

**DIAGNOSIS OF MULTIPLE INDEPENDENT AND
COEXISTING MECHANICAL AND HYDRAULIC FAULTS
IN CENTRIFUGAL PUMPS USING SUPPORT VECTOR
MACHINE BASED ALGORITHMS**

*A Thesis Submitted in Partial Fulfillment of the Requirements for
the Degree of*

DOCTOR OF PHILOSOPHY

by

RAPUR JANANI SHRUTI



**DEPARTMENT OF MECHANICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
GUWAHATI 781039, INDIA**

MAY 2019



**DIAGNOSIS OF MULTIPLE INDEPENDENT AND
COEXISTING MECHANICAL AND HYDRAULIC FAULTS
IN CENTRIFUGAL PUMPS USING SUPPORT VECTOR
MACHINE BASED ALGORITHMS**

*A Thesis Submitted in Partial Fulfillment of the Requirements for
the Degree of*

DOCTOR OF PHILOSOPHY

by

RAPUR JANANI SHRUTI

(Roll No. 146103018)



**DEPARTMENT OF MECHANICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
GUWAHATI 781039, INDIA**

MAY 2019





**Department of Mechanical Engineering
Indian Institute of Technology Guwahati
Guwahati-781039, INDIA**

CERTIFICATE

It is certified that the work contained in the thesis titled, **Diagnosis of Multiple Independent and Coexisting Mechanical and Hydraulic Faults in Centrifugal Pumps using Support Vector Machine Based Algorithms**, submitted by **Ms. Rapur Janani Shruti** (Roll. no. 146103018) to the Indian Institute of Technology Guwahati for the award of the degree of Doctor of Philosophy has been carried out under my supervision in the Department of Mechanical Engineering, Indian Institute of Technology Guwahati. This work has not been submitted elsewhere for the award of any other degree or diploma.

May 2019

Dr. Rajiv Tiwari

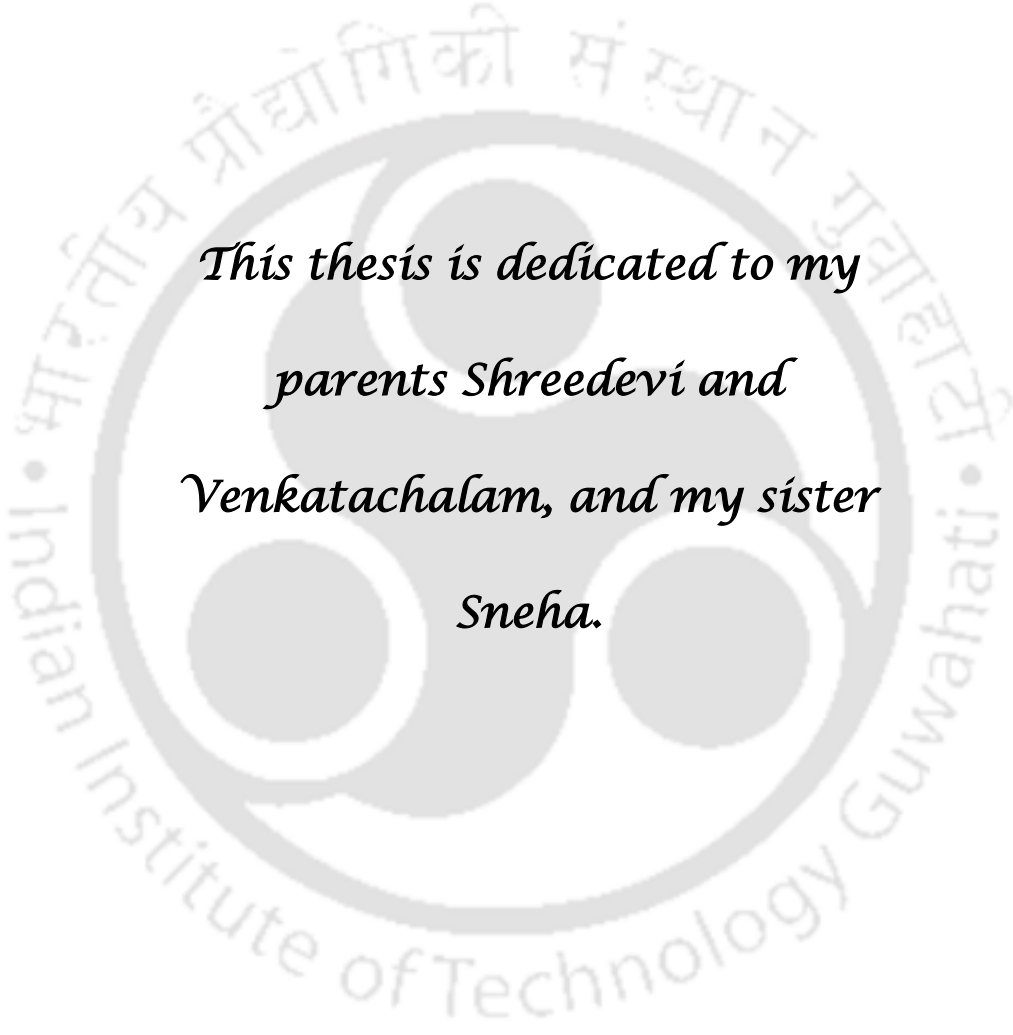
Professor

Department of Mechanical Engineering

Indian Institute of Technology Guwahati

Guwahati – 781039, INDIA





*This thesis is dedicated to my
parents Shreedevi and
Venkatachalam, and my sister
Sneha.*



Acknowledgment

*“Gratitude makes sense of our past, brings peace for today,
and creates a vision for tomorrow”*

At the outset, I would like to express a profound sense of gratitude to my thesis supervisor, Prof. Rajiv Tiwari, who left no stone unturned to pave the way for this research. He ignited the idea, stimulated my imagination, supported and guided me through the highs and lows of the study and most importantly, taught me many essential life lessons. His articulate work plans made the most complicated things look very simple many times. I always believed that – “Research is like a roller-coaster ride; it is always exciting but is full of surprises.” My Professor made this ride feel decidedly less tiring and more exciting. I place on record the time and efforts Prof. Tiwari put in to mold me into a better person, in general, and a focused researcher in particular. It will be an honor to collaborate with him in the future as well, to work on some more interesting problems in the area of ‘Rotor dynamics.’

Besides my advisor, I would like to thank the rest of my doctoral committee comprising of Prof. Santosha K. Dwivedy, Dr. Atanu Banerjee, and Prof. Chitrlekha Mahanta, not only for their insightful comments and encouragement but also for their questions which incited me to widen my research from various perspectives. I will be failing in my duties if I don’t thank Dr. Dhrubajyoti Bordoloi, who was instrumental in helping me smoothly conduct the various lab experiments. I would also like to thank all the anonymous reviewers for taking time off to go through my research and giving me thought-provoking ideas to better the understanding of the problem under consideration.

I am deeply indebted to the Department of Mechanical Engineering, IIT Guwahati for providing me with necessary infrastructure to carry out this research. The financial support they provided gave me an opportunity to attend many prestigious national and international conferences, where I could showcase my research to the veterans in the fields of Rotor dynamics, Vibrations and Condition monitoring from all over the world.

I would like to thank all my friends who have actively or passively supported me through this journey. I would like to specially mention the contribution of my lab mates: Dipendra bhayya, Purushottam, Nilakshi, Gyan, Gargi, and a few others, with whom I had interesting discussions, sleepless nights before deadlines, and lots of fun in the last four years. Apart from my lab mates, my other friends, Ruchika, Peeyushi, Trishna, Sunil, Indrani, helped me with my presentations, in proof-reading my technical documents and have remained as a constant source of encouragement.

My parents, Shreedevi, Venkatachalam, my husband Prithvi, my sister Sneha, and brother-in-law Nitin, stood as my pillars of support in the hardest of times. They made sure that I had the necessary emotional stability and supported me every day all through my journey. They believed in me and never let me lose my focus towards work. I am ever grateful for their enormous love and patience.

“Little drops of water make the mighty ocean,” I would like to thank the Almighty God for making me a channel to contribute to the might of this enormous ocean of knowledge.

March 2019

Rapur Janani Shruti

IIT Guwahati

Abstract

Reliable detection and isolation of centrifugal pump (CP) faults is a challenging and important task in the modern industries. Hence, this study proposes, machine learning based multi-fault detection of CPs driven by induction motor. The intelligent fault detection methodology is developed based on the multi-class support vector machine (SVM). The mechanical and hydraulic faults in CPs are mutually dependent and therefore may exist concurrently. Hence, in the present research, an assortment of various flow instabilities like, the suction flow blockages, discharge flow blockages, pseudo re-circulation and dry runs are considered coexisting with mechanical faults, like the impeller cracks and pitted cover plate faults. In addition, the suction and discharge CP blockages are considered with five levels of varying severity. A total of thirty-three faults are considered on the CP. In order to generate relevant fault signatures of CP faults, vibration signals in three orthogonal directions and current signals of all three phases have been acquired for a wide range of operating conditions (i.e., the speed) in time domain using an experimental test rig.

The fault diagnostics of CPs has been performed using features extracted independently from the time domain, frequency domain and time-frequency domains. The time-frequency domain data is obtained by transforming the time domain data using the continuous wavelet transform (CWT) and the wavelet packet transform (WPT) approaches. As the prediction accuracy highly depends on features, the wrapper model is used to select the most suitable features in each of the considered domains. These features are further used as inputs to SVM classifiers. The grid-search technique along with the cross-validation is adapted to select optimal SVM

parameters required to build a reliable SVM model. This model is used for fault predictions. The advantage of simultaneously using the vibration and current based features on the fault classification accuracy of mechanical and hydraulic CP faults has been established. Prediction performances have been checked with the low and high sampling rate of data acquisition using time domain and frequency domain data. If a defect exists at any stage the classifier identifies it, isolates it and establishes its severity. In the case of wavelet transform, several mother wavelets are considered to study the effects of them on the CP fault diagnosis.

The fault diagnosis is performed for wide range of distinct operating speeds of CPs. In order to check the robustness of the proposed methodology, the trained classifier is tested with features extracted from corrupted data. The data corruption is performed by adding white Gaussian noise at different signal to noise ratios to the baseline signals acquired from experimentation. The fault diagnosis in this study is also extended to the intermediate speed cases to check the classification ability of the developed method to account for unavailability of data at certain speeds of CP operation.

This investigation establishes that the vibration signatures are very useful and in most cases sufficient for the CP fault diagnosis. However, the addition of the features extracted from the motor line-current signatures improves the classification performance of the developed algorithm. All the methodologies developed based on different domains demonstrated that there is no or less operational speed dependence on the fault classification performance. However, for an intermediate speed case it is found that there is some operation speed dependence as the fault manifestation changes significantly with the CP speed. The

identification of fault severity is much more complicated than identifying faults belonging to different fault families. This is because faults that belong to same family with slight variations in severity produce similar fault signatures and therefore similar fault features. The developed methods also show robustness in performance even when the data is corrupted with additive noise.





Table of Contents

Acknowledgment	i
Abstract	iii
Table of Contents	vii
List of Figures	xiii
List of Tables	xvii
Nomenclature	xxi
Abbreviations	xxiii
 CHAPTER 1 Introduction and Literature Review	 1
1.1 Introduction.....	1
1.2 Working of CPs.....	2
1.3 Performance of CPs.....	3
1.3.1 Energy Conservation.....	4
1.3.2 CP Performance Parameters.....	4
1.4 CP Applications.....	9
1.5 CP Failure Modes.....	10
1.5.1 Mechanical Vibration Factors.....	10
1.5.2 Hydraulic Vibration Factors.....	14
1.5.3 Motor Factors.....	16
1.5.4 Degradation of Pumping Systems.....	17
1.6 Consolidated List of CP Problems and Their Causes.....	17
1.7 Condition Based Monitoring.....	17
1.8 Philosophy of CBM.....	22
1.9 Signals Used for CBM Analysis.....	22
1.10 Pre-Processing of Signals.....	23
1.11 Methods of CBM Analysis.....	26
1.12 CP Fault Diagnosis Using CBM Techniques.....	26
1.13 Visual Inspection of Sensor Data.....	26

1.13.1 Visual Inspection Using Spectral Studies.....	27
1.13.2 Visual Inspection Using Statistical Methods.....	36
1.13.3 Conclusions - Visual Inspection	37
1.14 Mathematical Modeling.....	38
1.14.1 Mathematical Modeling for CP Fault Detection.....	41
1.14.2 Conclusions – Mathematical Modeling	45
1.15 Machine Learning Techniques.....	45
1.15.1 Logic Based Classifiers.....	47
1.15.2 Neural Network Based Classifier.....	51
1.15.3 Statistical Learning Based Classifier.....	54
1.15.4 Support Vector Machine Classifier.....	56
1.15.5 Conclusions – Machine Learning	60
1.16 Outcomes of the Literature Review.....	67
1.17 Aims and Objectives of the Current Research.....	70
1.18 Organization of Thesis.....	72
CHAPTER 2 Laboratory Rig and Experimental Procedure for Centrifugal Pump Faults.....	73
2.1 Introduction.....	73
2.2 Laboratory Rig and Experimentation.....	73
2.2.1 Machine Fault Simulator.....	74
2.2.2 Measurement Sensors.....	79
2.2.3 Faults Considered on the CP.....	82
2.3 Experimental Procedure.....	88
2.3.1 CP Mounting and Assembly.....	88
2.3.2 Data Collection.....	90
2.4 Experimental Observations.....	94
2.4.1 Interpretation of the Experimental Observations.....	95
2.5 Experimentation Phases.....	98
2.5.1 Experimentation - Phase One.....	98
2.5.2 Experimentation - Phase Two.....	98

2.5.3 Experimentation - Phase Three.....	99
2.6 Summary.....	101
CHAPTER 3 Centrifugal Pump Fault Classification Methodology based on SVM.....	103
3.1 Introduction.....	103
3.2 Support Vector Machine Classifier.....	103
3.3 Multiclass Support Vector Machine Classifier.....	110
3.4 C-SVM: Soft Margin Approach.....	113
3.5 CV Technique and GST for Optimal Parameter Selection...	116
3.6 Wrapper Model for Feature Selection.....	118
3.7 Fault Diagnosis Procedure for the Current Study.....	120
3.7.1 Noise Corruption Analysis.....	122
3.8 Summary.....	126
CHAPTER 4 Preliminary Time, Frequency Domain based Centrifugal Pump Fault Diagnosis.....	127
4.1 Introduction.....	127
4.2 Phase I Experimentation – Time Domain Analysis.....	128
4.2.1 Fault Feature Extraction and Selection.....	128
4.2.2 Binary Classifications of Flow Blockages.....	135
4.2.3 Multiclass Classification of Blockages.....	136
4.2.4 Binary Classification of Impending Cavitation (C0)...	136
4.2.5 Conclusions: Phase I Time Domain Results.....	139
4.3 Phase I Experimentation – Frequency Domain Analysis.....	139
4.3.1 Fault Feature Extraction and Selection.....	140
4.3.2 Binary Classifications of Flow Blockages.....	140
4.3.3 Multiclass Classifications.....	141
4.3.4 Conclusions: Phase I Frequency Domain Results.....	143
4.4 Phase-II Experimentation – Time Domain Analysis.....	144
4.4.1 Phase II Feature Selection.....	144

