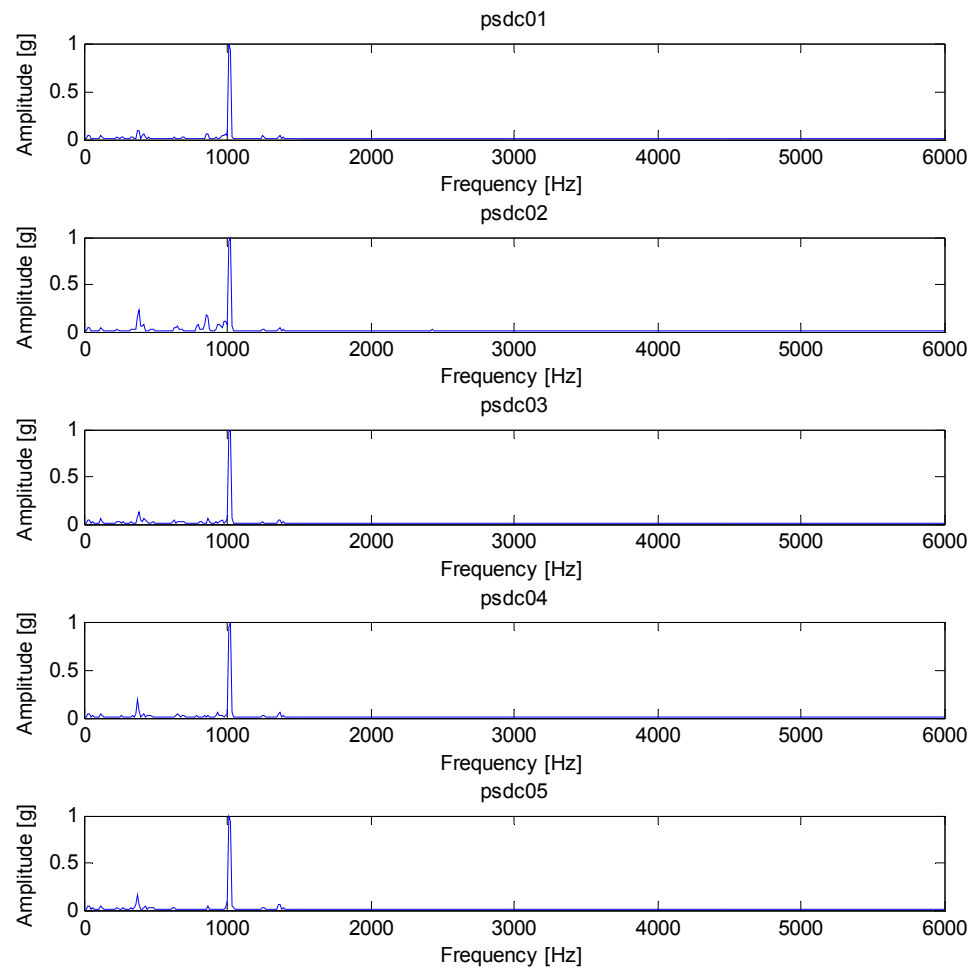


Figure B.1 : PSD1-PSD5

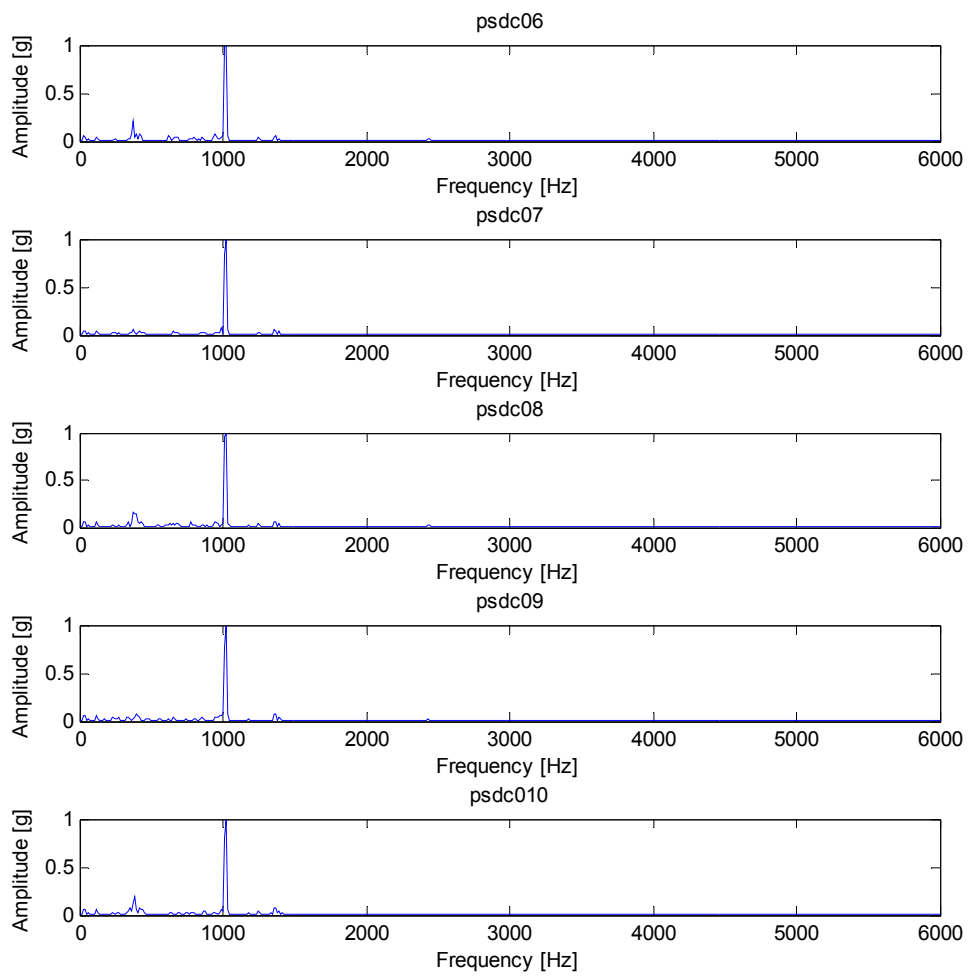


Figure B.2 : PSD6-PSD10

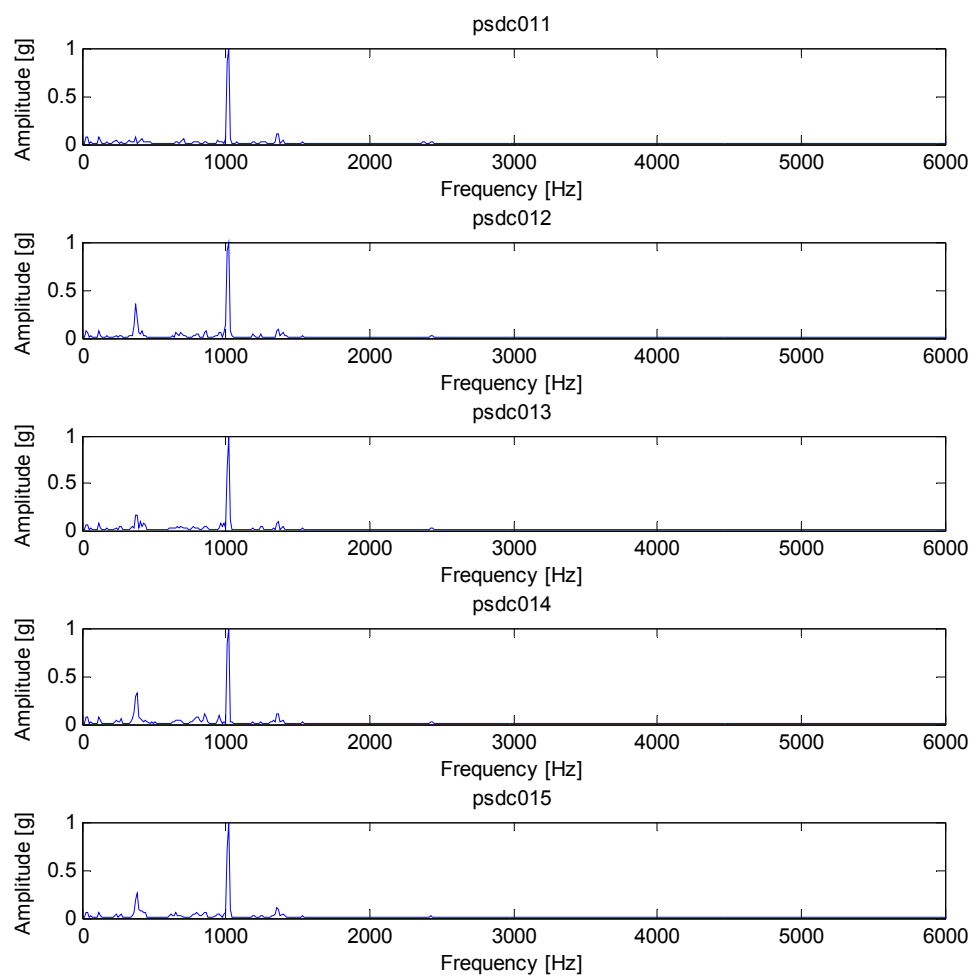


Figure B.3 : PSD11-PSD15

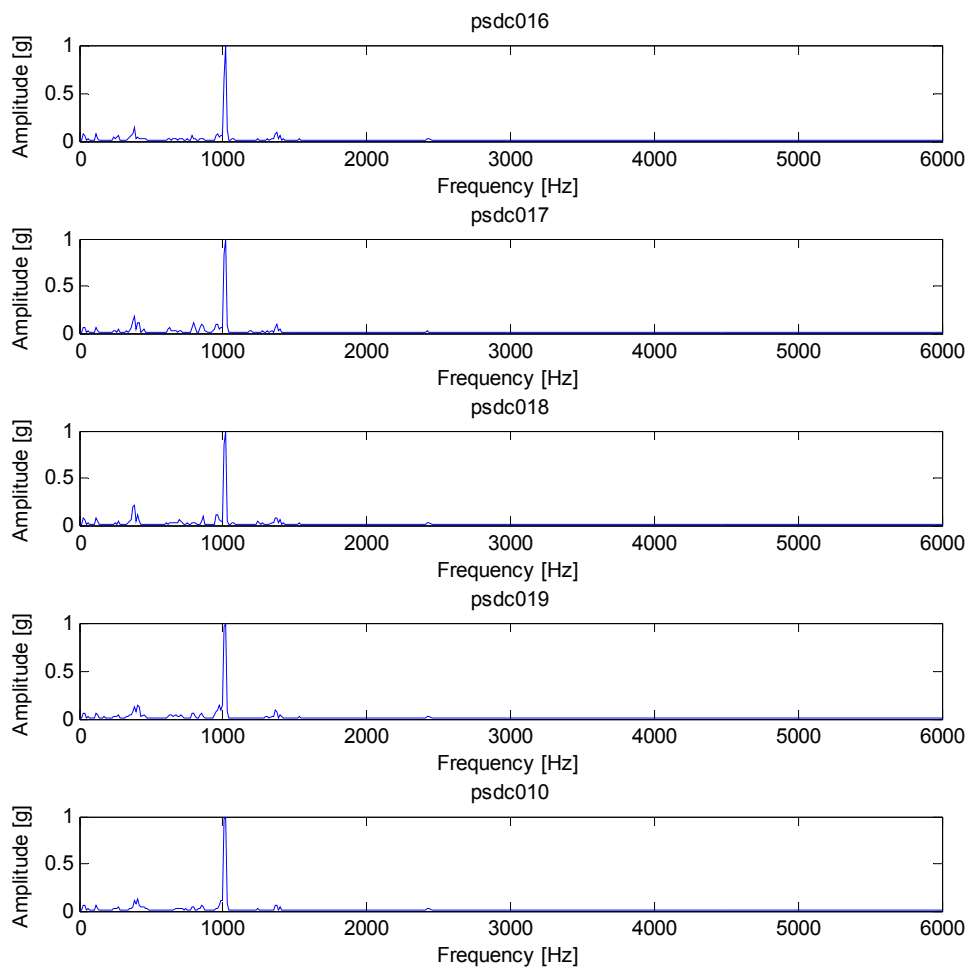


Figure B.4 : PSD16-PSD20

APPENDIX C: Training Set of the Simulation Signal - I (PSD of CWT)