4.4.2 Binary Classification - Case A, Case B and Case C.....	147
4.4.3 Prediction Performance - IF and IFSB - Case D.....	148
4.4.4 Prediction Performance - HP, SB, IF, IFSB – Case E	148
4.4.5 Multiclass Classification – Case F	149
4.4.6 Conclusions: Phase-II Time Domain Results.....	152
4.5 Summary.....	154
CHAPTER 5 Multi-Fault Diagnosis of Centrifugal Pumps in Time	
Domain using Data Acquired at Low Sampling Rate.....	157
5.1 Introduction.....	157
5.2 Fault Feature Extraction.....	158
5.3 Fault Feature Selection.....	161
5.3.1 Data Resolution Effect.....	162
5.4 Fault Diagnosis Using the SVM.....	166
5.4.1 Case A:Same Speed Classification.....	168
5.4.2 Case B:Fault Identification at Intermediate Test Speed	176
5.4.3 Case C:Fault Identification with Noisy Test Data	178
5.5 Comparison of the results of the methodology with other works in literature.....	183
5.6 Summary.....	185
CHAPTER 6 Multi-Fault Diagnosis of Centrifugal Pumps in Frequency	
Domain using Data Acquired at Low Sampling Rate.....	187
6.1 Introduction.....	187
6.2 Fault Feature Extraction.....	188
6.3 Fault Feature Selection.....	191
6.4 Fault Diagnosis Using the SVM.....	194
6.4.1 Case A:Same Speed Classification.....	195
6.4.2 Case B:Fault Identification at Intermediate Test Speed	203
6.4.3 Case C: Fault Identification with Noisy Test Data.....	205
6.5 Comparison of the results in frequency-domain with other similar work in literature.....	210

6.6 Summary.....	213
CHAPTER 7 Multi-Fault Diagnosis of Centrifugal Pumps using Continuous Wavelet Transform Analysis.....	217
7.1 Introduction.....	217
7.2 Fault Feature Extraction.....	218
7.2.1 Continuous Wavelet Transform.....	218
7.2.2 Scale Selection and Feature Extraction.....	221
7.2.3 PCA Based Feature Selection.....	223
7.2.4 Feature and Mother Wavelet Selection.....	225
7.3 SVM Based Fault Prediction Performance.....	228
7.3.1 Case A: Same Speed Classification.....	228
7.3.2 Case B: Intermediate Speed Fault Classification.....	239
7.3.3 Case C: Fault Identification with Noisy Test Data....	241
7.4 Summary.....	243
CHAPTER 8 Multi-Fault Diagnosis of Centrifugal Pumps using Wavelet Packet Transform Analysis.....	245
8.1 Introduction.....	245
8.2 Fault Feature Extraction.....	246
8.2.1 Wavelet Packet Transform.....	246
8.2.2 Scale Selection and Feature Extraction.....	248
8.2.3 PCA Based Feature Selection.....	251
8.2.4 Feature and Mother Wavelet Selection.....	254
8.3 SVM Based Fault Prediction Performance.....	255
8.3.1 Case A: Same Speed Classification.....	256
8.3.2 Case B: Intermediate Speed Classification.....	266
8.3.3 Case C: Fault Identification with Noisy Test Data ...	268
8.4 Comparison of the developed wavelet based methodologies to similar work from literature.....	270
8.5 Summary.....	273

CHAPTER 9	Multi-Fault Diagnosis of Centrifugal Pumps: A	
	Comparative Analysis.....	275
	9.1 Introduction.....	275
	9.2 Feature(S) Used For the Fault Classification.....	275
	9.3 CP Fault Classification Performance in Various Domains....	278
	9.3.1 BF Fault Family.....	278
	9.3.2 IFBF Fault Family.....	279
	9.3.3 PCBF Fault Family.....	282
	9.3.4 Dry Run Fault Family.....	283
	9.3.5 AF Fault Family.....	283
	9.3.6 Overall Comparison of Fault Classification Cases....	290
	9.4 Summary.....	292
CHAPTER 10	Concluding Remarks.....	295
	10.1 Introduction.....	295
	10.2 Major Conclusions from the Current Work.....	296
	10.3 Main Contributions of the Thesis Work.....	299
	10.4 Recommendations from the Current Work.....	300
	10.5 Limitations of the Thesis Work.....	302
	10.6 Scope for Future Work.....	302
	Appendix A: Centrifugal Pump Applications.....	305
	Appendix B: Time-domain plots of various CP faults high sampling rate...	307
	Appendix C: Time-domain plots of various CP faults low sampling rate...	313
	Appendix D: PSD plots of various CP faults high sampling rate.....	319
	Appendix E: PSD plots of various CP faults high sampling rate.....	325
	Appendix F: CWT plots of various CP faults low sampling rate.....	331
	Appendix G: WPT plots of various CP faults low sampling rate.....	335
	Appendix H: Feature calculation for low sampling rate data.....	339
	References.....	341
	Publications from the Present Research.....	359

List of Figures

Figure 1.1	The section (left) and plan (right) views of an impeller, LE: Leading edge, TE: Trailing edge.....	3
Figure 1.2	Streamline and incompressible fluid flow across a pipe.....	5
Figure 1.3	Schematic diagram of CP working in a typical industrial application	6
Figure 1.4	Unbalance in a rotor, (a) Balanced rotor, (b) unbalanced rotor.....	12
Figure 1.5	Misalignment in a rotor (a) correct alignment, (b) parallel misalignment, (c) angular misalignment, (d) combined angular and parallel misalignment.....	13
Figure 1.6	Plant fault maintenance strategies.....	21
Figure 1.7	Flowchart on working of CBM.....	23
Figure 1.8	CP sensor locations for acquiring different signals.....	24
Figure 1.9	Axial vibration power spectrum of a healthy pump at 30 Hz operational speed.....	30
Figure 1.10	Axial vibration power spectrum of a clogged pump at 30 Hz operational speed.....	30
Figure 1.11	Axial vibration power spectrum of a pump with cracked impeller at 30 Hz operational speed.....	31
Figure 1.12	Axial vibration power spectrum of a clogged pump with cracked impeller at 30 Hz operational speed.....	31
Figure 1.13	Vertical transverse vibration power spectrum of a healthy pump at 30 Hz operational speed.....	32
Figure 1.14	Vertical transverse vibration power spectrum of a pump with cracked impeller at 30 Hz operational speed.....	32
Figure 1.15	Flow chart of a model based fault detection system.....	40
Figure 1.16	Structure of a decision tree algorithm.....	48
Figure 1.17	Rough set representation.....	49
Figure 1.18	ANN architecture.....	52
Figure 1.19	SVM classifier.....	56
Figure 2.1	CP fault diagnosis experimental set-up.....	74
Figure 2.2	A schematic diagram of the experimental set-up.....	76
Figure 2.3	(a) LabVIEW environment, (b) samples raw data plot.....	78
Figure 2.4	Accelerometers mounting position.....	82
Figure 2.5	Current probes measuring motor line current.....	82
Figure 2.6	(a) Mechanical modulating valve with markings; (b) cracks on the impeller; (c) pitted cover plate.....	84
Figure 2.7	Line diagram showing different blockage levels.....	84
Figure 2.8	Assembly drawing with dimensions (all dimensions in mm).....	90
Figure 2.9	Pump mounted on the MFS with a poly-carbonate transparent cover	91
Figure 2.10	Raw power spectral density data in linear scale from the accelerometer-1 for HP fault at 30 Hz with (a) high frequency resolution data sampling and (b) low frequency resolution data sampling.....	92

Figure 2.11	(a) Flow pattern over healthy impeller blades; (b) flow pattern over cracked impeller blades.....	96
Figure 2.12	Stages of material erosion due to bubble formation (a) Collapse of bubble cloud, (b) pressure wave emission, (c) micro-jet formation, (d) pit formation and material erosion.....	97
Figure 2.13	CP vicious fault cycle.....	97
Figure 3.1	(a) Various orientations of the data separating planes; (b) the optimal orientation of the SVM hyperplane.....	105
Figure 3.2	(a) Intermeshed non linearly separable data (b) using a Gaussian kernel the data gets separable using a non-linear boundary in the feature space.....	106
Figure 3.3	Working of the DAG.....	112
Figure 3.4	The OVO classifier.....	113
Figure 3.5	Different types of classifier fits, (a) under-fit classifier, (b) just-fit classifier and (c) over-fit classifier.....	117
Figure 3.6	The CV method along with GST for optimum selection of SVM parameters.....	118
Figure 3.7	Feature selection using the wrapper model.....	120
Figure 3.8	The flowchart for the proposed methodology.....	123
Figure 3.9	The flowchart for fault classification using noisy test data.....	125
Figure 4.1	Time domain features with dataset number for (a) mean (b) standard deviation (c) skewness (d) kurtosis (e) crest factor and (f) entropy...	131
Figure 4.2	Binary classification of faults with respect to no fault condition, speed (Hz) versus percentage classification accuracy of, (a) SB1 and HP, (b) SB2 and HP, (c) SB3 and HP, (d) SB4 and HP, (e) C0 and HP.....	142
Figure 4.3	Multi-class classification of blockage: percentage of classification accuracy versus speed (Hz).....	143
Figure 4.4	Speed versus classification accuracy of different features: Case D with B1.....	146
Figure 4.5	Blockage level versus average classification accuracy over entire speed range Case D.....	148
Figure 4.6	Classification accuracy of each fault at different speeds for (i) SB1, (ii) SB2, (iii) SB3, (iv) SB4; with $(\mu-\sigma-S)$ feature.....	150
Figure 4.7	Classification accuracy by testing with one fault at a time for different speeds with $(\mu-\sigma-S)$ feature Case F.....	151
Figure 5.1	Feature classification performance.....	162
Figure 5.2	Feature classification performance of different powers of σ	164
Figure 5.3	Feature values versus dataset number for σ (top) and σ^{-1} (bottom)...	164
Figure 5.4	Faults scatter plot for the low resolution data.....	165
Figure 5.5	Fault scatter plot for the high resolution data.....	165
Figure 5.6	Performance of the SVM classifier using vibration data alone and with both vibration and current data for the PCBF.....	167
Figure 5.7	Effect of number of training/ testing datasets on the classifiers performance for AF.....	168
Figure 5.8	Performance of the SVM classifier at intermediate speeds.....	178

Figure 5.9	Percentage overall classification accuracy versus the quality of data used for testing the algorithm.....	183
Figure 6.1	Feature values of σ (above) and $1/\sigma$ (below) for each data set of HP at 30 Hz.....	190
Figure 6.2	Average classification performance of various features to segregate HP, SB2, SB5, DB1, DB5, IF, IFSB2, IFSB5, IFDB1, IFDB5, PC, PCSB2, PCSB5, PCDB1 and PCDB5 faults.....	193
Figure 6.3	Performance comparison of the classifier using the data sampled at high frequency resolution and low frequency resolution to segregate HP, SB2, SB5, DB1, DB5, IF, IFSB2, IFSB5, IFDB1, IFDB5, PC, PCSB2, PCSB5, PCDB1 and PCDB5 faults.....	194
Figure 6.4	Scatter plot of HP, SB2, SB5, DB1, IF, IFSB2, PC, and PCSB2 faults at 45 Hz using features μ , σ , and $1/\sigma$	194
Figure 6.5	Mean versus sample set.....	206
Figure 6.6	Standard deviation versus sample set.....	208
Figure 6.7	Performance of vibration and current signals.....	208
Figure 6.8	Overall classification of uncorrupted, 3AWGN and 5AWGN data...	210
Figure 7.1	CWT spectrum - scale versus space for (a) HP, (b) SB2, (c) IF, (d) IFSB2, (e) PC, and (f) PCSB2 faults using db9 wavelet.....	221
Figure 7.2	Eigenvalues versus Principal component for, (a) HP, (b) IF, (c) PC and (d) SB5 faults.....	226
Figure 7.3	Feature selection using AF fault set and db9 wavelet.....	227
Figure 7.4	The classification accuracies of various wavelet families at 35 Hz, 50 Hz, (a) bi-orthogonal family, (b) reverse bi-orthogonal family, (c) coiflet, (d) Daubechies, (e) Gaussian, (f) Mexican hat, Morlet, Meyer, (g) symlets.....	229
Figure 7.5	Overall classification accuracies of the selected wavelets for BF faults case.....	231
Figure 7.6	Overall classification accuracies of the selected wavelets for IFBF faults case.....	233
Figure 7.7	Overall classification accuracies of the selected wavelets for the PCBF faults case.....	235
Figure 7.8	Overall classification accuracies of the selected wavelets for the dry run faults case.....	237
Figure 7.9	Comparison of the fault classification accuracies obtained for the AF fault classification using 5%, 10%, 25% AWGN and the baseline data.....	242
Figure 7.10	Comparison of the standard deviation feature for no-noise, 5% AWGN, 10% AWGN, 25% AWGN for the HP fault at 30 Hz speed	242
Figure 8.1	WPT decomposition.....	248
Figure 8.2	Feature scatter plot for HP, SB2, SB5, DB5, IF, IFSB2, PC and PCSB2 faults. Plot showing 4-D of mean (x-axis), standard deviation (y-axis), entropy (z-axis) and RMS (colour bar).....	253
Figure 8.3	Fault classification ability of different wavelets of Haar and Daubechies family at 35, 50, and 65 Hz.....	255

Figure 8.4	Average fault classification ability of different wavelets belonging to Symlets, Coiflets, Dmey, biorthogonal and reverse biorthogonal families.....	256
Figure 8.5	WPT based fault classification methodology.....	258
Figure 8.6	Percentage classification accuracy versus CP operating speed. Algorithm trained and tested at the same speed. The performance comparison of WPT-BE-MSVM and WPT-PCA-MSVM for BF fault case.....	260
Figure 8.7	Percentage classification accuracy versus CP operating speed. Algorithm trained and tested at the same speed. Comparison of performance of WPT-BE-MSVM and WPT-PCA-MSVM for IFBF fault case.....	261
Figure 8.8	Percentage classification accuracy versus CP operating speed. Algorithm trained and tested at the same speed. Comparison of performance of WPT-BE-MSVM and WPT-PCA-MSVM for PCBF fault case.....	264
Figure 8.9	Percentage classification accuracy versus CP operating speed. Algorithm trained and tested at the same speed. The performance comparison of WPT-BE-MSVM and WPT-PCA-MSVM for DR fault case.....	264
Figure 8.10	Percentage classification accuracy versus CP operating speed. Algorithm trained and tested at the same speed. The fault classification performance comparison of WPT-BE-MSVM and WPT-PCA-MSVM for AF fault case.....	267
Figure 8.11	Percentage classification accuracy versus CP operating speed. Algorithm trained and tested at different speeds (30, 40/35 stand for algorithm training speeds of 30, 40 Hz and algorithm test speed of 35 Hz).....	268
Figure 9.1	Comparison of AF fault family classification using features extracted from various domains at an intermediate speed.....	288

List of Tables

Table 1.1	Types of bearings used in CP system. Color code: Red – No capacity, Yellow – Low capacity, Orange – Medium capacity, Blue – High capacity, Green – Very high capacity.....	13
Table 1.2	The problems in a CP and their probable causes	18
Table 1.3	Typical signals used to monitor various mechanical/ electrical components.....	24
Table 1.4	Typical signals used to monitor various mechanical/ fluid faults in CPs.....	25
Table 1.5	Conventional techniques used for diagnosis of some mechanical/ fluid faults in CPs	27
Table 1.6	CP fault frequencies	28
Table 1.7	Various visual inspection techniques for classification of CP faults	39
Table 1.8	Comparison of researches on application of machine learning to fault diagnosis of CPs	61
Table 2.1	Technical specification of various MFS and allied components	75
Table 2.2	Details of the accelerometers	80
Table 2.3	Three CP configurations used in the current research	87
Table 2.4	Fault combinations considered in the proposed work	87
Table 2.5	Nomenclature, characteristics and description of various CP faults considered.....	89
Table 2.6	Details of data resolution and frequency content with different sampling rates.....	92
Table 2.7	Total data collected using different sampling rates of data acquisition	93
Table 3.1	SNR _{dB} value for different speeds using accelerometers and line current signals.....	125
Table 4.1	Multiclass fault classification at 40 Hz for feature finalization.....	134
Table 4.2	Percentage prediction accuracy in the binary classification of data	137
Table 4.3	Percentage accuracy in multiclass classification	138
Table 4.4	Percentage prediction accuracy in the binary classification HP and C0.....	139
Table 4.5	Percentage prediction accuracy of various features for binary and multiclass classifications at 40 Hz.....	140
Table 4.6	Fault classification cases and their description	145
Table 4.7	Calculation of $O_{(\mu-\sigma-S)}$ with Case F at SB2.....	147
Table 4.8	O_j , of different features Case F	149
Table 4.9	Confusion matrix with a prescribed γ value, for 30 Hz CP speed... ..	152
Table 4.10	Confusion matrix with a prescribed γ value, for 35 Hz and 40 Hz... ..	152
Table 4.11	Confusion matrix with a prescribed γ value, from 45 Hz to 55 Hz... ..	153
Table 4.12	Confusion matrix with a prescribed γ value, from 60 Hz to 65 Hz... ..	154
Table 5.1	Total data collected and features extracted for different data acquisition rates used	163

Table 5.2	Fault classification cases.....	167
Table 5.3	Case A performance of the classified with BF and HP faults.....	170
Table 5.4	Case A performance of the classified with IFBF and HP faults.....	172
Table 5.5	Case A performance of the classified with PCBF and HP faults.....	174
Table 5.6	Case A performance of the classified with dry runs and HP faults...	175
Table 5.7	Case A performance of the classified with AF and HP faults.....	177
Table 5.8	Prediction performance of classifier when testing with uncorrupted data and training with 5% AWGN data	180
Table 5.9	Prediction performance of classifier when testing with uncorrupted data and training with 10% AWGN data	181
Table 5.10	Prediction performance of classifier when testing with uncorrupted data and training with 25% AWGN data	182
Table 5.11	Comparison of prediction performance of the current methodology with other works in literature in time-domain.....	184
Table 6.1	Statistical features extracted in the frequency domain.....	189
Table 6.2	Total data collected using different sampling rates of the data acquisition.....	192
Table 6.3	Fault classification cases considered	195
Table 6.4	Case A performance of the classified with BF and HP faults.....	198
Table 6.5	Case A performance of the classified with IFBF and HP faults.....	200
Table 6.6	Case A performance of the classified with PCBF and HP faults.....	202
Table 6.7	Case A performance of the classified with dry runs and HP faults...	203
Table 6.8	Case A performance of the classified with AF and HP faults.....	204
Table 6.9	Case B AF fault classification performance at intermediate speeds	207
Table 6.10	Multi fault classification with 3AWGN corrupted data.....	211
Table 6.11	Multi fault classification with 5AWGN corrupted data.....	212
Table 6.12	Comparison of fault prediction performance in frequency domain to that presented in literature.....	214
Table 7.1	List of wavelet families and order of wavelets considered	220
Table 7.2	Scales giving the top three energies for various fault conditions in x, y, z directions of the accelerometer 1 (A1x, A1y, A1z, respectively), the accelerometer 2 (A2x, A2y, A2z, respectively) and the line phases (C1, C2, C3).....	224
Table 7.3	Fault classification cases considered	230
Table 7.4	BF fault classification using morl mother wavelet.....	232
Table 7.5	IFBF fault classification using the bior3.3 mother wavelet	234
Table 7.6	PCBF fault classification using the gaus1 mother wavelet.....	236
Table 7.7	Dry run fault classification using the gaus1 mother wavelet.....	238
Table 7.8	AF fault classification using the guas1 mother wavelet	240
Table 7.9	Case B the AF fault classification performance at intermediate speeds.....	241
Table 8.1	Families and order of wavelets considered in the present work	249
Table 8.2	The feature and energy level selection for WPT-BE-MSVM.....	250
Table 8.3	Binary feature permutations and combinations performance for WPT-BE-MSVM	252
Table 8.4	Feature selection for the WPT-PCA-MSVM	254

Table 8.5	Fault classification cases considered	257
Table 8.6	The BF fault family classification using WPT-BE-MSVM and WPT-PCA-MSVM	259
Table 8.7	IFBF fault family classification using WPT-BE-MSVM and WPT-PCA-MSVM	262
Table 8.8	PCBF fault family classification using WPT-BE-MSVM and WPT-PCA-MSVM	263
Table 8.9	Dry run fault family classification using WPT-BE-MSVM and WPT-PCA-MSVM	265
Table 8.10	AF fault family classification using WPT-BE-MSVM and WPT-PCA-MSVM.....	269
Table 8.11	Comparison of fault classification accuracies obtained for AF fault classification using 5%, 10% and 25% AWGN.....	271
Table 8.12	Comparison of fault prediction performance using wavelets to that presented in literature.....	272
Table 9.1	Features used for CP multiclass fault classification in time-domain, frequency-domain and time-frequency domain	277
Table 9.2	Comparison of BF fault family classification using features extracted from various domains.....	280
Table 9.3	Comparison of IFBF fault family classification using features extracted from various domains.....	286
Table 9.4	Comparison of PCBF fault family classification using features extracted from various domains.....	287
Table 9.5	Comparison of dry run fault family classification using features extracted from various domains.....	288
Table 9.6	Comparison of AF fault family classification using features extracted from various domains.....	289
Table 9.7	Comparison of AF fault family classification using features extracted in various domains from corrupted testing data.....	291
Table 9.8	Bird's eye view of various fault classification cases.....	292



Nomenclature

b	Bias
BPF	Blade pass frequency
C	SVM penalty parameter
C_a	Percentage classification accuracy
G	Acceleration due to gravity
$\bar{h}$	Mother wavelet
H	Total head at the considered point
I	Impulsion index
K	Kurtosis index
LCF	Line current frequencies
$\dot{m}$	Mass flow rate
N	Speed of the rotor
N_s	Specific speed
O_a	Percentage overall classification accuracy
P	Crest factor
P	Pressure at the chosen point
p_v	Absolute vapor pressure
PA	Power to average ratio
R	Power consumption
RSS	Root sum of squares
S	Entropy
SNR	Signal to noise ratio
T	Tolerance index
V	Velocity of the fluid
$\dot{V}$	Volumetric flow rate at the discharge
$\mathbf{w}$	Normal to hyperplane constructed by SVM
W	Specific work
$W_{p,q}^n$	Wavelet packet transform
$\mathbf{X}$	Feature points located on hyperplane
Y	Class label
Z	Elevation of the point above the reference
β_i	Lagrange multiplier
Γ	Gaussian RBF kernel parameter
H	Efficiency
K	Kurtosis
M	Mean
μ_k	k^{th} statistical moment
P	Density of fluid at all points
Σ	Standard deviation
X	Skewness
ψ	Continuous wavelet transform
ω	Shape factor



Abbreviations

A1x, A1y, A1z	Accelerometer -1 Measurement in x, y, z Directions
A2x, A2y, A2z	Accelerometer -2 Measurement in x, y, z Directions
AE	Acoustic Emission
AWGN	Additive White Gaussian Noise
AC	Alternating Current
ABCA	Artificial Bee Colony Algorithm
ANN	Artificial Neural Network
NPSH _A	Available NPSH
BNC	Bayonet Neill Concelman Connector
BEP	Best Efficiency Point
BF	Blockage Fault family
BPF	Blade Pass Frequency
C0	Beginning of Cavitation
CP	Centrifugal Pump
CBM	Condition Based Maintenance
CWT	Continuous Wavelet Transform
CV	Cross Validation
C1, C2, C3	Current Measurements in 1, 2, 3 Phases
DAQ	Data Acquisition System
DAG	Direct Acyclic Graph
DC	Direct Current
DB	Discharge Blockage
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
EMD	Empirical Mode Decomposition
GA	Genetic Algorithm
GST	Grid-Search Technique
HP	Healthy Pump
HPZ	High Pressure Zone
IFBF	Impeller Fault Blockage Fault family
IF	Impeller Faults
IFDB	Impeller Faults with Discharge Blockage
IFSB	Impeller Faults with Suction Blockage
IAS	Instantaneous Angular Speed
IMF	Intrinsic Mode Functions
ITD	Intrinsic Time-Scale Decomposition
kNN	k Nearest Neighbors
LE	Leading Edge
LSSVM	Least Square Support Vector Machine
LTSA	Load Torque Signature Analysis
LAN	Local Access Network
LLE	Local Linear Embedding
LPZ	Low Pressure Zone

MFS	Machine Fault Simulator
MVU	Maximum Variance Unfolding
MSB	Modulation Signal Bispectrum
NPSH	Net Positive Suction Head
NN	Neural Network
NSP	Non-Dimensional Symptom Parameters
OVA	One versus all
OVO	One versus one
OEM	Original Equipment Manufacturer
PNN	Partially Linearized Neural Networks
PSO	Particle Swarm Optimization
PCDB	Pitted Cover fault with Discharge Blockage
PCSB	Pitted Cover fault with Suction Blockage
PCBF	Pitted Cover-plate Blockage Fault family
PC	Pitted Cover-plate fault
PCA	Principal Component Analysis
PRC	Proper Rotation Component
RBF	Radial Basis Function
NPSH _R	Required NPSH
RMS	Root Mean Square
SNR	Signal To Noise Ratio
SLT	Statistical Learning Technique
SB	Suction Blockage
SVM	Support Vector Machine
SSP	Synthetic Symptom Parameter
TE	Trailing Edge
USB	Universal Serial Bus
VFD	Variable Frequency Drive
WPT	Wavelet Packet Transform
WT	Wavelet Transform

CHAPTER 1: Introduction and Literature Review

1.1 Introduction

Centrifugal pumps (CPs) belong to the family of turbo-machines. They are used for the transportation of fluid, achieved by raising the defined volume flow to a desired pressure level. It is estimated that, in a standard chemical plant, the number of CPs used approximately equals its employee headcount (Hennecke, 2000). Also, roughly, twenty-percent of the energy generated globally is utilized in driving the CPs (Hart, 2002). Thus, it can be safely said that CPs are critical plant equipment and their efficient working is essential for the persistent functioning of the plant.

In recent years the research on fault diagnosis and monitoring of CPs has picked up the pace. The industrialists and users of the pumps are now enthusiastic to embrace diagnostic features in both software/ hardware form to improve equipment reliability and safety of personnel. Condition based monitoring (CBM) is one of the solutions to machinery maintenance requirements. The traditional CBM methods of fault diagnosis include locating fault specific harmonic components from the acquired signals such as vibration, line current, pressure and acoustic emission. Also, the physical model of the system derived from mathematical equations based on a set of assumptions has also proved to be very useful in CP fault diagnosis. However, the latest approaches to CBM aims at reducing human engagement in the real-time fault detection and decision making. CBM can be accomplished by using machine learning techniques like fuzzy-logic-based systems, neural networks, and support vector machines; among many others. It is undeniably apparent that this area has a vast scope. This chapter is aimed at describing the fundamental working principle of CPs, their performance measures,

various kinds of faults in them and the signals they generate and their CBM based diagnostics' schemes. Also, the primary objectives and scope of the thesis are presented in this chapter.

1.2 Working of CPs

The energy transfer in turbo-machines directly depends upon the rotor speed. On the contrary, positive displacement pumps always deliver the same amount of stroke volume irrespective of the speed of the rotor. The CP primarily comprises a pump shaft, a bearing housing, an impeller and a casing. A shaft seal in a CP averts the fluid from discharging into the surroundings or the bearing housing. The CP is energized by a motor, which is coupled to the pump shaft. The fluid that needs to be pumped runs past the suction nozzle into the impeller. Impeller then transmits the necessary energy to convey the fluid by accelerating it in the circumferential direction. The inner surface of the casing is called the volute. This volute has an expanding cross-sectional area as it envelops around the pump casing. The objective of the volute is to gather the high-speed fluid dispensed by the impeller from its periphery and to gradually reduce the velocity head of the fluid, thereby increasing its pressure head. This fluid is later discharged from the CP through the discharging connections.

Figure 1.1 shows the plan view and section of the impeller blade. The convex surface of the impeller blade feels the maximum pressure for a given radius. It is thus called the *pressure side*. The concave surface of the blade has a lower pressure and therefore is termed as the suction side. In Figure 1.1 the LE stands for leading edge, and the TE stands for trailing edge of the blade.

In a CP the impeller does all the work, and therefore much research focuses on the optimization of the design of the impeller. The volute is often regarded as a mere collector of the fluid. However, it is not entirely accurate to consider the independent operation of the impeller and the volute. The pump casing defines the environment in which the impeller works and thus has a profound effect on the performance of the CP.

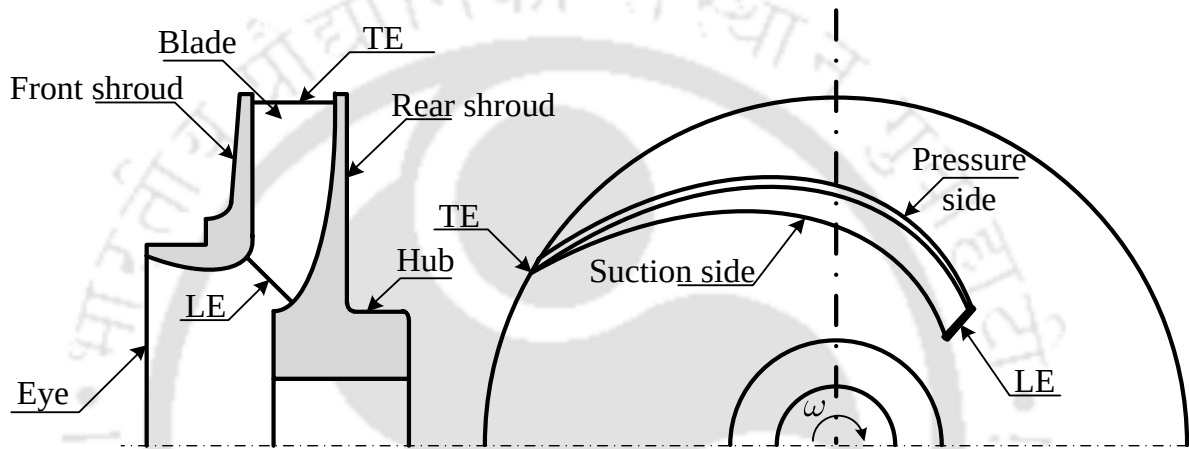


Figure 1.1: The section (left) and plan (right) views of an impeller, LE: Leading edge, TE: Trailing edge

Improper design of the impeller may even cause inefficient working of the CP (Worster, 1963). Consideration of impeller-volute interaction is essential in the design of the CPs. The description of the design aspects of these is beyond the scope of this work. The readers' may find the following references useful to understand their design philosophy (Worster, 1963; Lobanoff and Ross, 2013).

1.3 Performance of CPs

This section describes briefly the fundamental principles that govern the working of CPs and the various performance parameters that decide the functioning of them.

1.3.1 Energy Conservation

In fluid dynamics, the energy conservation is given by Bernoulli's principle (Gulich, 2008; Elger and Roberson, 2016). It states that with an increment in the velocity of the fluid, there is a corresponding decrease in its pressure head. In general, for a streamline flow of an incompressible fluid, the Bernoulli's equation is given as,

$$\frac{v^2}{2g} + z + \frac{p}{\rho g} = H \quad (1.1)$$

Where, v is the velocity of the fluid, g is the acceleration due to gravity, z is the elevation of the point above the reference (negative value of z is considered if the point is below the reference), p is the pressure at the chosen point, ρ is the density of fluid at all points, and H is the total head at the considered point.

For a flow across a pipe, as shown in Figure 1.2, the total head developed is given by the difference between heads developed at section 2 and section 1. It can be represented as,

$$H_2 - H_1 = \frac{v_2^2 - v_1^2}{2g} + z_2 - z_1 + \frac{p_2 - p_1}{\rho g} \quad (1.2)$$

where, v_1, v_2 are the velocities, z_1 and z_2 are the elevations from the datum ($z = 0$), and p_1 and p_2 are pressures at pipe sections 1 and 2, respectively.