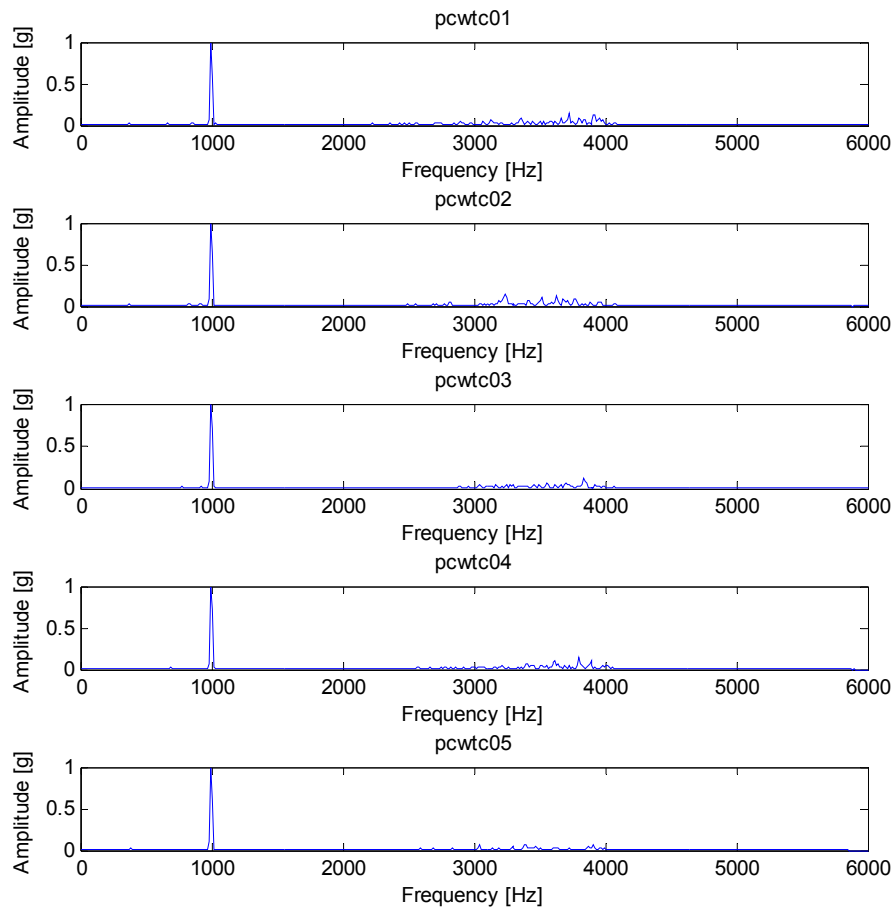


Figure C.1 : Pcwt1-Pcwt5

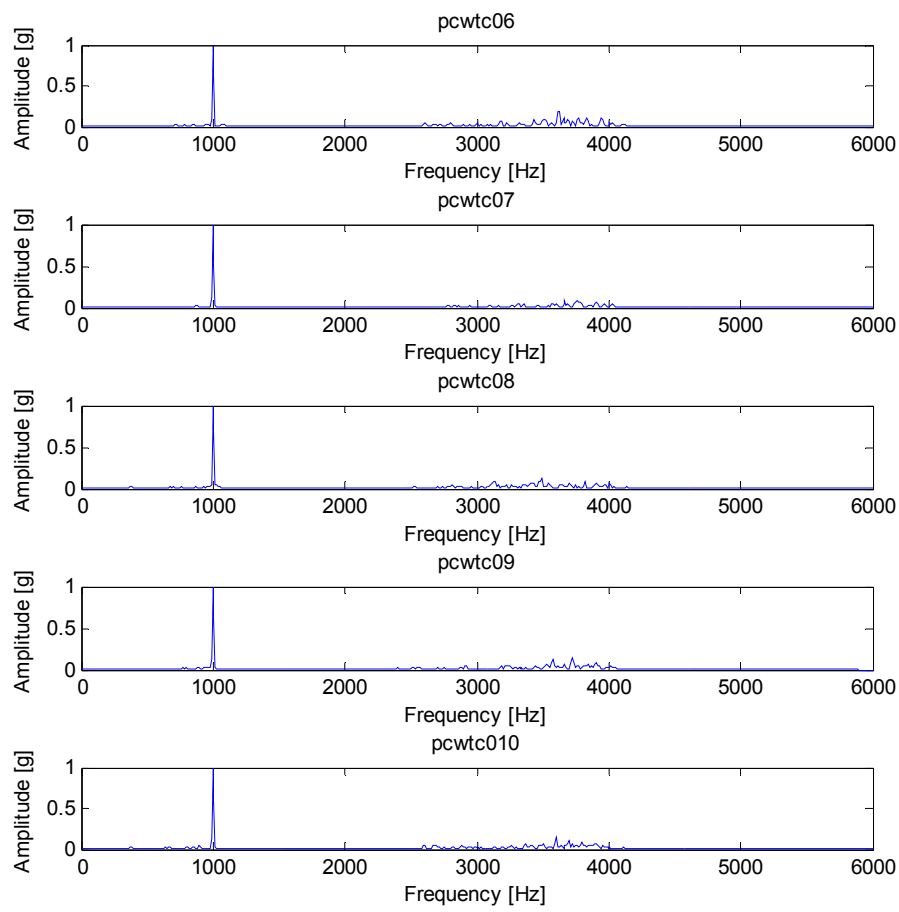


Figure C.2 : Pcwt6-Pcwt10

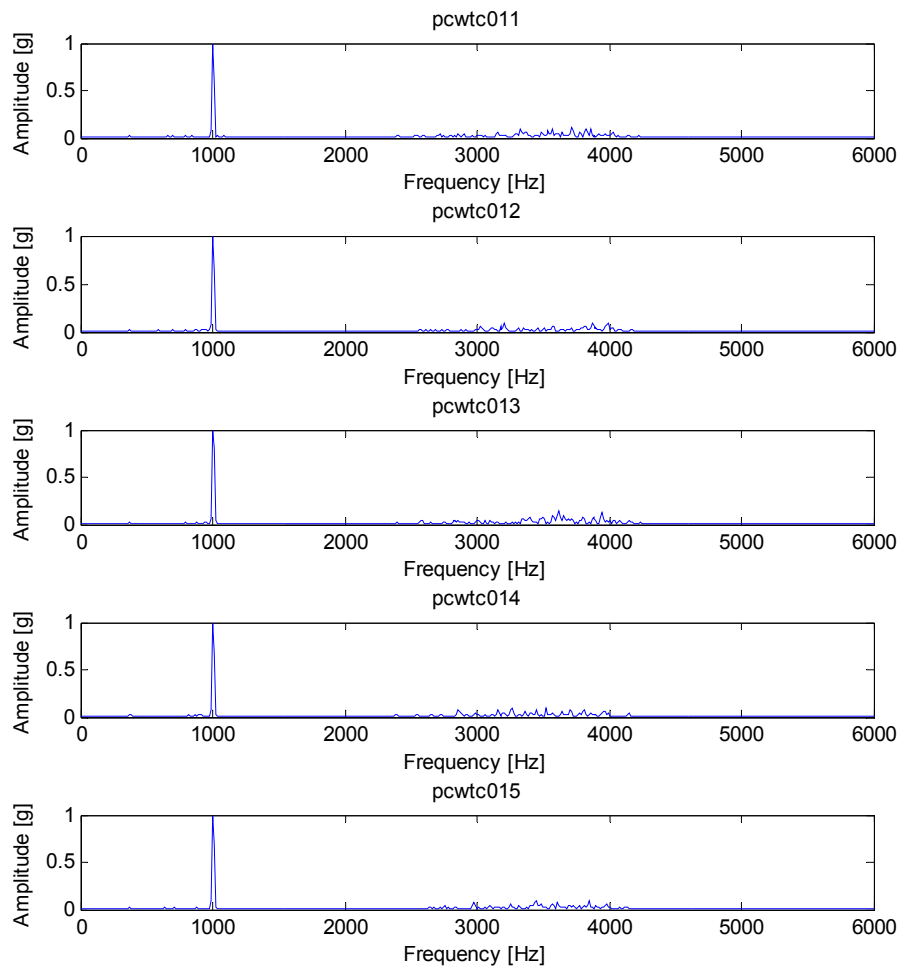


Figure C.3 : Pcwt11-Pcwt15

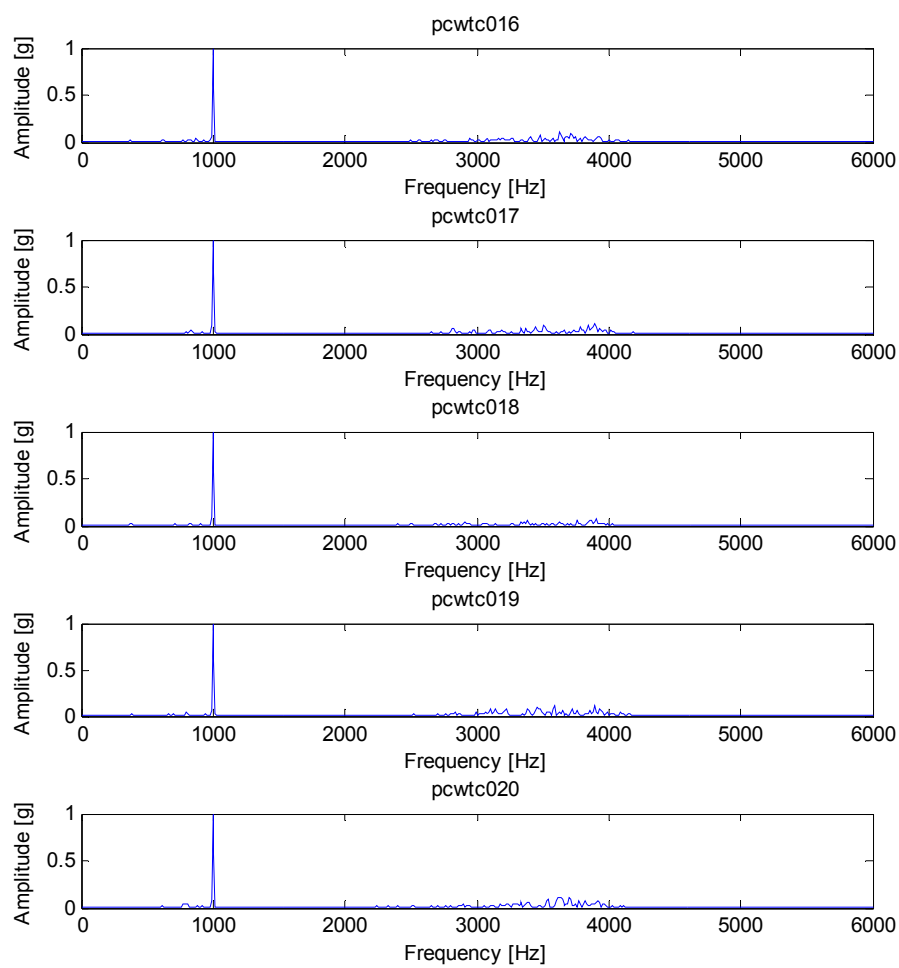


Figure C.4 : Pcwt16-Pcwt20

APPENDIX D: Test Set of the Simulation Signal - I (PSD)