1.3.2 CP Performance Parameters

The performance of the CPs is quantified by the following variables (Gulich, 2008):

- 1) Volumetric flow rate at the discharge, $\dot{V}$
- 2) Specific work, W or head developed $H = W/g$

- 3) Power consumption, R
- 4) Efficiency, η
- 5) Net positive suction head, NPSH,
- 6) Speed of the rotor, N

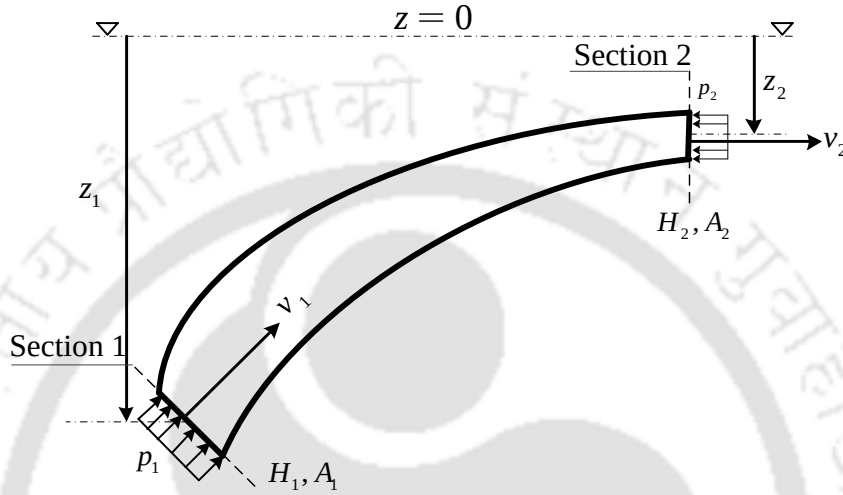


Figure 1.2: Streamline and incompressible fluid flow across a pipe

In a CP, the specific work W , is given as the effective energy transferred by the pumping action to the fluid per unit mass. It is calculated between the suction and the discharge sides of the CP. W can be given as the product of total dynamic head H and acceleration due to gravity g . For a pumping system shown in Figure 1.3, the total dynamic head is $H = H_d - H_s$

In which, the subscripts d stands for discharge and s stands for suction. The values of H_d and H_s can be found out from Eq. (1.1). Thus, the head at the suction is given as,

$$H_s = \frac{v_s^2}{2g} + z_s + \frac{p_s}{\rho g} \quad (1.3)$$

Also, the head at the discharge is given as,

$$H_d = \frac{v_d^2}{2g} + z_d + \frac{p_d}{\rho g} \quad (1.4)$$

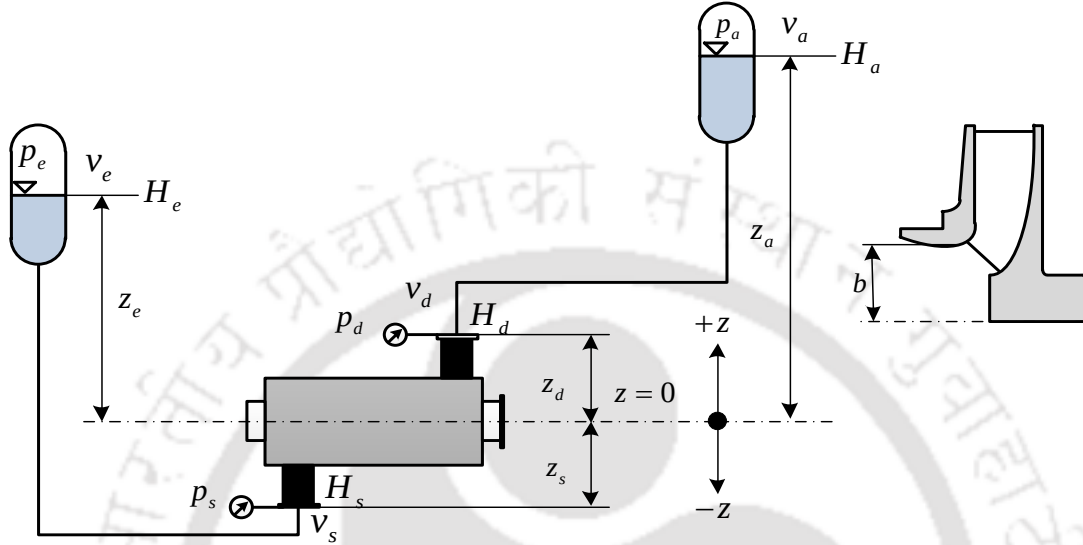


Figure 1.3: Schematic diagram of CP working in a typical industrial application

Therefore, the total dynamic head is given as,

$$H = H_d - H_s = \frac{v_d^2 - v_s^2}{2g} + z_d - z_s + \frac{p_d - p_s}{\rho g} \quad (1.5)$$

If the supply to the suction and delivery of the discharge are at levels e and a , respectively, then there are losses in the piping connections to the CP and from the CP. Let, $H_{l,s}$ and $H_{l,d}$ represent the loss in head in the suction pipe and discharge pipe, respectively.

Therefore, the suction head and discharge head can be re-written as,

$$H_s = H_e - H_{l,s} \quad (1.6)$$

$$H_d = H_a + H_{l,d} \quad (1.7)$$

where,

$$H_e = \frac{v_e^2}{2g} + z_e + \frac{p_e}{\rho g} \quad (1.8)$$

$$H_a = \frac{v_a^2}{2g} + z_a + \frac{p_a}{\rho g} \quad (1.9)$$

Apart from total dynamic head, another very important parameter for the CP design is the NPSH (net positive suction head). This is important because it helps in estimating the approach flow conditions of the CP so that appropriate design and layout may be chosen. The study of flow condition is vital because whenever the fluid pressure in the CP becomes lower than the vapor pressure of the pumping fluid, a partial vaporization of fluid takes place. This phenomenon is called as the *cavitation* (Young, 1999; Brennen, 2013). The occurrence of cavitation in a CP can be detrimental to CP's life and can have adverse effects on its performance (Gulich, 2008; Lobanoff and Ross, 2013). The NPSH that is required by the CP layout to completely or partially avoid cavitation is called as the *required NPSH* and is given by the symbol $NPSH_R$ and the NPSH that is available from the design is called the *available NPSH* and is given by $NPSH_A$. If p_v is the absolute vapor pressure of the fluid, then the NPSH can be given as,

$$NPSH = \frac{p_{s,abs} - p_v}{\rho g} + z_s + \frac{v_s^2}{2g} \quad (1.10)$$

where, the absolute static pressure at the suction is given as,

$$p_{s,abs} = p_{e,abs} + \rho g (z_e - z_s - H_{l,s}) + \frac{\rho}{2} (v_e^2 - v_s^2) \quad (1.11)$$

The available NPSH for the given plant conditions can be derived from Eq. (1.10) as,

$$NPSH_A = \frac{p_{e,abs} - p_v}{\rho g} + z_e + \frac{v_e^2}{2g} - H_{l,s} \quad (1.12)$$

From the Bernoulli's equation (Eq. (1.1)), the absolute value of pressure at the uppermost point of the impeller, located at a distance 'b' from the rotor axis can be calculated. This value should in all cases be greater than the vapor pressure so that there is no evaporation of large volumes of fluid. Any given pump design has a defined $NPSH_R$ value and it is expected that $NPSH_A > NPSH_R$. The liquid level Z_e required to avoid cavitation is given as,

$$Z_e = NPSH_R + H_{l,s} - \frac{p_{e,abs} - p_s}{\rho g} - \frac{v_e^2}{2g} + b \quad (1.13)$$

The useful power of a pump R_u is given as,

$$R_u = \dot{m}W \quad (1.14)$$

where, $\dot{m}$ is the mass flow rate, at it is given by, $\dot{m} = \rho \dot{V}$. If R is the total power required to drive the CP, the useful power R_u is always less than R owing to the losses in pumping components. Therefore, the efficiency of pumping system η is given as,

$$\eta = \frac{R_u}{R} = \frac{\dot{m}W}{R} = \frac{\rho \dot{V} H g}{R} \quad (1.15)$$

From the above equations, it is clear that with the variation in H the values of R_u and η also change considerably. The *pump characteristics* can be found by charting these values against the flow rate, $\dot{V}$. The flow-rate at which the CP efficiency is highest is called the *best efficiency point* (BEP) of it. All the CPs are designed to operate at their BEPs. At the BEP the pump parameters are given as $\dot{V}_{bep}$, H_{bep} , R_{bep} , and η_{bep} at a given speed N . All these performance parameters are unified by a quantity called the *specific speed*, N_s . It is given by,

$$N_s = N \frac{\sqrt{\dot{V}_{\text{bep}}/f_q}}{H_{\text{bep}}^{0.75}} \quad (1.16)$$

Here, N is in rpm, $\dot{V}_{\text{bep}}$ is in m^3/s and H_{bep} in m. f_q is the number of entries to the impeller, $f_q = k$ implies k entries to the impeller. The number of entries to the impeller refers to the number of suctions a pump can have. In general, it is expected that a CP has one suction and one discharge, called as “single suction pump”. But, there can even be two suctions of the CP where fluid enters axially from both the directions, called as a “double suction pump”. In double suction pump the axial thrust cancels out, and thus is employed for large quantities of fluid. The choice of the impeller type (radial, axial or semi-axial) depends on specific speed and the type of the CP application.

1.4 CP Applications

CPs have great importance in varied fields. It is estimated that the world market volume for CPs is 20 billion US dollars a year (Gülich, 2008). Their usages range from miniature CPs of a few Watts to large pump turbines of more than 250 MW. CPs are developed for flow rates ranging from 0.001 to 60 m^3/s , head ranges of 1 to 5000 m and operating speeds from a 100 to as high as 30000 rpm (Gülich, 2008). This speaks volumes about the wide-variety of CPs available for users having a wide range of pumping requirements. Some CP applications are given in Appendix A - Table A.1. They are taken from ref. (Gülich, 2008). From the table, it becomes apparent that there are extensive applications of CPs owing to their versatility, simple design, and efficient working. As it can be seen from the table, in the most of the applications CPs are not just some additional equipment, but they are some of those critical components that ensure the process flow of the industry. Hence, it is imperative that they be sufficiently

maintained to have sustained operation of the plant and to ensure operators safety. However, before understanding the maintenance strategies that can be used on CPs, it is essential to know what their various failure mechanisms are.

1.5 CP Failure Modes

Green and Casada (1995) in their book presented documentation on the possible causes of a decrease in performance/ availability of a CP. McKee et al. (2011) presented that there are 13 major failure causes of CPs applied in sewage industries. However, for studying the failure modes of a CP, the whole CP system needs to be considered. A typical CP system comprises of a foundation, driver, power transmission system, CP, suction pipes and valves, discharge pipes and valves, instrumentation for controlling CP flow and alignment anchoring devices (Nelson, 1992). The most common reasons for CP failures can be broadly divided into,

1. Mechanical vibration factors
2. Hydraulic vibration factors
3. Motor factors
4. Degradation of pumping system

1.5.1 Mechanical Vibration Factors

The primary causes of mechanical vibrations (Jackson, 1979; Greene et al., 1995; Lobanoff and Ross, 2013) in CPs are: (1) impeller and seal rubs, (2) unbalanced rotor, (3) misalignment, (4) damaged bearings, and (5) miscellaneous causes.

Impeller and seal rubs: Mechanical faults in a CP tend to cause a chain of faults in it subsequently. For example, the misalignment of the driver and the CP shaft or unbalanced impeller causes an increase in vibration of the system. This lets the rotating parts *rub* with the

stationary parts. Rubbing increases, the clearance in the wear rings resulting in leakage and loss of CP head. Rubbed parts generate high temperatures and when they are contacted with pumped fluid create the *cavitation*.

Unbalanced rotor: Rotating unbalance is a non-uniform distribution of mass around the axis of the rotor (Tiwari, 2017). A schematic of unbalance present on a rotor shaft is shown in Figure 1.4. The gap between the CP axis of rotation and its center of gravity is called as the *eccentricity*. This generally manifests itself at a frequency equal to the rotational speed of the CP. Vibrations at this frequency also appear for a bowed shaft or an eccentrically machined impeller. The unbalance in the rotors could be because of reasons like the manufacturing defects, a damaged rotor (shaft or impeller), improper balancing practices, etcetera.

Misalignment: The misalignment is a common reason for rotor failure (Patel and Darpe, 2009; Tiwari, 2017). The types of misalignments in a rotor are shown in Figure 1.5. When the axis of the driver and driven shafts are in line with each other, the rotor system is said to be in the aligned form. If the axis of the driver and driven shafts are parallel and offset, the misalignment is called the *parallel misalignment*. If the axes are intersecting at an angle, then the misalignment is termed as the angular misalignment. If both offset and angular axes are present, they give rise to the *combined angular and offset misalignment*. Misalignments can be caused due to improper axis setting of the CP system and/or due to improper couplings used. Vibrations due to misalignments manifest themselves in the frequency spectra at twice the operating speed of the CP. If the drive is transmitted using a coupling having pins/ teeth, misalignment can induce vibrations at frequencies that are multiples of the product of the number of pins (or teeth) and the operating frequency of the CP (Yedidiah, 2012).

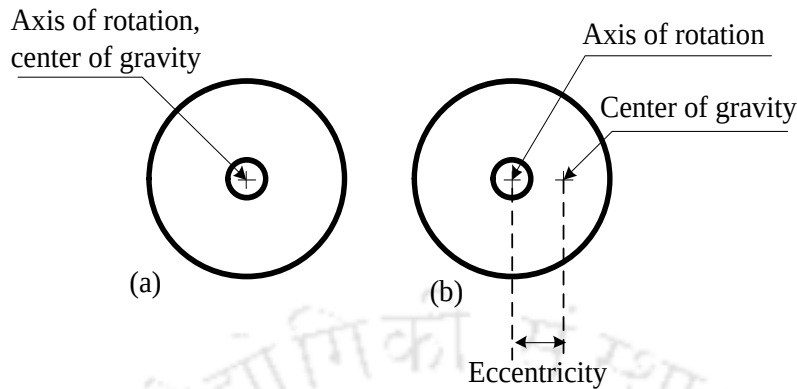


Figure 1.4: Unbalance in a rotor, (a) Balanced rotor, (b) unbalanced rotor

Damaged bearing: The choice of the bearings used in the CPs depends upon the application it is used in and the types of loads it is designed to take. The standard choices of bearings used in CP systems are tabulated in Table 1.1 (source: *SKF bearings*). It can be seen from the table that the amount of load (and its directions), speed and amount of misalignment handled changes mainly with the bearing chosen. In a CP, the axial load is due to hydraulic forces acting on the shrouds, moment force due to change in the direction of fluid flow through the impeller, hydrostatic force acting on the impeller (suction). The radial load in a CP is due to uneven distribution of pressure around the circumference of an impeller.

The failures of bearings can be due to: (1) excessive loading, (2) overheating, (3) brinelling, (4) fatigue failure, (5) contamination, (6) lubricant failure, (7) corrosion, (8) misalignment, and (9) improper fits. In CPs with ball bearings, defective races may cause vibrations whose frequency equals the product of the rotational speed of the CP unit and the number of balls in the bearing with the faulty races (Yedidiah, 2012).

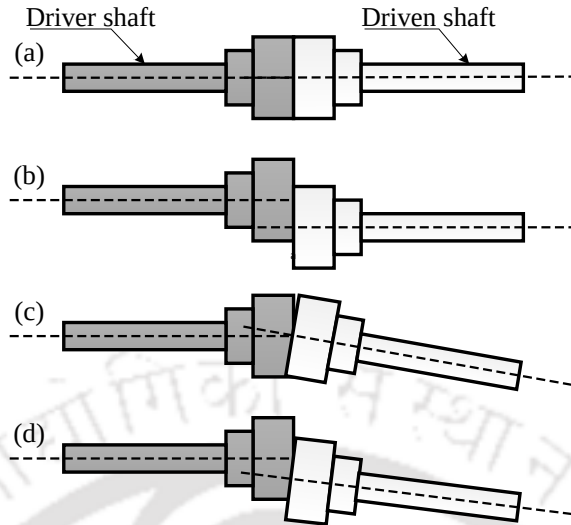


Figure 1.5: Misalignment in a rotor (a) correct alignment, (b) parallel misalignment, (c) angular misalignment, (d) combined angular and parallel misalignment

Table 1.1: Types of bearings used in CP system. Color code: Red – No capacity, Yellow – Low capacity, Orange – Medium capacity, Blue – High capacity, Green – Very high capacity

Type of bearing	Characteristics			
	Axial Load	Radial Load	Speed	Misalignment
Single row deep groove ball bearings	Yellow	Yellow	Green	Orange
Double row angular contact ball bearing	Orange	Orange	Blue	Yellow
Single row angular contact ball bearing pair	Orange	Green	Blue	Yellow
Cylindrical roller bearing	Blue	Red	Blue	Yellow
Spherical roller bearing	Green	Orange	Orange	Green
Taper roller bearing set	Green	Green	Orange	Yellow
Spherical roller thrust bearing	Red	Green	Orange	Green

Miscellaneous causes: The different causes of CP vibrations include a vibrating piping system, too loose or too tight fixtures, flexible foundations, unequal tightening of bolts, damaged impeller, shaft cracks, CP internal surface damages, etcetera. The frequencies of vibration of these causes (the most of them) are not directly related to the CP operating speed. They depend upon the natural frequencies of vibrating components (Yedidiah, 2012).

1.5.2 Hydraulic Vibration Factors

The source of vibrations or noise can also be due to hydraulic causes. The pressure variations generated by the liquid motion causes them. The vortex development can generate noise due to turbulence, stall, pulsations, flashing, cavitation, flow separation, impeller interaction with the pump cutwater, water hammer, and dry run (Lobanoff and Ross, 2013). The details of the mechanism of formation of various flow instabilities and pressure pulsations are given in refs. (Makay, 1980; Lobanoff and Ross, 2013). These pressure pulsations produce either discrete or broadband frequency components. They can be divided into four types based on the frequencies they produce (Szenasi et al., 1986),

1. Spasmodic broadband energy bursts caused due to the cavitation, water hammer and flashing
2. Discrete frequency components produced at blade pass frequency
3. Broad-band energy resulting from high-velocity turbulent flows
4. Flow-pulsations induced because of flow past the impasse and side branches

Here, blade pass frequency (BPF) is given by,

$$BPF = i \times n \times N \quad (1.17)$$

where, n is the number of impeller blades and N is the rotational speed, i is the i^{th} harmonic excited.

Water hammer: Start or stop of the CP with sudden opening or closure of the valve results in high-pressure surge called the water hammer. It applies a sudden impact force to the internal surfaces of the CP. It has the capability of cracking concrete structures to which pipes are anchored (Lobanoff and Ross, 2013).

Stall: When a CP operates in off-design conditions, the flow from the impeller does not agree with the volute requirement, thereby changing the pressure pulsation attributes. When operating at low flow rates, a rotating stall phenomenon may take place. Stall produces low-frequency vibration signals in the CP (Zhang et al., 2014).

Cavitation and Flashing: The mechanism of cavitation is already explained in the script. A detailed discussion on cavitation assessment, detection and damage are given in refs. (Sloteman, 2007) and (Ashokkumar, 2011). Flashing occurs in CP systems while pumping hot water. When hot pressurized water flows past a blockage, there is a reduction in pressure making the fluid to vaporize or flash. It gives similar noise and vibrations as the cavitation.

Turbulence: Obstruction to flow results in a drop in the flow rate and a development of secondary flow called the recirculation flow. As the separation of flow increases, the vortex is formed which results in local pressure drop and thus formation of vapor bubbles (Tabar et al., 2011). There are different types of recirculation in a CP, including suction recirculation and discharge recirculation (Fraser, 1981). Suction recirculation results in the cavitation damage on the pressure side of the vane at the inlet of the impeller and, random crackling

sound in the suction, surging in the suction of pump, noise, and vibrations (Fraser, 1981). Discharge recirculation gives rise to the cavitation damage on the pressure side of the impeller vane at its discharge and inlet of the diffuser vanes of the casing. It also causes the axial shaft movement, cracking and failure of impeller shrouds, shaft failures, noise and excessive vibrations (Fraser, 1981).

Dry run: Many CPs are designed to work when primed. That is, before the start of the CP the suction line and the CP itself must be fully filled with the fluid to be pumped. This procedure is called the *priming*. However, sometimes the machinists overlook the drying up of the reservoir or dearth of priming fluid in CPs and this results in a *dry run* condition, which causes excessive heat generation. In turn, this causes bearing, seal failures, and high vibrations.

Apart from all the causes mentioned above for CP failures, another significant reason for the deterioration in CP performance and thus a failure of it could be when it operates away from the BEP of the CP can also cause an increase in vibrations in a CP (Behzad et al., 2004).

1.5.3 Motor Factors

To maintain the integrity of the pumping system it needs to be ensured that the driving motor is also defect-free. This is because the faults on the motors cause anomalies in the operation of the CP and vice versa. The common causes of failures of motors are (Greene et al., 1995),

1. Bearing faults (~41% failures)
2. Stator faults (~37% failures)
3. Rotor faults (~10% failures)
4. Unbalance, misalignment, etc. (~12% failures)

The further details of motor failures are not in the scope of the present script. Some works on the electric motor fault diagnosis are given in refs. (Nandi and Toliyat, 1999; Filippetti et al., 2000; Nandi et al., 2005).

1.5.4 Degradation of Pumping Systems

The CPs may fail because of the degradation of their internal components due to any of the causes mentioned above or due to the environments they are used in. The common causes of CP degradation are the corrosion, erosion, abrasion, wear, fractures, cracks, extrusions, and dents. Fractures, cracks, dents, and extrusions usually occur due to excessive stresses in the CP system. The types of stresses could be tension, compression, bending, shear, and fatigue. Blows usually cause dents and extrusions. When a fluid interacts with a solid surface for a long time, the dampened surface of that solid may degenerate due to one or a combination of the following reasons: chemical reactions, electrolytic action, and crevice corrosion (Yedidiah, 2012).

1.6 Consolidated List of CP Problems and Their Causes

List of problems observed on a CP and their causes are given in Table 1.2 (Yedidiah, 2012).

1.7 Condition Monitoring

After understanding of different failure mechanisms in CPs, the next logical step would be to design a maintenance strategy to safeguard the CP and its allied components to improve its longevity, ensure operators safety and ensure good monetary returns. The fault management procedures used in plants take the form of maintenance or repair or reconstruction (Isermann,

2011). The different types of fault management strategies are depicted in Figure 1.6. The maintenance in a plant for a process or equipment could be broadly classified as planned maintenance or unplanned maintenance. Planned maintenance could be of three types, preventive maintenance, corrective maintenance or on-demand maintenance. It is set duration and/ or set working-hours based. Advancement of this is to employ maintenance on demand. This relies on the real-time data to describe the operating state of the system. On finding an anomaly, the maintenance is requested. These types of planned maintenances are generally done within the overhaul period.

Table 1.2: The problems in a CP and their probable causes

CP Problems	Causes
CP does not develop head nor delivers fluid	1. Dry run 2. Broken shaft 3. Disengaged driver and CP 4. Impeller key broken 5. CP without impeller
CP develops head but does not deliver fluid	1. Entrapped air 2. Clogged suction 3. Strainer clogged 4. High discharge pressure (greater than max pressure developed by CP) 5. Low operating speed 6. Wrong direction of operation 7. $NPSH_A$ inadequate
Less fluid delivered than expected (or) Enough CP pressure not developed (or) Shape of performance curve deviates from the rated curve	1. Air entrapment 2. Low operating speed 3. Wrong direction of impeller rotation, damaged impeller 4. High pressure demand 5. $NPSH_A$ inadequate 6. Impeller or casing clogged 7. Leakage via wear rings/ seal faces 8. Casting imperfections in impeller or casing 9. Blockage/ Obstruction on suction or discharge pipes 10. Strainers clogged 11. Recirculation 12. CP operating away from the BEP conditions

Table 1.2: The problems in a CP and their probable causes (contd.)

CP Problems	Causes
CP consumes a lot of power	1. High operating speed 2. Fluid at higher specific gravity or viscosity than originally mentioned 3. Oversized impeller, damaged impeller 4. Misalignment 5. Rubbing 6. Damaged bearings 7. Bowed shaft 8. Incorrect direction of rotation 9. NPSH_A inadequate 10. CP operating away from the BEP conditions 11. Failure of discharge line
CP performance deteriorates in a short time after starting	Unwanted air entrapment, leakage, air funnels in the CP system
CP operation develops noise and vibrations	1. Misalignment 2. Rubbing 3. Worn out bearings 4. Incorrect direction of rotation 5. NPSH_A inadequate 6. Cavitation 7. Impeller damage 8. CP operating at critical speed 9. Suction/ discharge pipes clogged 10. Bowed shaft 11. Unbalance 12. Entrapped air 13. Recirculation 14. CP operating away from the BEP conditions 15. Improper foundation 16. Resonance between CP and base plate or foundations 17. Casing or impeller erosion 18. Uneven thermal expansion of CP parts
Short life of mechanical seals (or) Seal leaks excessively	1. Bearing wear 2. Unbalance 3. Bowed shaft 4. Misalignment 5. Debris in liquid flushing the seal 6. Seal not perpendicular to CP axis 7. Abrasion 8. Improper installation

Table 1.2: The problems in a CP and their probable causes (contd.)

CP Problems	Causes
Bearing damages	1. Impeller damage 2. Clogged impeller 3. Unbalanced rotor 4. Excessive axial or radial loads (not in line with bearing chosen) 5. Misalignment 6. CP operates at low flow rates for a long time 7. Improper installation 8. Bearing housing damages 9. Insufficient cooling 10. Improper lubrication 11. Dirt in bearings 12. Excessive suction pressure
CP overheats	1. Dry run 2. Air entrapment 3. Misalignment 4. Damaged bearings 5. Improper lubrication 6. CP operates near shut-off
Short life of impeller or casing	1. Corrosion 2. Fatigue from thermal shocks/ vibrations 3. Cavitation damages 4. Abrasion 5. Excessive stresses from bolts in piping installations or foundation 6. Transient stresses during start and stop of CP
Casing bursts at start or stop of CP	1. Water hammer 2. Slam pressure
Flow decreases periodically then returns to normal	Change in liquid level in the suction tank

On the other hand, unplanned maintenance strategies are based upon reacting to damage when there is almost a shutdown of the process (or the equipment). A re-configuration can be applied if there are redundant components in the system and if they can be put to use. Therefore, advanced maintenance strategies involve prevention of failure, fixing measures to

take out faults and restructuring or redesigning to avert failures because of additional mechanisms with degrading functions.

In a typical industry, 15 – 40% of the manufacturing costs are attributed to maintenance (Rao, 1996). Therefore, it is vital that industries device a proper maintenance strategy to cut down the extra expenditures. The scope of this work has been limited to on-demand maintenance scheme, also known as the *condition based maintenance* (CBM) or *structural health monitoring*. The CBM has gained popularity with industries as it helps in reducing unwanted break-down of the plant and recommends maintenance only when required. This thereby saves a lot of time and cost for the plant. It is estimated that an investment of 10,000 to 20,000 USD on CBM per year would save up to 5,00,000 USD per year (Rao, 1996). The CBM involves a holistic approach of integrating varied fields of expertise including, the management, economics, information and technology, instrumentation engineering, mechanical and structural engineering, electrical engineering, etc. (Rao, 1996). The subsequent section would discuss the working of CBM.

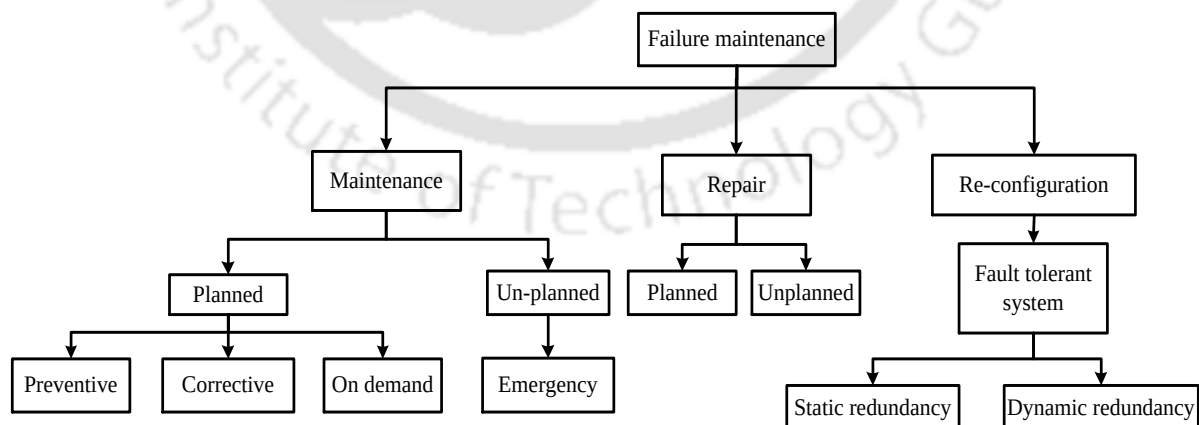


Figure 1.6: Plant fault maintenance strategies

1.8 Philosophy of CBM

Condition based monitoring encompasses evaluating the ‘condition’ of the system. This is done by carefully observing the current working status of the system via some useful measured parameters and/or signals. The necessary flow of a CBM working is given in Figure 1.7. It involves,

- Step 1:** Acquiring data from sensors attached to the system to be monitored
- Step 2:** Pre-processing the data to required format
- Step 3:** Comparing this data with that of the parameters at standard operation of the system
- Step 4:** Estimating the deviation observed
- Step 5:** If the deviation observed is above a threshold (or acceptable) limit, identify the reason for the deviation and alarm a fault to the maintenance engineer.

From the flowchart, it is clear that the choice of the signals (or sensors) used in the CBM is crucial and needs to be judiciously made. The details on the types of signals that can be used are given in the subsequent section.

1.9 Signals Used for CBM Analysis

There are numerous choices of signals that can be acquired for the CBM of mechanical systems. A consolidated list of the popular signal choices used to monitor various mechanical and electrical components is given in Table 1.3 (Alpaydin, 2008).

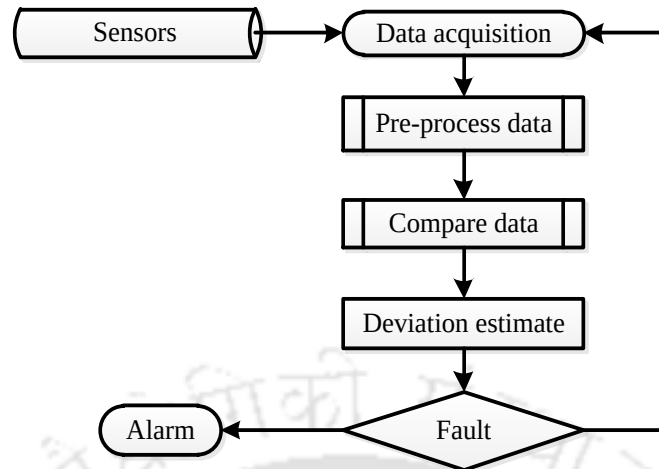


Figure 1.7: Flowchart on working of CBM

Grey color in the table shows the usefulness of a particular signal in fault identification of particular component. It is evident from the table that more than one signal type can be used to monitor each of the equipment. However, the choice of the signal depends upon various factors like, the type of measuring equipment (intrusive/ non-intrusive, small/ large, etc.), ease of access of the measuring point, cost of the sensors, sensitivity of the sensors, availability of the sensors, proficiency of technicians in handling the sensors, etc. Figure 1.8 gives the details about different measuring locations for various signals used for CP monitoring.