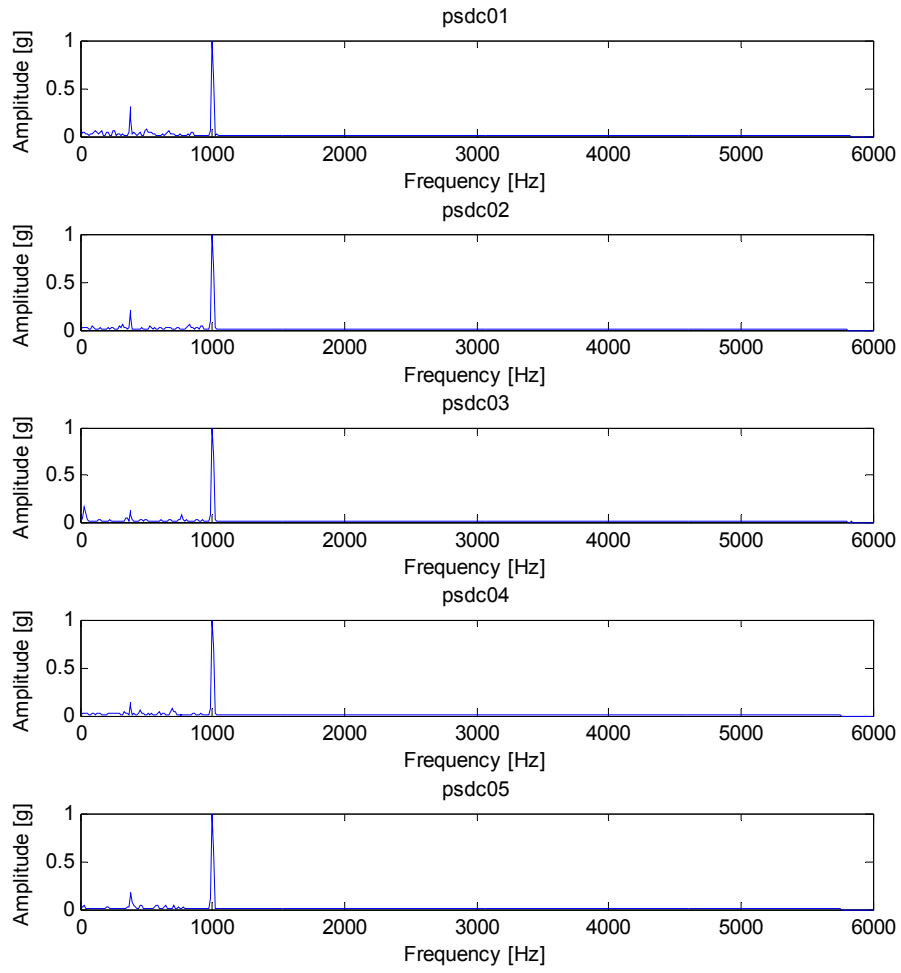


Figure D.1 : PSD1-PSD5

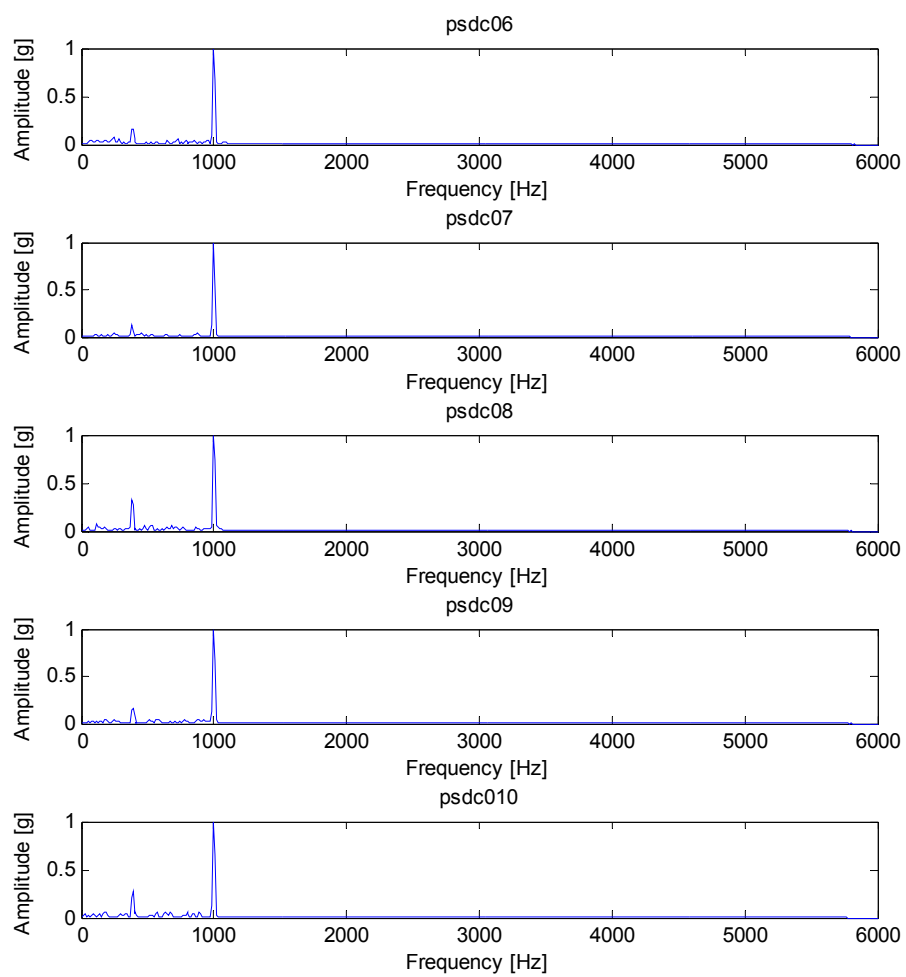


Figure D.2 : PSD6-PSD10

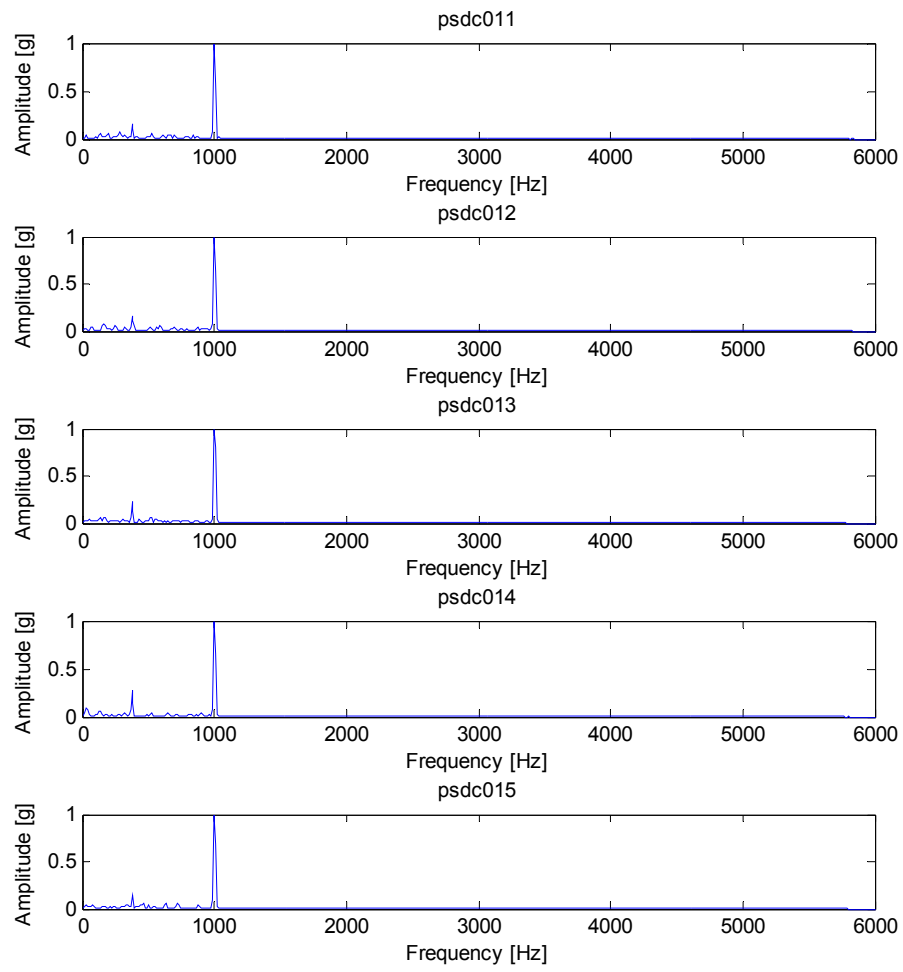


Figure D.3 : PSD11-PSD15

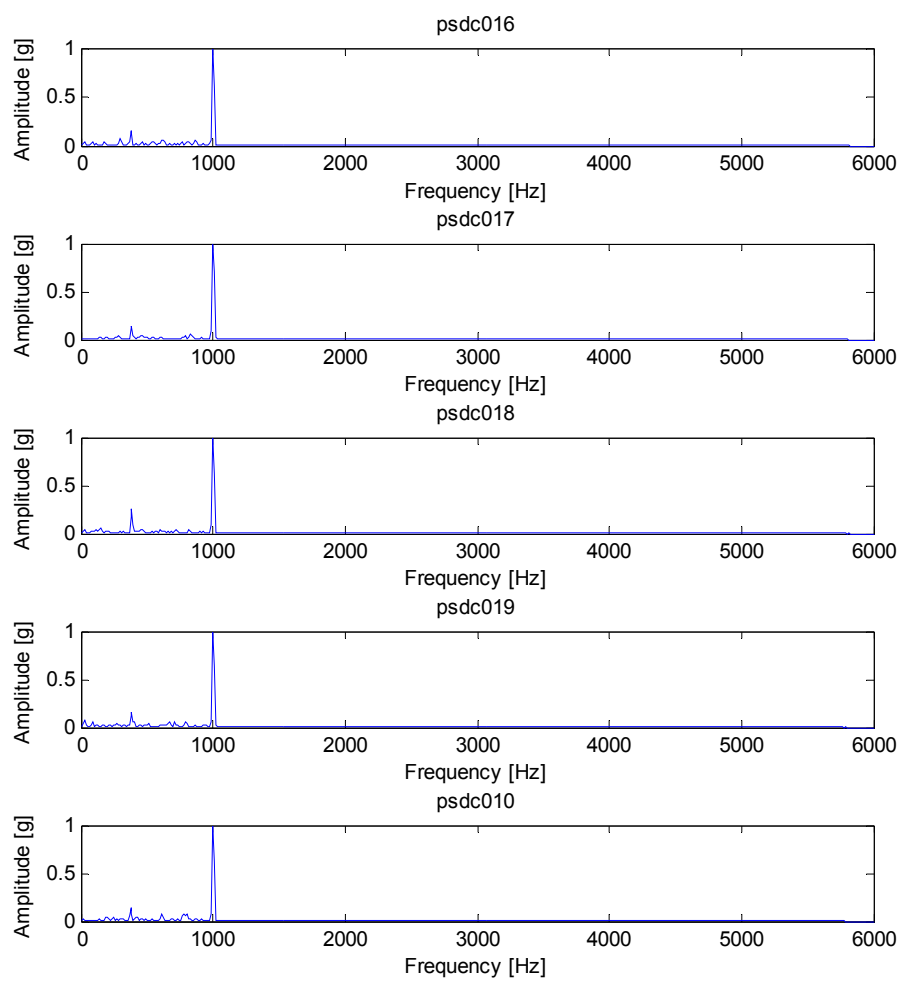


Figure D.4 : PSD16-PSD20

APPENDIX E: Training Set of the Simulation Signal - II (PSD of CWT)