Also, Table 1.4 gives the list of signals that can be used for different CP fault types (Dister, 2003). Once the choice of the signals to be acquired is made, the next step would be to extract useful information from the signals.

1.10 Pre-Processing of Signals

Extracting useful information from the raw-signal obtained from the sensors is called as the *pre-processing* of the data. This is an important step because it is imperative that the

unnecessary, redundant information is removed from the data and only useful information is extracted from it.

Table 1.3: Typical signals used to monitor various mechanical/ electrical components

	Vibration	Humidity	Temperature	Pressure	Acoustic	Thermography	Motor current	Insulation resistance	Electrical capacitance	Electrical Inductance
Pumps										
Valves										
Motors/ Fans										
Heat exchangers										
Steam turbines										
Electronic equipment										
Cables										
Pump seals										
Piping/ Structures										
Compressors										

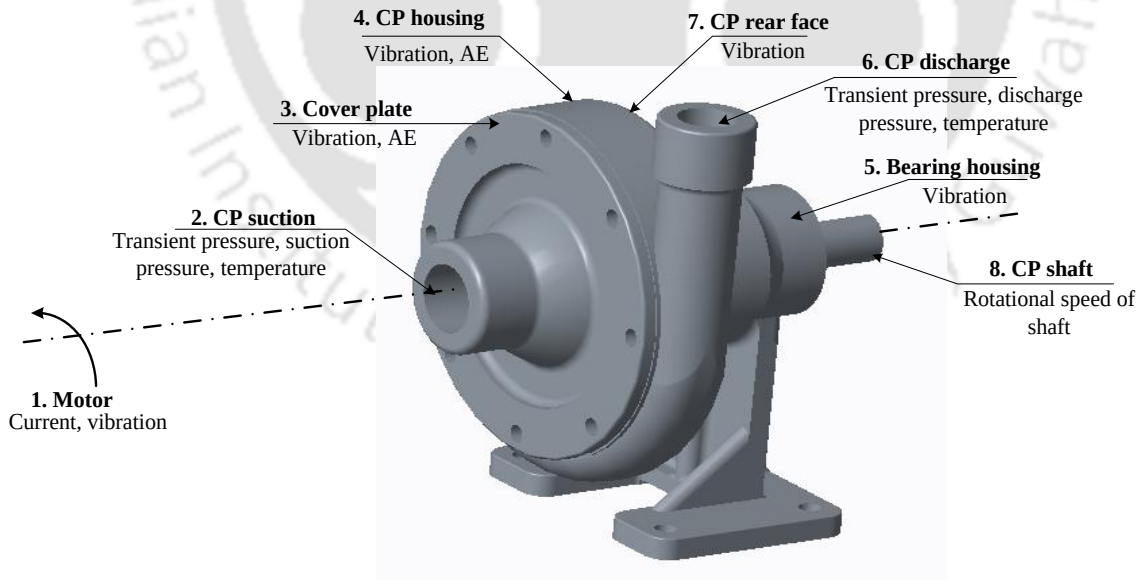


Figure 1.8: CP sensor locations for acquiring different signals

The idea is that the state of the system must be legibly reflected in the pre-processed data.

Some of the basic elements of data pre-processing are (Verma et al., 2013),

1. Filtering (remove unwanted noise from original signal),
2. Clipping (removing regions of no significant information from the signal),
3. Normalization (reducing large variations in data),
4. Smoothing (removing distortions in the signals).

The raw signals acquired from the CP are in general in time-domain. They can, however, be used in the frequency domain or can be transformed using wavelets to the time-frequency domain. The selection of the domain depends on the nature of the faults and how best their characteristics can be obtained (Jardine et al., 2006). Comprehensive reviews on the application of wavelets for the fault diagnosis of machinery are given in refs. (Peng and Chu, 2004; Yan et al., 2014).

Table 1.4: Typical signals used to monitor various mechanical/ fluid faults in CPs

Fault type	On-line detection signal(s)
Impeller cracks, worn gear, cracked mounting foot	Vibration
Cavitation, water hammer, blade wear, cracked housing	Vibration, pressure
Offset impeller air gap	Current, pressure, vibrations
Leakage	Pressure, current
Bearing defects	Vibration, thermal
Plugged orifice	Pressure, current

1.11 Methods of CBM Analysis

There are different approaches to CBM, which can be listed as,

1. Visual inspection of sensor data methods
2. Physical-model based methods
3. Expert system based data driven methods

In the subsequent section the application of these methods for the fault diagnosis of the CPs is explained.

1.12 CP Fault Diagnosis Using CBM Techniques

The usual techniques used for the health-monitoring of CPs for various faults are given in Table 1.5 (Dister, 2003). Most of these techniques use evaluation of geometric or hydraulic conformity through manually inspecting the respective components. However, these methods are time-consuming and may not be very beneficial for online monitoring of CP systems. The alternative is to use sensor data to understand the absence or presence of the faults.

Ning et al. (2011) presented a discussion on the usage of various hardware-software combinations to acquire data from a CP system. McKee et al. (2014) reviewed the various state of the art techniques used in the fault diagnosis of CPs.

1.13 Visual Inspection of Sensor Data

Visual inspection of sensor data is a classical approach to the fault diagnosis. It is generally done either by observing the spectral patterns of the acquired signals or by extracting some useful statistical features from the signals and studying their variation over time.

Table 1.5: Conventional techniques for diagnosis of some mechanical/ fluid faults in CPs

Faults	Conventional detection techniques
Impeller cracks	Visual inspection, fluorescent dye testing
Cavitation	Air bubbles in flow stream, rattling sound
Water hammer	Fluid surging acoustic sound
Offset impeller air gap	Poor performance
Blade wear	Excessive vibration, acoustic noise
Leakage	Loss in CP efficiency
Bearing defects	Visual inspection, seized bearing
Bent or broken rotor	Inoperable CP, rotor impeller rubbing
Worn gear	Loss in performance, excessive vibrations
Cracked mounting foot	Rotation during start, excessive vibrations
Cracked housing	Visual inspection, fluorescent dye testing, fluid loss, X-ray

1.13.1 Visual Inspection Using Spectral Studies

Most faults in a mechanical system have specific fault signatures manifesting in different forms based on the type of signal extracted. Many industrial veterans use an experience based visual inspection of these fault signatures to evaluate the faults in the system and/or the severity of the fault. For example, using vibration signatures, the prominent fault frequencies are the rotational frequencies (1x, 2x, 3x, etc.), BPFs, the line current frequencies (LCF) and the natural frequencies (1n, 2n, etc.). To find out the natural frequencies of the set-up under study, the installation is subjected to an impact test. The CP is hit with an impact hammer from different directions and the accelerometer data is collected. FFT of the accelerometer's

time-domain data reveals the natural frequencies of the installation. Three trials are performed for each impacting direction. Based on the observations, the natural frequencies are tabulated. The three principal harmonics of the BPF are predominant. In the subsequent discussions, a case study of a CP driven at 30 Hz speed using a 3-phase induction motor is discussed. The different fault frequencies are shown in **Error! Reference source not found..**

Whenever there is a flow instability associated bubble formation in the CP, the amplitude of axial vibrations at the BPF increases. Figure 1.9 shows the typical power spectrum plot of the axial vibration from a good pump. This is plotted for 30 Hz operating speed. It can be observed from the graph that the significant peaks are at different harmonics of the rotational speed, i.e., LCFs and BPFs. Figure 1.10 shows a typical vibration power spectrum plot of a clogged CP. Blockage in flow results in bubble formation. These bubbles rotate along with the impeller vanes, and cause a ‘rotating cavitation’ phenomenon. Because of this they result in increased blade pass frequency (Čudina and Prezelj, 2009). Now, if the BPF amplitudes of both Figure 1.9 and Figure 1.10 are compared, it can be observed that though the first and the second BPFs have similar magnitudes, at the third harmonic of the BPF (at 450 Hz) in the axial vibration of the clogged CP is much higher than the health pump, clearly indicating the effect of bubble formation.

Table 1.6: CP fault frequencies

Fault frequency	Value
1x, 2x, 3x	30 Hz, 60 Hz, 90 Hz
1LCF, 2LCF	50 Hz, 100 Hz
1BPF, 2BPF, 3BPF	150 Hz, 300 Hz, 450 Hz
1n, 2n, 3n	107 Hz, 125 Hz, 322 Hz

Similarly, the axial vibration power spectrum of a CP with cracked impeller can be compared with a clogged pump with the cracked impeller in Figure 1.11 and Figure 1.12. Observations similar to that of the previous two figures can be made. Whenever cracks on the impeller develop they increase the transverse vibrations of the pump due to the unbalance effects. Figure 1.13 shows the vibration power spectrum of a healthy pump in the vertical transverse direction. Figure 1.14 shows the vertical transverse vibration power spectrum plot of a pump with the cracked impeller. On comparing both the graphs, it can be observed that the 1BPF and 2BPF amplitudes have increased when the pump develops cracks. Also, the overall vibration level has also increased significantly. Note that all the plots presented here are not taken from any work in literature and are given based on the experimental observations made in the present research. Here the impeller defects are given by cutting notches on each impeller vane, the blockage faults are created by clogging the flow into and out of the CP using a modulating valve.

Similar to the case study presented, many researchers have reported fault diagnosis of various CP conditions using spectral studies of vibration signatures (Sinha and Rao, 2006; Tan and Leong, 2008; Mohanty et al., 2012; Abdulkarem et al., 2014; Hamomd et al., 2017), acoustic emissions (AE) (Alfayez et al., 2004; Alfayez and Mba, 2005; Alfayez et al., 2005), motor current signals (Mohanty et al., 2012; Stopa et al., 2014; Luo et al., 2015; Alabied et al., 2017; Lima et al., 2017), instantaneous angular speed (IAS) (Al-Hashmi et al., 2004), noise (Chudina, 2003; Chini et al., 2005; Čudina and Prezelj, 2009), and pressure pulsations (Albraik et al., 2012; Lu et al., 2016).

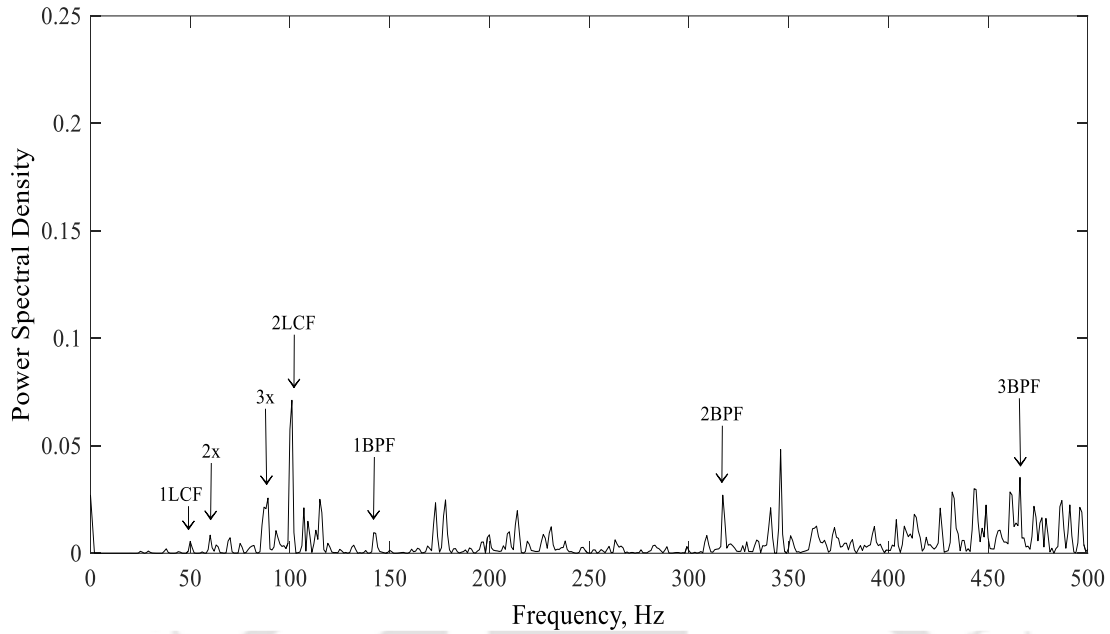


Figure 1.9: Axial vibration power spectrum of a healthy pump at 30 Hz operational speed

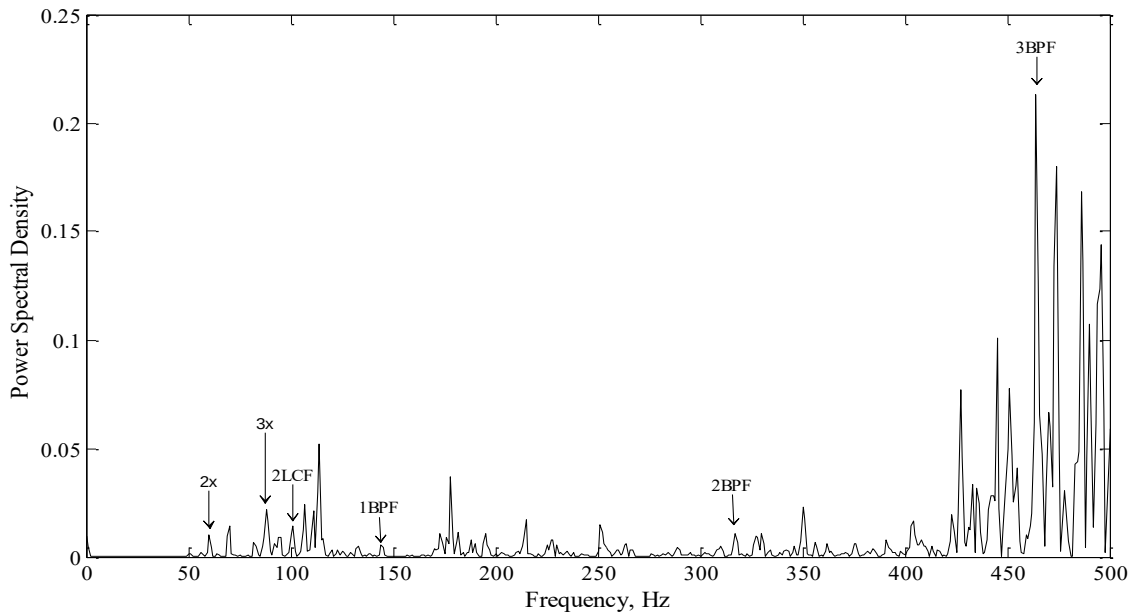


Figure 1.10: Axial vibration power spectrum of a clogged pump at 30 Hz operational speed