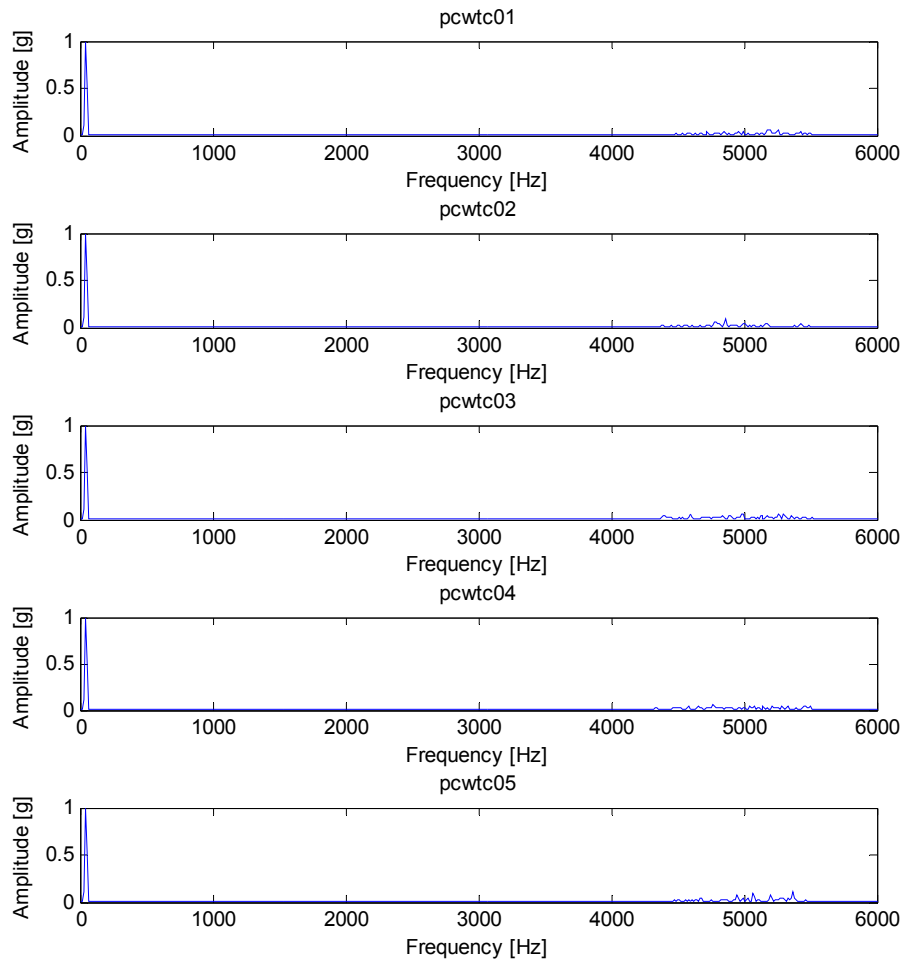


Figure E.1 : Pcwt1-Pcwt5

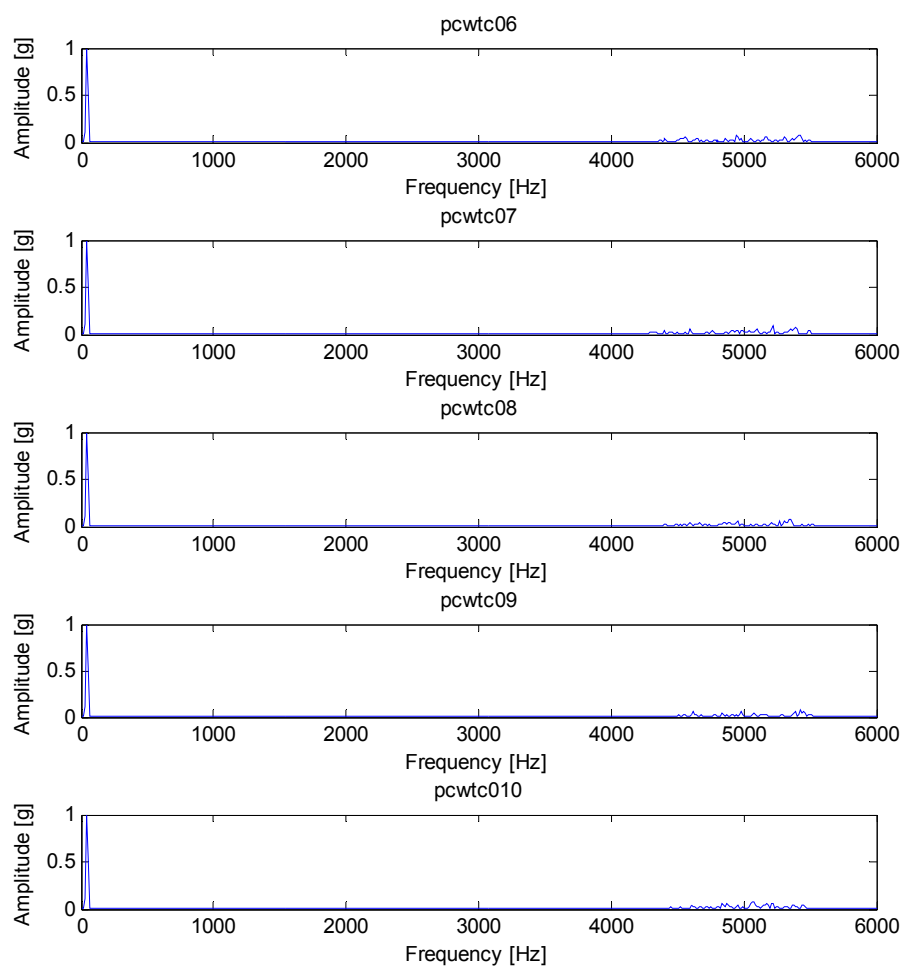


Figure E.2 : Pcwt6-Pcwt10

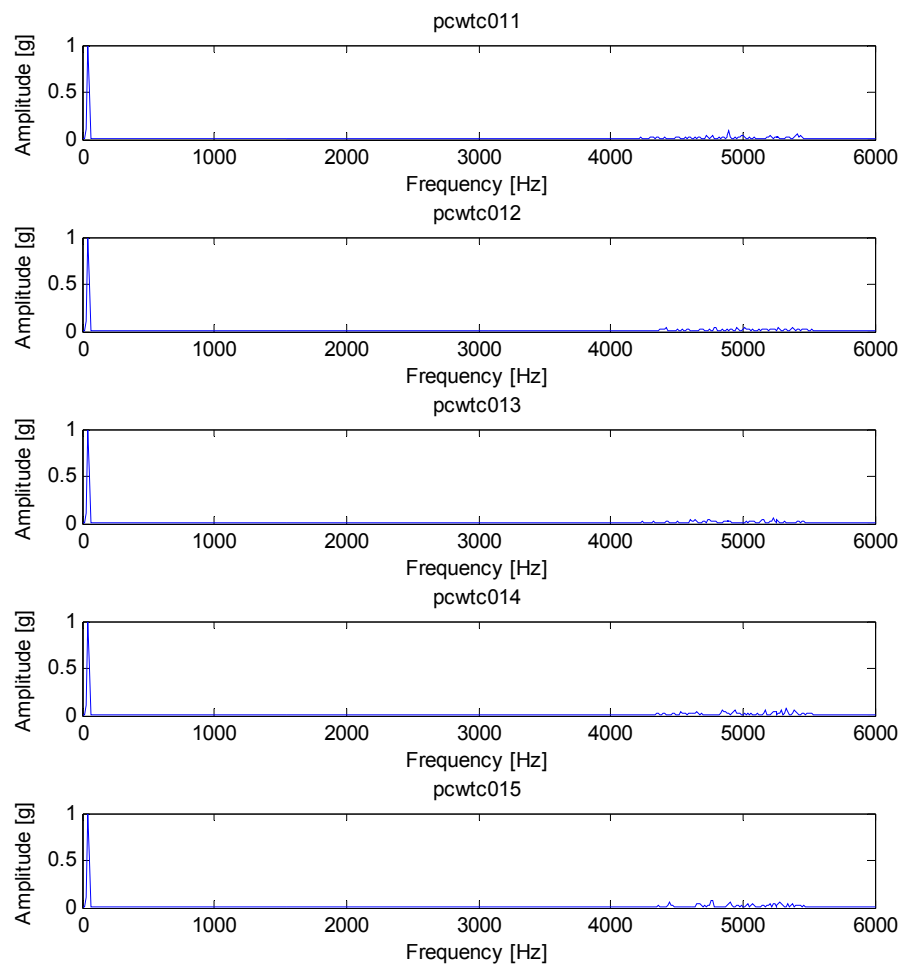


Figure E.3 : Pcwt11-Pcwt15

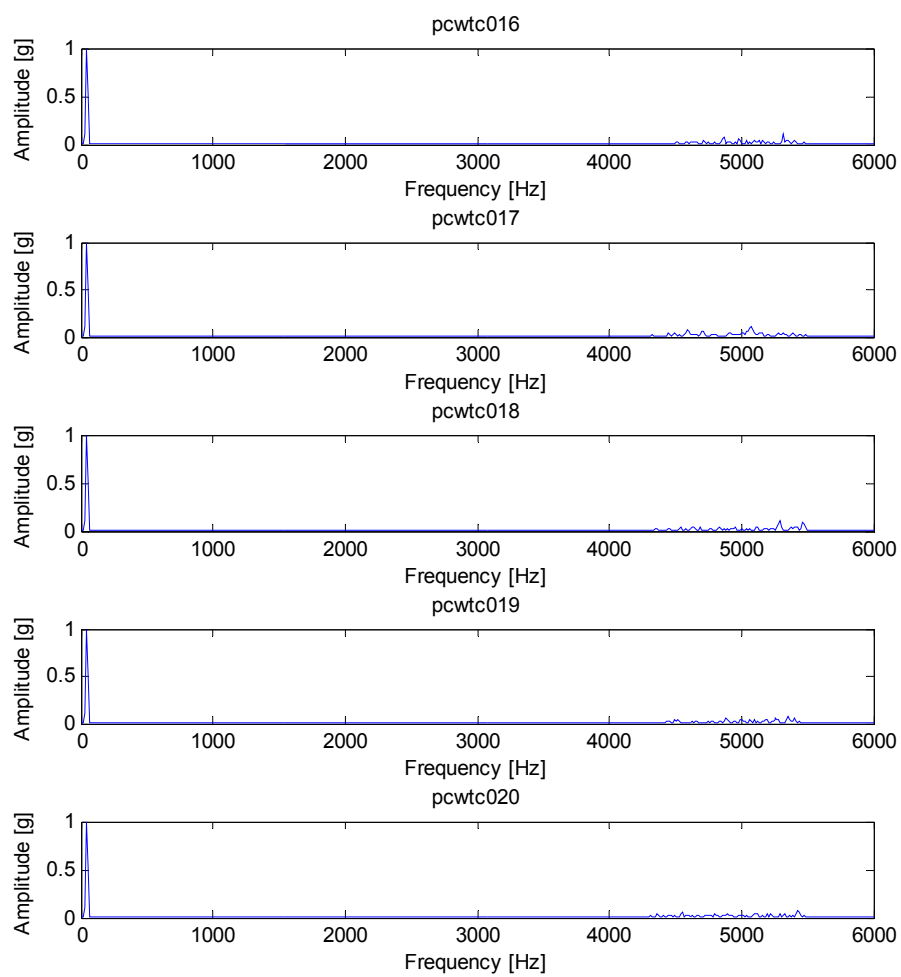


Figure E.4 : Pcwt16-Pcwt20

APPENDIX F: Test Set of the Simulation Signal - II (PSD)

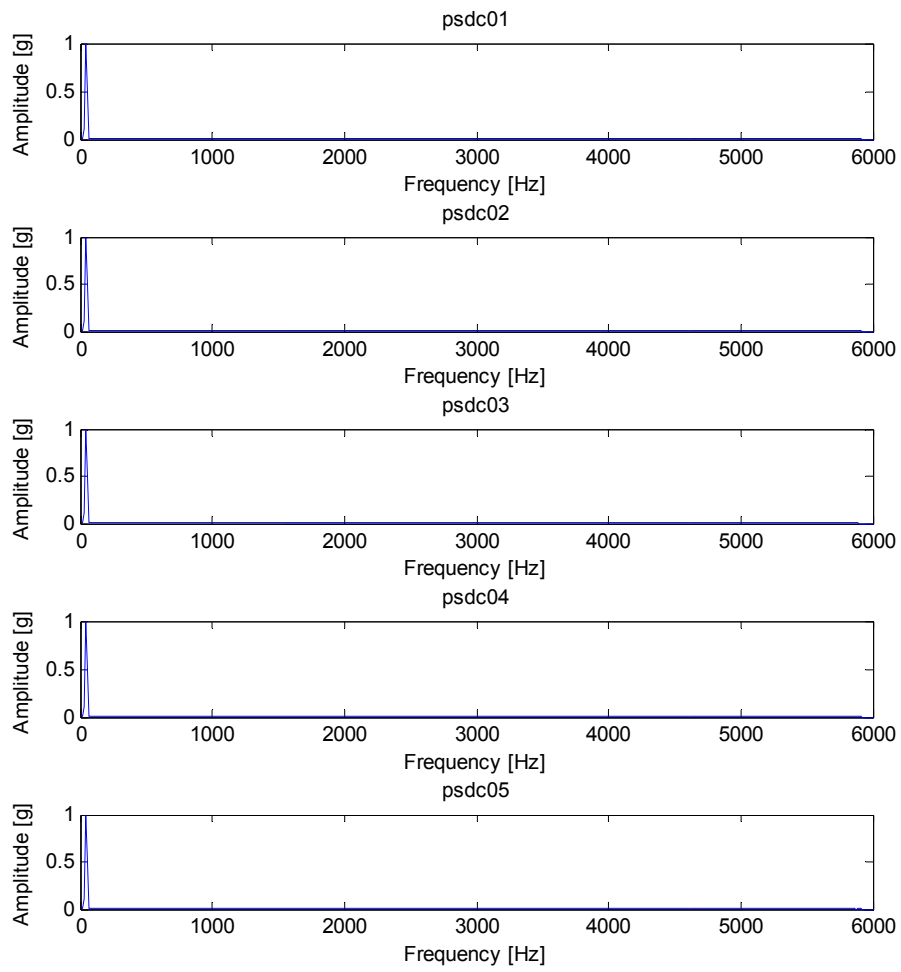


Figure F.1 : PSD1-PSD5

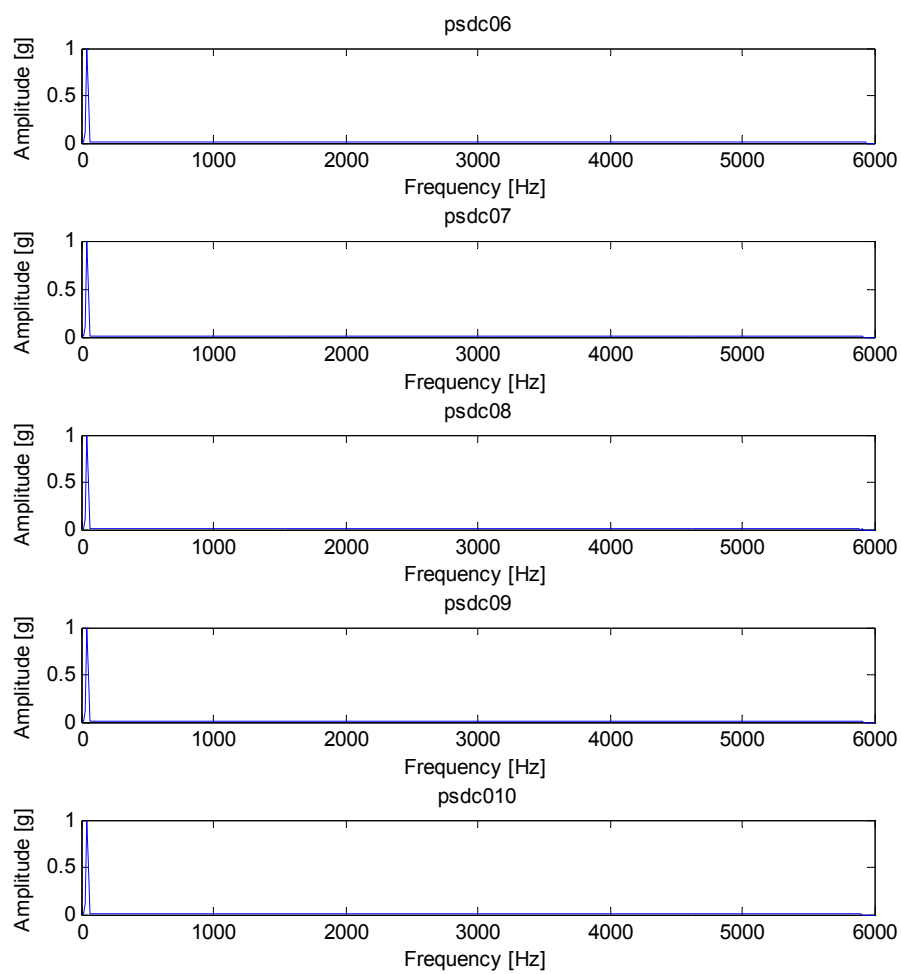


Figure F.2 : PSD6-PSD10

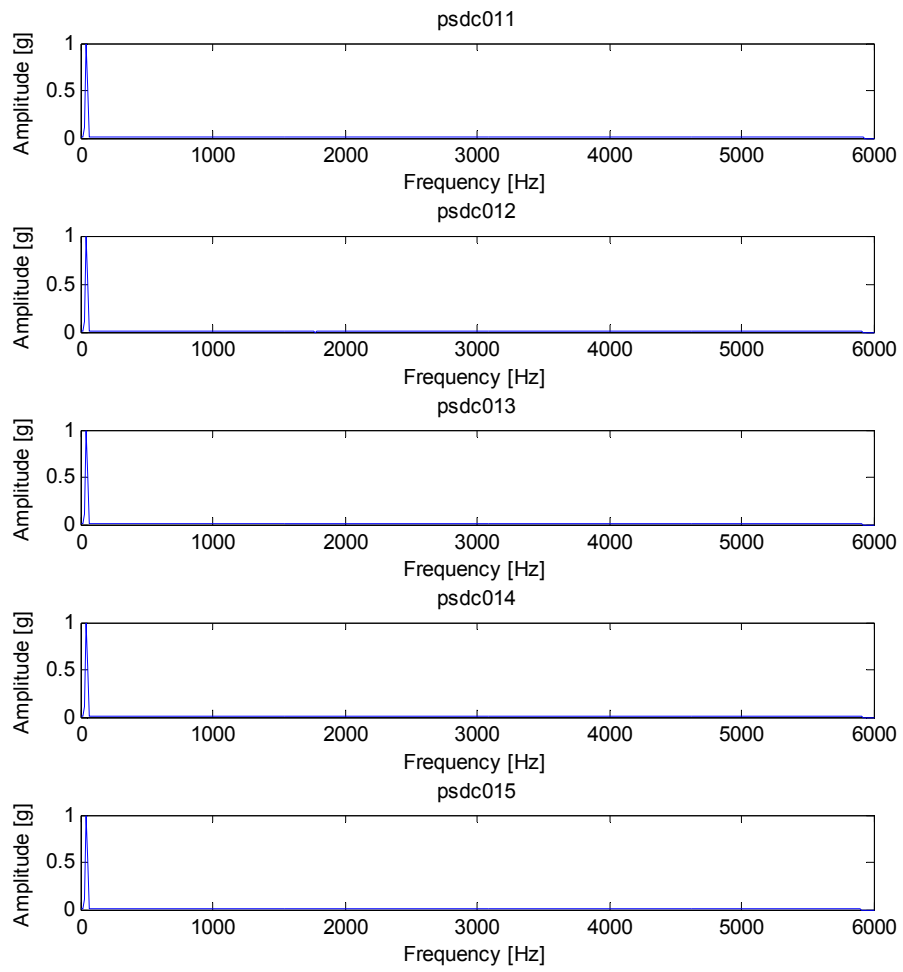


Figure F.3 : PSD11-PSD15

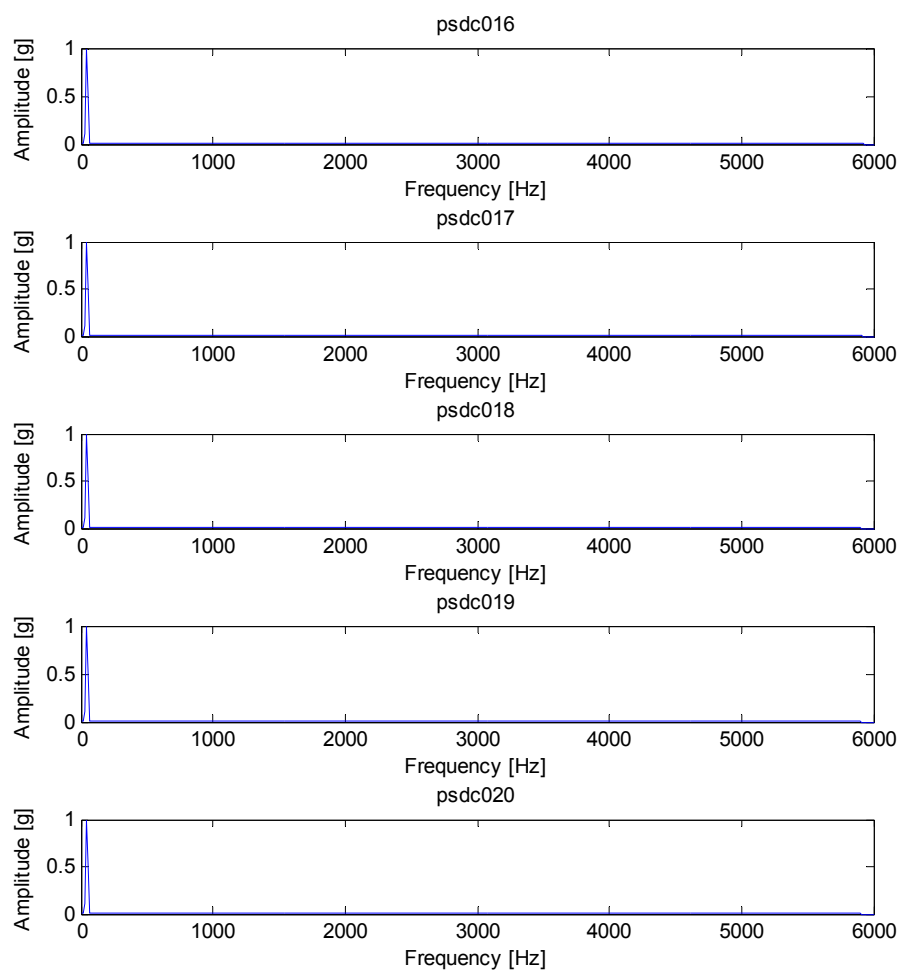


Figure F.4 : PSD16-PSD20

APPENDIX G: Graphical User Interface (GUI) Outputs

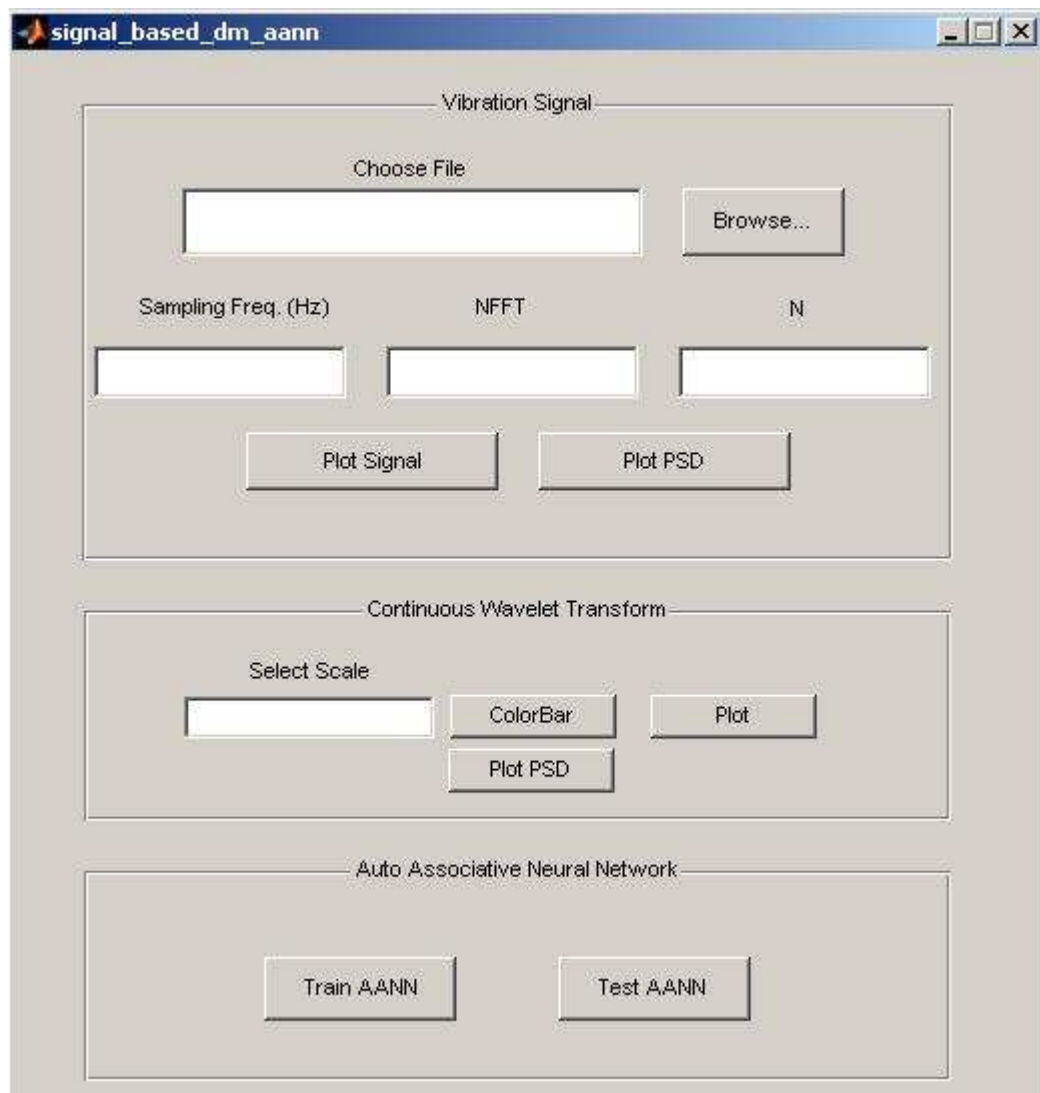


Figure G.1 : Graphical user interface of application

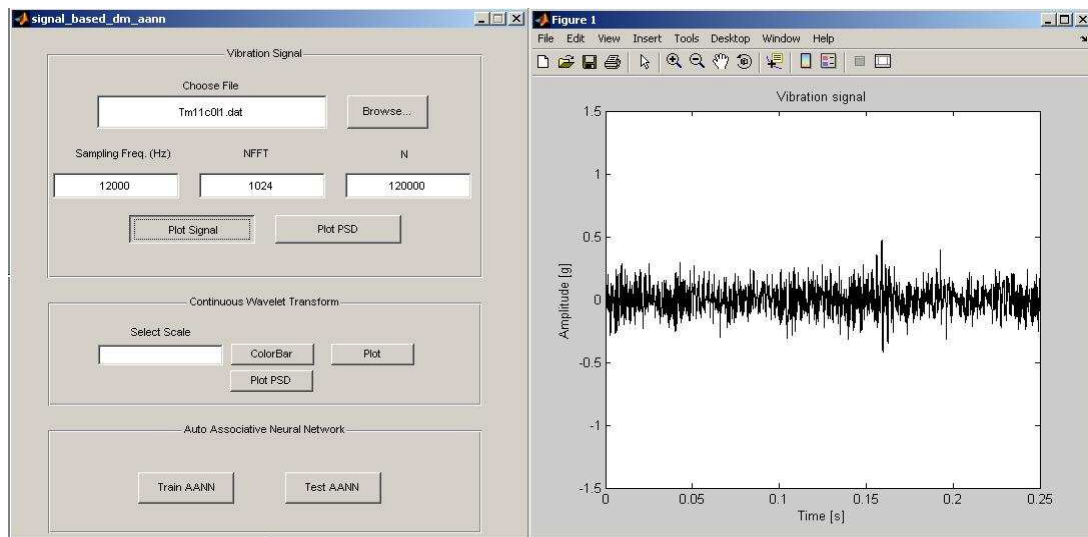


Figure G.2 : Vibration signal

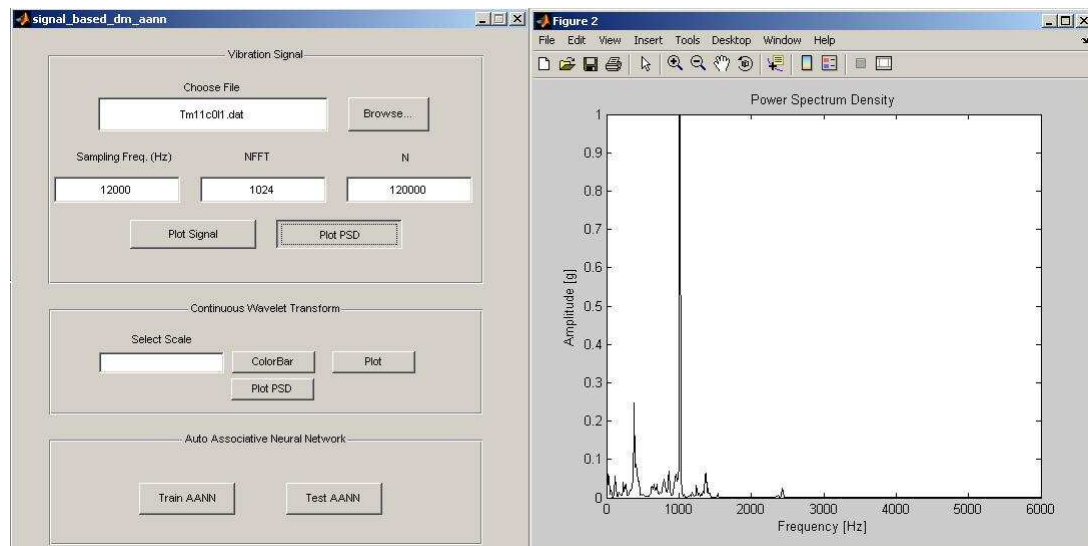


Figure G.3 : PSD of the vibration signal