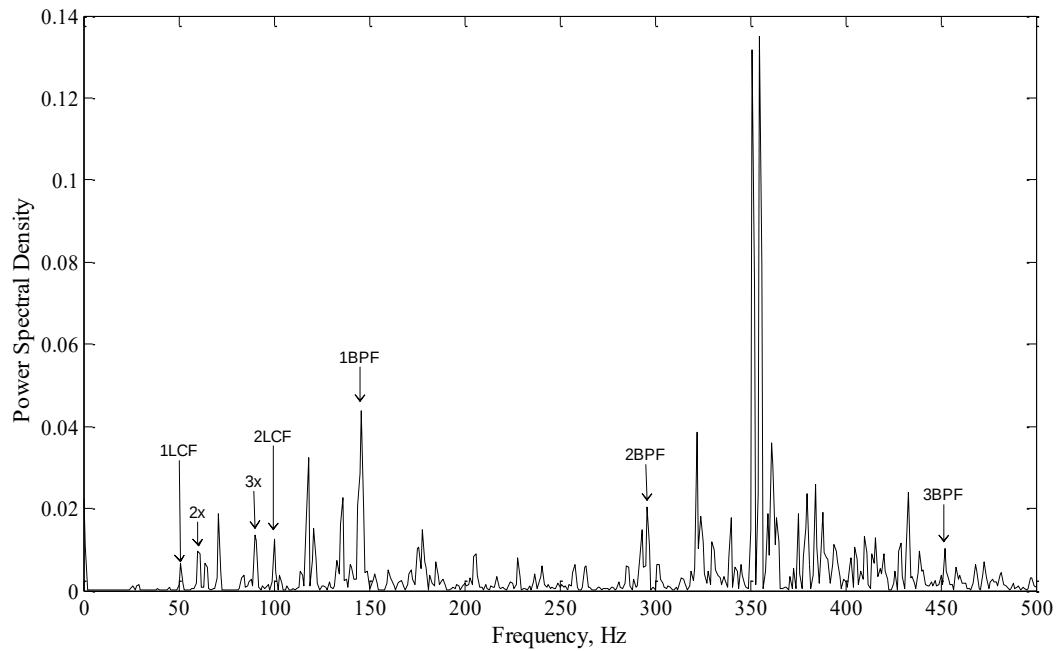


Figure 1.11: Axial vibration power spectrum of a pump with cracked impeller at 30 Hz operational speed

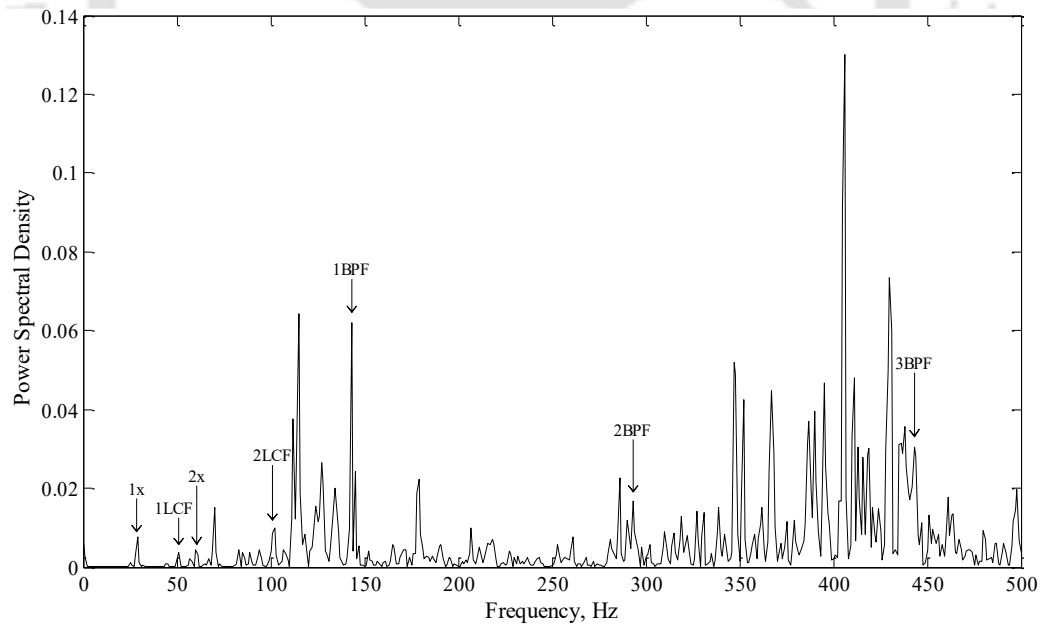


Figure 1.12: Axial vibration power spectrum of a clogged pump with cracked impeller at 30 Hz operational speed

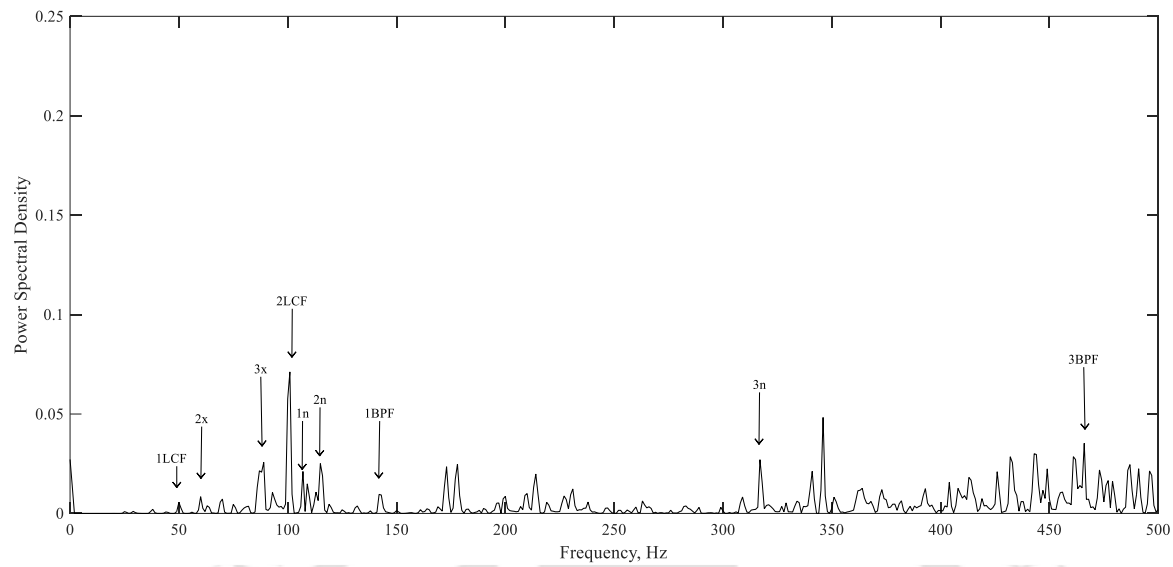


Figure 1.13: Vertical transverse vibration power spectrum of a healthy pump at 30 Hz operational speed

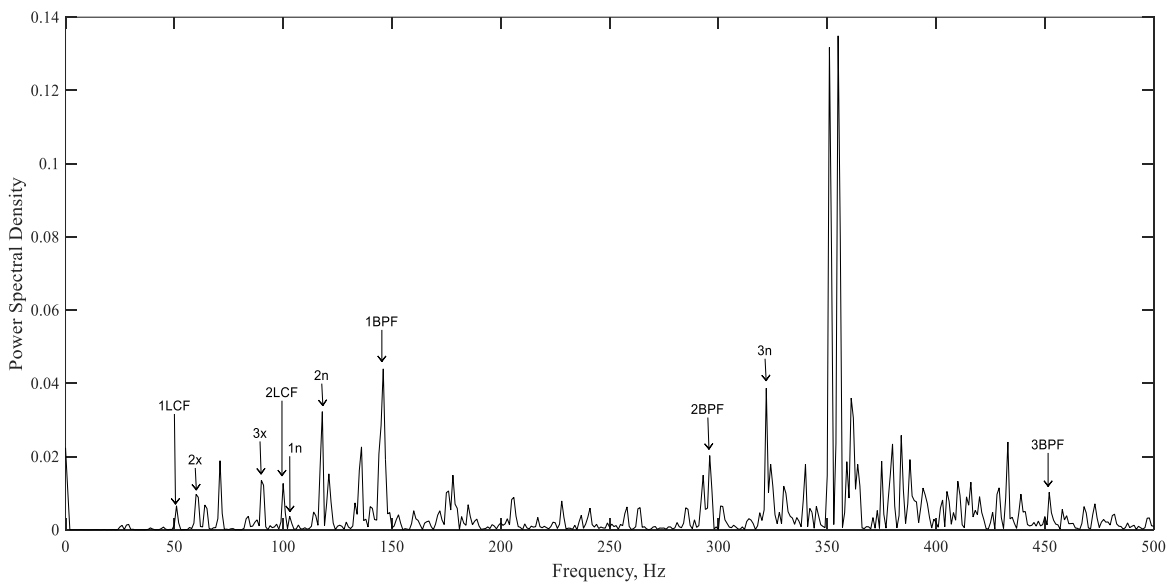


Figure 1.14: Vertical transverse vibration power spectrum of a pump with cracked impeller at 30 Hz operational speed

Vibration based spectral analysis

Sinha and Rao (2006) presented a vibration based fault diagnosis of CP in a typical industrial application. Modal analysis of the CP system was done to comprehend the dynamics of the entire assembly. The main reason for bearing failure was found to be the resonance of the bearing pedestals with 2x component of the response during CP operation due to the non-linear interaction between the CP foundation and the concrete floor. Tan and Leong (Tan and Leong, 2008) presented an experimental investigation of cavitation condition in a CP. Vibration spectra from various locations on CP were captured, and envelope spectra were extracted for various faults considered. It was observed that during normal operation the peak of vibration spectrum coincident to the CP operating speed was more pronounced. However, when cavitation was present the half order, sub-harmonic of blade pass frequency (BPF) was showing.

Abdulkarem et al. (2014) applied vibration time-index parameter and power spectrum analysis to identify different the severity levels of CP impeller cracks. They observed an increment in amplitude of the vibration time-index parameter and power spectrum amplitude with the increase in the severity of the impeller defects. Hamomd et al. (2017) developed a modulation signal bispectrum (MSB) approach to extract characteristic features of bearing and impeller faults in low-frequency range using vibration signals.

AE spectral analysis

Alfayez et al. (2004), Alfayez and Mba (2005), and Alfayez et al. (2005) investigated the usefulness of acoustic emission (AE) to detect the early cavitation of CPs. Before 3% head drop criterion, a surge in AE root mean square value was observed. Also, when the NPSH

value was high (where cavitation is known to develop) a substantial increase in the AE value was found. The authors also suggested that the position of the AE sensor influenced the usefulness of data acquired. A sensor in the vicinity of the impeller was said to capture the emission better.

Motor current spectral analysis

Mohanty et al. (2012) described the use of vibration and motor current signature analysis to identify various types of impeller defects in CPs. It was found that with the increase in the severity of the impeller defects the amplitude of sidebands increased. Also, the RMS value of the vibrations was found to increase with the fault severity. The current spectrum was used to identify the running speed of the CP, and the BPF could be identified by plotting the demodulated current spectrum. Stopa et al. (2014) presented the effectiveness of load torque signature analysis (LTSA) in the identification of cavitation in a CP. The LTSA employs the motor current signals to estimate the CP torque developed. This also gives information about the occurrence or severity of cavitation from the spectral analysis. The LTSA was also found to have a strong correlation with the suction pressure spectra. Lima et al.(2017) also used motor-current LTSA to evaluate cavitation condition in a CP.

Luo et al. (2015) employed the motor current signature analysis to understand the cavitation condition in a CP. It was found that the RMS value of the current fluctuated with different flow patterns. Apart from this, noise level, stator current spectrum, rotor speed, noise distribution, and BPF components showed to be useful features in cavitation detection. Alabied et al. (2017) used the motor line current data to diagnose various CP faults including, bearing inner race faults, bearing outer race faults and impeller defects. Intrinsic time-scale

decomposition (ITD) was adapted to process motor current signals. The diagnosis was accomplished by combining the RMS values of the first proper rotation component (PRC) with the raw signal RMS values.

IAS spectral analysis

Al-Hashmi et al. (2004) presented a cavitation detection technique using the instantaneous angular speed (IAS) measurement. The normalized amplitude at the 3x speed was found to be an excellent indicator to detect the beginning of cavitation and to enumerate the severity of it.

Noise spectrum analysis

Chudina (2003) experimentally established that the inception of cavitation in a CP could be found out by inspecting its noise spectrum. The author showed that the commencement of cavitation could be observed as a discrete frequency tone at $\frac{1}{2}$ BPF on the noise spectrum and also the difference in noise level between the formative stage and fully developed cavitation was around 12 to 20 dB. The author suggested that this tool could also be useful to determine the NPSH_R for a CP operation. Chini et al. (2005) presented a noise spectrum based technique for the recognition of cavitation faults in the CP. The authors observed that at frequencies equal to the odd multiples of the number of vanes in the CP there is shoot in noise level associated with cavitation.

Cudina and Prezelj (2009) explored the mechanism involved for the generation of discrete frequency in the audible noise spectrum with the inception of cavitation. For experimentation measurements including, the sound pressure level in the surrounding air, underwater acoustics and structural vibration were made. It was observed that the frequency tone was a result of

vibrations (modes) caused by collapse and bombardment of bubbles on the inner surfaces of the CP.

Pressure spectrum analysis

Albraik et al. (2012) observed that there was a change in the pressure fluctuations of a CP with different impellers faults and at different flow rates. The parameters of the faults that were changed on the impeller were: the depth of the dents and the number of dents. The results suggested that the CP vibration level amplified with the increase in flow rates, and was dissimilar for each of the defective impellers though used in geometrically similar CPs.

Lu et al. (2016) presented a numerical and experimental simulation to estimate the cavitation in a CP. The degree of CP cavitation was monitored using the suction and discharge pressure pulsations in the CP. It was found that both these parameters were good indicators of the formation of cavitation; however, suction pressures very more sensitive. It was found that the characteristic frequency of CP suction pressure pulsations was generated around 30 Hz for severe cavitation.

1.13.2 Visual Inspection Using Statistical Methods

Instead of spectral inspection of the signal data, researchers have also worked on observing patterns of the features extracted from these signals.

Kamiel et al. (2005) used principal component analysis (PCA) fault monitoring method with Hotelling's T^2 statistic to identify impeller fault. Vibration signature was used to extract the features. The accelerometer at the volute casing gave the most superior result for the fault detection. Halligan and Jagannathan (2011) used the linear and kernel PCAs to isolate and

detect various CP faults such as the seal failure, sensor failure, impeller failure and filter clog. The various faults were forming distinct clusters when the principal component values were plotted on x and y-axes. However, the performance using Gaussian kernel PCA was found to be better than the linear PCA for this system. Similarly, the authors' research team Soylemezoglu et al.(2011) used Mahalanobis-Taguchi based multi-sensor system to identify the faults mentioned above in the CP. In this Mahalanobis distance was used for the fault feature clustering and the change of this value was used for the prognostics and diagnosis.

McKee et al. (2015) developed a new method of vibration cavitation detection based on adaptive octave band analysis, principal component analysis, and statistical metrics. Threshold limits for different cavitating conditions were stated, and they were tested on a real industrial data.

1.13.3 Conclusions - Visual Inspection for CP Fault Detection

A comparison of different works done by researchers in the area of CP fault diagnosis using visual inspection methods are tabulated in Table 1.7. Visual inspection techniques are instrumental in identifying faults when their characteristic frequencies are known. However, there are four significant disadvantages of these methods,

- Highly skilled and experienced professionals are necessary to monitor the signal
- It is challenging for a human being to monitor the signal continuously and may be subject to human errors
- Any change in the operating speed of the CP the spectrum pattern would change
- Proper filtering techniques need to be employed, to keep the spectrum immune to noise.

1.14 Mathematical Modeling

Mathematical modeling of a system involves systematic representation of the physical attributes of the system so that it imitates the actual operating system as closely as possible. To simplify the system modeling, many assumptions are taken into consideration. The mathematical models usually use stiffness (spring elements), dampers and distribution of loads or point loads (masses). For instance, a four-wheeler can be modeled as a mass being suspended on four springs and four dampers. Thus, giving rise to a second order differential equation to describe the system. Even a system as complex as a human body can be modeled mathematically (Yang and Lövsund, 1997). The advantage of mathematical modeling techniques is that they give a deeper understanding of the system functioning. Some reviews on different mathematical modeling techniques employed in fault diagnosis of machinery are given in refs. (Frank, 1996; Isermann and Ballé, 1997; Isermann, 2005).