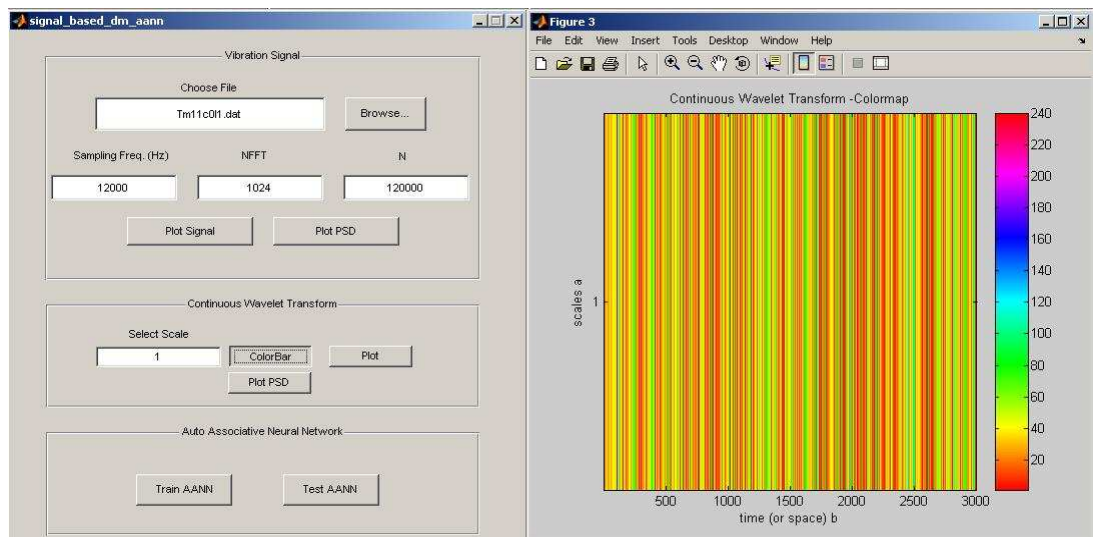


Figure G.4 : CWT at 1st scale

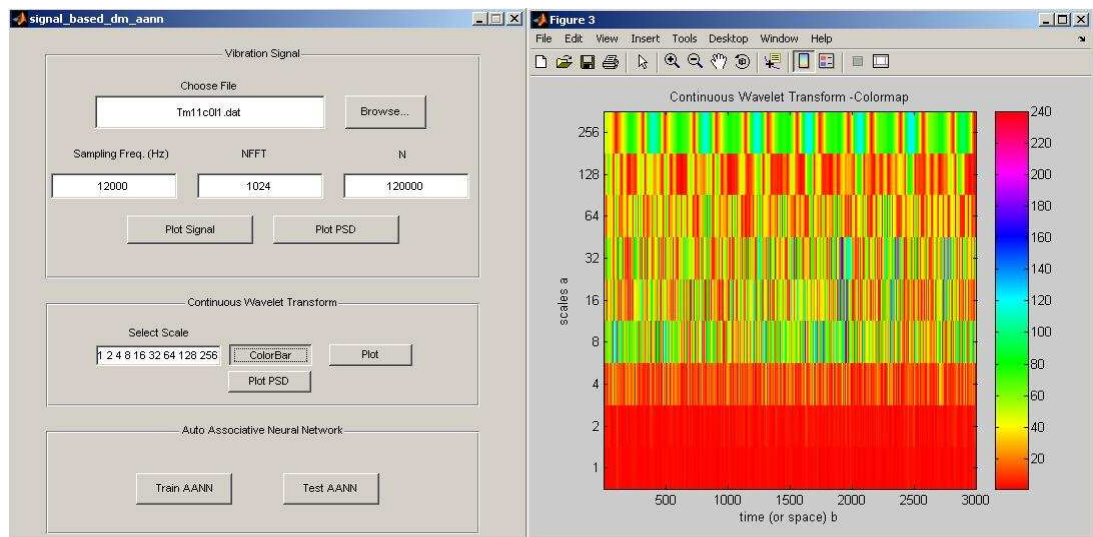


Figure G.5 : CWT at multiple scales

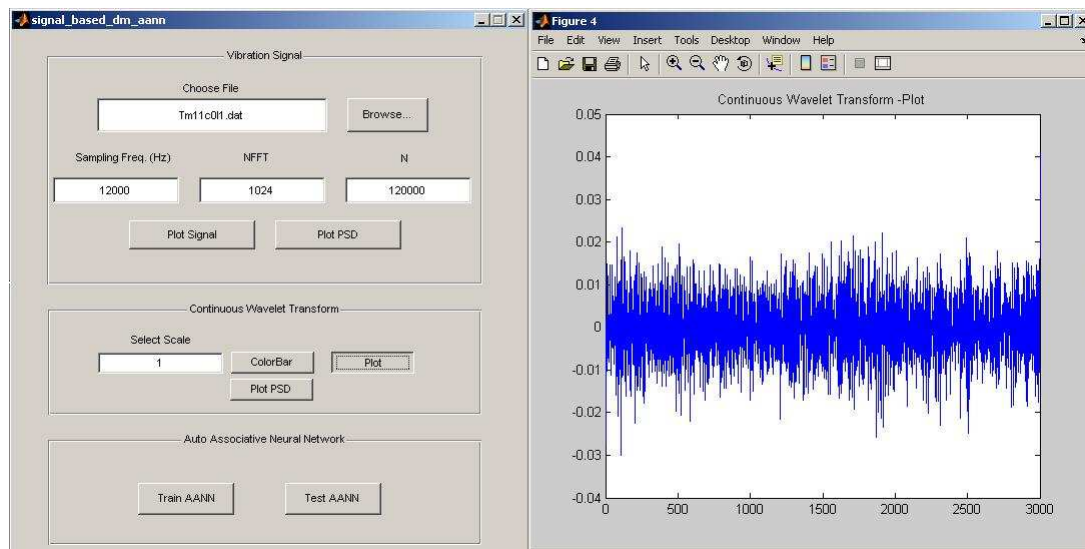


Figure G.6 : CWT at 1st scale

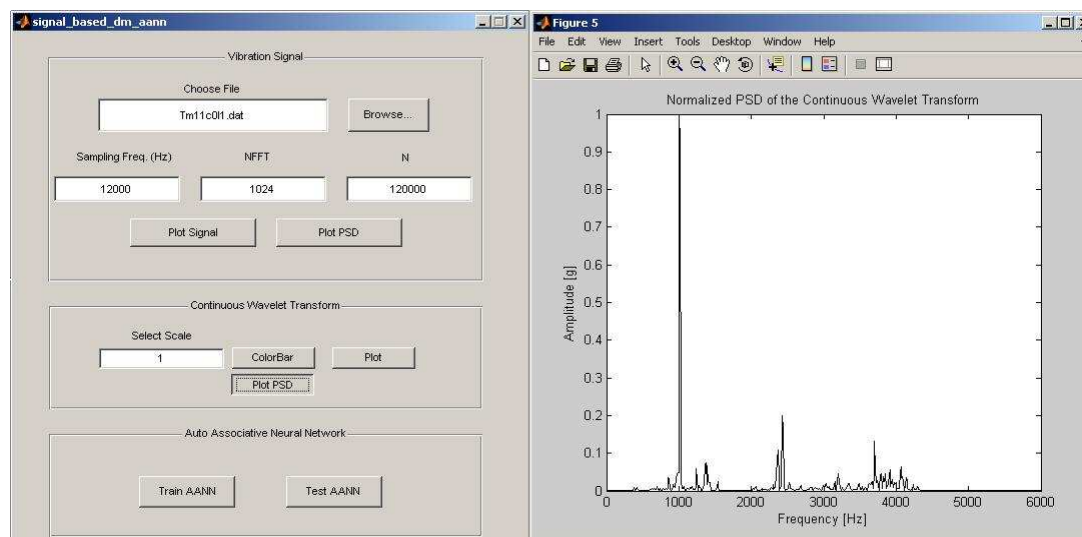


Figure G.7 : PSD of the CWT at 1st scale

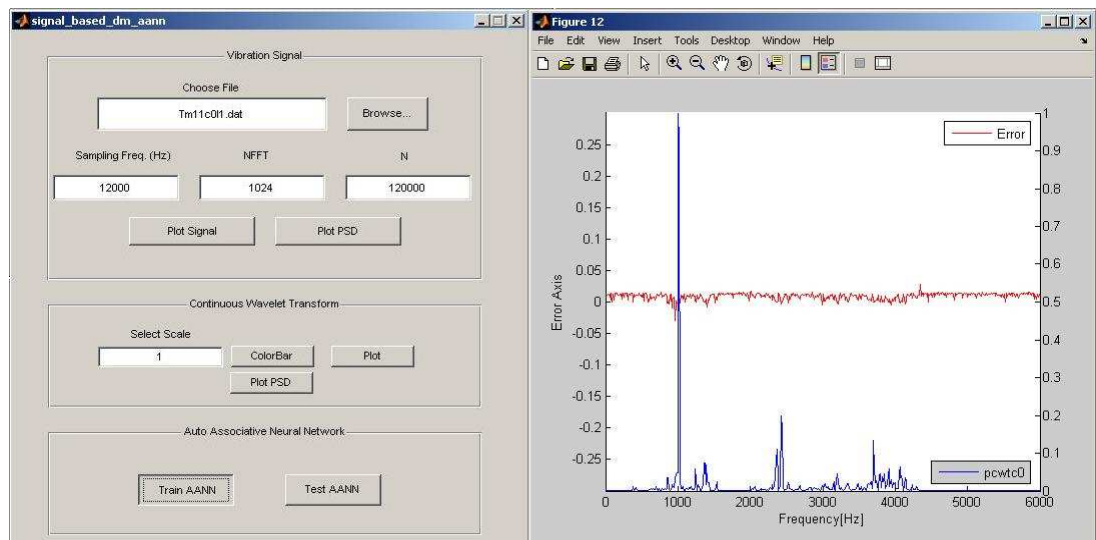


Figure G.8 : Training of the AANN

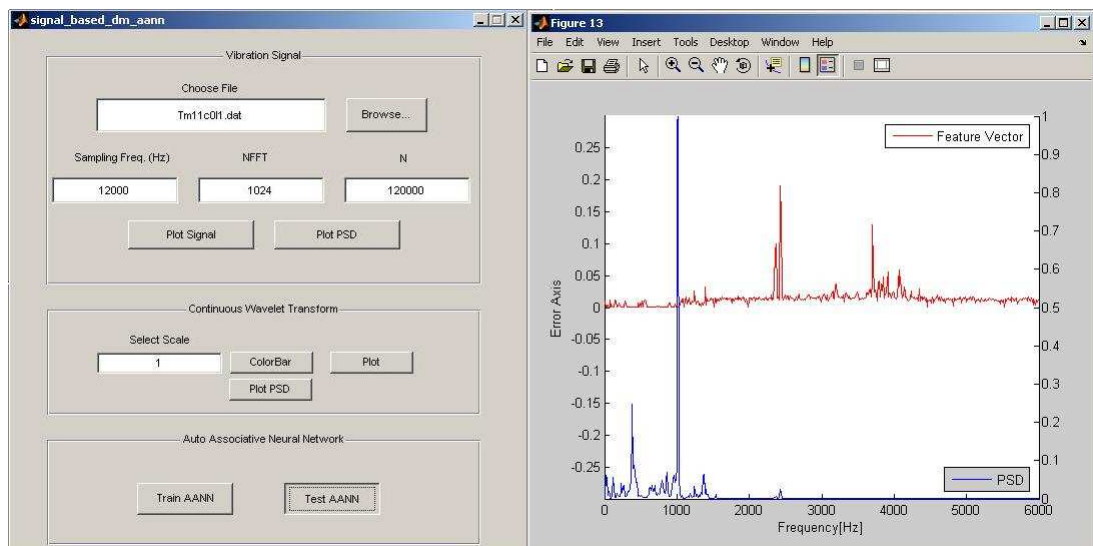


Figure G.9 : Test of the AANN

CURRICULUM VITAE



Name Surname: Selim GÜLLÜLÜ

Place and Date of Birth: Erzurum/ TURKEY, 1976

Address: Ataköy 9.Kısım Bakırköy/İstanbul

E-Mail: selim.gullulu@gmail.com

B.Sc.: Istanbul Technical University, Electronics and Telecommunications Engineering 1994-1998

M.Sc.: Istanbul Technical University, Information Technologies – Defence Technologies 1999-2003

Professional Experience and Rewards: Alcatel-Lucent Global Network Engineering and Integration Center Engineer

PUBLICATIONS/PRESENTATIONS ON THE THESIS

- **Güllülü S.**, Şeker S., 2011: Signal Based Approach for Data Mining in Fault Detection of Induction Motor. *Scientific Research Essays*, Vol. 6(22), 7 October, 2011, 4720-4731
- Şeker S., **Güllülü, S.**, Ayaz, E., 2008: Transfer Function approach based upon wavelet transform for bearing damage detection in electric motors. *IEEE International Symposium on Industrial Electronics*, June 30 - July 2, 2008 Cambridge, UK, 749-752
- Şengüler T., Karatoprak E., Ayaz E., **Güllülü, S.**, Şeker S., 2007: Entropy Approach Using PCA for Sequential Accelerated Aging Processes in Electrical Motors. *IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*, Sept 6-8, 2007 Cracow, Poland, 501-505