To develop a precise system model, a lot of experience and expertise relevant to the system is required. This would ensure that the model is neither too simplified nor is it left highly complex. To validate these models sizeable experimental data sets are used. Model-based fault detection techniques rely on the generation of residuals also called as fault indicators. They are acquired as deviations in individual attributes of the process found from process variables, like the state variables, or coefficients or assessed bounds. For this effect, data acquired from the system process is measured against the data supplied by models demonstrating the benchmark condition. The subsequent step involves residual estimation to decide about fault(s) presence. Methods like threshold based judgment, statistical decision theory, fuzzy decision making, pattern recognition or neural networks are employed to identify faults.

Table 1.7: Various visual inspection techniques for classification of CP faults

Reference	Signal type	Faults detected	Method used
Chudina, 2003	Noise	Cavitation	Noise spectrum
Alfayez et al., 2004, Alfayez and Mba, 2005, Alfayez et al., 2005	AE	Cavitation	AE RMS
Al-Hashmi et al., 2004	IAS	Cavitation onset and severity	IAS
Chini et al., 2005	Noise spectrum	Cavitation	
Kamiel et al., 2005	Vibration	Impeller defects	Hotelling's T^2 statistic
Tan and Leong, 2008	Vibration spectra	Cavitation	Envelope spectra
Čudina and Prezelj, 2009	Noise	Cavitation	Noise spectrum
Halligan and Jagannathan, 2011	Inlet pressure, inlet temperature, vertical and lateral acceleration, outlet pressure, temperature and flow	Seal failure, sensor failure, impeller failure and filter clog	Linear and kernel PCA
Mohanty et al., 2012	Vibration, motor current signature analysis	Impeller defects with varying severities	Vibration and current spectrum
Abdulkarem et al., 2014	Vibration	Impeller cracks with varying severity	Time vibration index parameter and power spectrum
Stopa et al., 2014	LTSA	Cavitation inception and severity	Torque, flow rate and frequency spectra
McKee et al., 2015	Vibration	No cavitation, incipient cavitation, full cavitation	Octave band analysis, PCA
Luo et al., 2015	Motor current signature analysis	Cavitation	RMS value of the current, stator current spectrum, noise level, noise distribution, rotor speed and vane pass frequency components
Lu et al., 2016	Suction and discharge pressure pulsations	Cavitation stages	Pressure spectrum
Alabied et al., 2017	Motor current	Bearing inner race faults, bearing outer race faults and impeller defects	ITD, PRC
Lima et al., 2017	LTSA	Cavitation	Torque, flow rate and frequency spectra
Hamomd et al., 2017	Vibration	Impeller blockage and bearing defects	MSB

The accepted ranges of the residual values need to be judiciously made based on experience to give rise to minimize false alarms (Harihara and Parlos, 2008b). Every system has its model, and there cannot be one model that can be applied to many systems. A typical mathematical model based fault diagnosis system's flowchart is given in Figure 1.15.

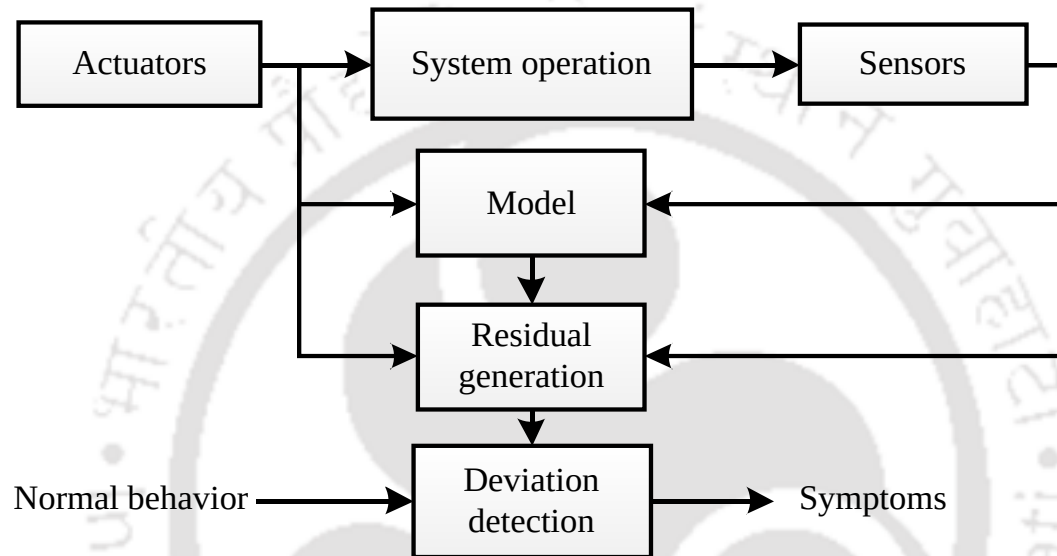


Figure 1.15: Flow chart of a model based fault detection system

The drawbacks of mathematical modeling techniques are, (1) these types of models are of high cost and component sensitive (McKee et al., 2014), (2) it is challenging to build a physical model that integrates all characteristics of a system and is valid over a wide operating range of the system (Wolfram et al., 2001a; Kallesoe et al., 2004; Peng et al., 2010), (3) while building a model for compound nonlinear, time-varying systems it is remarkably demanding to detect structural changes in the system and to obtain acceptable models for this purpose, (4) the available model is expected to represent benchmark conditions, and the impact of a deviation from these conditions on the model outputs is tough to predict (Du et al., 1995).

1.14.1 Mathematical Modeling for CP Fault Detection

CPs are majorly made up of the stationary and rotating mechanical components and fluid flow components. Hence, they involve complex fluid-structure interactions. They produce non-stationary signals, and therefore these systems are complicated to accurately model.

Isermann (1984) reviewed fault detection using methods based on modeling and estimation. Both determinate and indeterminate parameters of the model were used. Every parameter had a certain acceptance limit for the standard value. The model used data from the sensor to monitor essential parameters in the system. An alarm was activated if parameters were said to be outlying their tolerance limits. Later, the site and source of the fault were recognized. Next fault assessment was done. Lastly, a judgment was made whether to consider the fault tolerable or not. The developed technique was applied to a CP energized by a direct current motor. The drawback of this technique was that it indicated that leakage in a CP was not a defect. That is the model was sensitive to the process noise but insensitive to leaks.

In some of the approaches of modeling CP systems, blocks were created to identify the mechanical parts, fluid parts, etc. (Wolfram et al., 2001a; Kallesoe et al., 2004; Kallesoe et al., 2006; Samanipour et al., 2017). These methods involved developing models and parameters for each of the sub-systems and thereby calculating the residuals. The model developed by Wolfram et al. (2001a) consisted of three models for the CP, pipe, and mechanical subsystems, respectively. Separate neural-fuzzy models were used to denote corresponding normal state for each subsystem. To implement the fault detection residuals were calculated. The models were designed in such a way that the individual residuals were only sensitive to the used measurements and the underlying sub-process, and were insensitive

to the faults in measurements or other sub-processes. The faults that were considered are the sensor faults, gap losses, obstruction, cavitation, bearing friction and impeller defects.

Kallesoe et al. (2004) developed a model by dividing the CP system into four sub-systems where each subsystem was sensitive to a subset of faults considered. The developed model was based on the structural analysis (to identify the subsystems), analytical redundancy relation (ARR) and observers design. One of the residual observers was designed for each of the four subsystems. These residuals were further used for the identification/ isolation of the faults. On applying the developed methodology to an industry benchmark condition, it was observed that the algorithm was successful in identifying four out of five faults in the CP. Where, the five faults considered on the CP were clogging, bearing faults, leakage, cavitation and dry run. The signals that were acquired for the detection were shaft torque, pressure head developed by the CP and the flow through the CP. The method was however sensitive to the change in the operating parameter of the CP.

In another work, Kallesoe et al. (2006) developed a mathematical system model of the CP which needed only two inputs/ variables, and they are the electrical motor measurements (stator voltages, motor current) and the delivery pressure of the CP. The CP model was distributed into two sub-systems, the electrical system, and mechanical/ hydraulic systems. The faults accounted for were the CP clogging, friction generated because of either rub impact or bearing defects, degradation of CP performance because of cavitation, and dry run. Thus, the considered faults were identified using the ARR. A large residual implied a deviation of the system's actual behavior and that was expected by the developed model, suggesting a fault

in the system. Though this method detected the CP faults, the transient phases of the system created difficulties for the algorithm.

Samanipour et al. (2017) developed a nonlinear electro-pump model to diagnose cavitation defects in CPs. The model constituted three blocks, including blocks for the induction motor, mechanical CP parts, and hydraulic CP parts. The pressure and torque values were identified from the developed model, and their nonconformity to the performance curve of the CP was identified as the residual. The extracted features from the residuals were input to a self-organizing map neural network to identify the state of the system. The low cavitation was detected with 88% accuracy, while the high cavitation was identified with 96% accuracy.

Different other approaches used to model the CP system are given in refs. (Nordmann and Aenis, 2004; Harihara and Parlos, 2006, 2008b, 2008a). Nordmann et al. (2004) constructed a mathematical model of a rotor using the inertia, stiffness, damping, and loading function matrices. The input and output characteristics of the active magnetic bearings (AMB) on the CP were measured. The CP data, which were measured, included the rotor speed, CP flow rate, head developed, radial clearance of the piston seal and the impeller seal. From the AMB, the AMB force, pre-magnetization current, windings per pole pair, air gap, and the cross-sectional area of the pole were used. Using this data, the bearing force was estimated. Twenty-five quantifiable transfer functions created relating the inputs with the force excitation patterns worked as indications to the response of the system. Substantial differences from the model estimated parameters signified the presence of a fault in the system. Recognition of the fault was done by estimating the fault-symptom relations or by generating faults in the model

parameters and relating the faulty model parameters with the actual data. This method was tested on a CP with no fault, and dry-run faults.

A team of researchers developed a sensorless model for CPs using the motor current signature analysis. ‘Sensorless’ here meant that there was no add-on or intrusive sensor. None of CP or motor design attributes were used to develop the model. The developed model was successful in identifying various cavitation states in a CP with very few false alarms. The system model inputs used were different transformed signals processed from the voltages and currents, for example, the voltage level, voltage imbalance, etc. A function was constructed based on the time-varying inputs in the polynomial NARX form. The developed model constituted sampling the motor electrical signals, down-sampling the signals to lower frequency ranges, and later scaling these signals to per-unit values. At a specific CP operating state, the model projected the baseline system response. The residuals were generated by comparing this to that of the actual measured response of the system. When residual was found to be beyond a certain threshold fault was alarmed. Else, the system was considered healthy. Based on various experiments a 5% threshold was used in the study. They applied the aforementioned technique to detect the cavitation (Harihara and Parlos, 2006), successfully, identify varying severity of impeller cracks (Harihara and Parlos, 2008b) and detect motor faults and pump faults (Harihara and Parlos, 2008a). It was observed that the motor line current analysis based model could identify the faults in a shorter span when compared to the model using vibration parameters as an indicator.

1.14.2 Conclusions – Mathematical Modeling for the CP Fault Detection

Mathematical modeling of the CP system is very useful to get the physical feel of the processes taking place within it. However, from the survey, the following observations can be made,

- It requires a complex system model to define the CP system accurately
- The models developed are highly sensitive to the operating state (or change in operating state) of the CP
- Each fault in the CP needs to be modeled separately using separate functions
- Sometimes the model is sensitive to the noise
- Signals used and the threshold limit chosen to play a vital role in ensuring no false alarm generation
- It is complicated to model a combination of faults

1.15 Machine Learning Techniques

With the advent of high-performance computing and increased data storage capacities, industries are looking for ways and means to automate processes including maintenance. Machine learning and data-driven techniques utilize existing sensor data and perform fault detection based on various schemes. These techniques help in realizing quick and reliable fault detection with minimum human intervention. They, therefore, help in reducing the cost and improving the system performance.

For any successful application of machine learning algorithm, the pre-requisite is the availability of data. The machine learning algorithms employ a collection of features to represent the data. These features can have continuous values, discrete values or binary values.

In case, the examples are denoted by recognized labels (known correct outputs) then the

learning algorithm is said to be *supervised* (Kotsiantis et al., 2007). However, in *unsupervised* learning, the examples are not labeled (Jain et al., 1999). A brief review of machine learning algorithms is given in ref. (Dutton and Conroy, 1997).

Features play a vital role in understanding the fault characteristics and thereby enhance the performance of the machine learning algorithm. These features may be extracted from the signal in time domain (Zouari et al., 2004; Sakthivel et al., 2010a; Sakthivel et al., 2010b; Saberi et al., 2011; Sakthivel et al., 2012b; Sakthivel et al., 2014; Zhao et al., 2016; Azizi et al., 2017; Bordoloi and Tiwari, 2017; Kumar and Kumar, 2017), frequency domain (Perovic et al., 2001; Yuan and Chu, 2006; Zhao et al., 2010; Nasiri et al., 2011; Farokhzad et al., 2012; Yunlong and Peng, 2012; Xue et al., 2014) or in time-frequency space (Muralidharan and Sugumaran, 2012; Sakthivel et al., 2012b; Muralidharan and Sugumaran, 2013c, 2013b, 2013a; Muralidharan et al., 2014; Ebrahimi and Javidan, 2017; Kumar and Kumar, 2017) of the sensor data. Statistical features, like the mean, standard deviation, variance, etc. are extracted from these signals. The choice of features, feature subset selection, needs to be made judiciously. Feature subset selection involves removing unnecessary, redundant feature data, thereby reducing its dimensionality (Yu and Liu, 2004). Features, in general, can be relevant (influencing the outcome), irrelevant (having low or no influence on the outcome) or redundant (another can represent one feature) (Kotsiantis et al., 2006). Some of the linear and non-linear dimensionality reduction techniques used in the CP fault diagnosis are the principal component analysis (PCA), local linear embedding (LLE), Hessian LLE, kernel-PCA (kPCA), Laplacian Eigenmap, local tangent space analysis, manifold charting and maximum variance unfolding (MVU), etc. (Yuan and Chu, 2006; Sakthivel et al., 2014).

There are many techniques used to quantify the efficiency of a machine learning algorithm. One of those methods is usage of percentage classification accuracy. It is defined as,

$$\text{Accuracy} = \left(\frac{\text{number of correctly predicted data}}{\text{total number of testing data}} \right) \times 100\% \quad (1.18)$$

According to Kotsiantis et al. (Kotsiantis et al., 2006; Kotsiantis et al., 2007) supervised machine learning algorithms may be categorized as,

1. Logic based
2. Neural network based
3. Statistical learning based
4. Support vector machines

1.15.1 Logic Based Classifiers

Logic-based classification algorithms can be of many types, of which decision trees and rough sets are dealt in this section. A basic understanding of decision tree algorithm is given in (Murthy, 1998). Decision ‘trees’ are tree structures that sort examples based on their feature values. Here, each node signifies a feature value of a sample to be sorted, and every division (branch) denotes a value that the feature can take. Samples are sorted beginning from the origin/ root level depending on the value of features they assume. A general structure of the decision tree is given in Figure 1.16.

These algorithms are generally uni-variate as they branch on a unique feature value at every node. The popularly used decision tree algorithms are C4.5 (Quinlan, 2014) and J48

algorithms. Some of the researches that used decision trees for the classification of CP faults are: (Sakthivel et al., 2010a; Sakthivel et al., 2010b; Muralidharan and Sugumaran, 2013c, 2013a; Sakthivel et al., 2014).

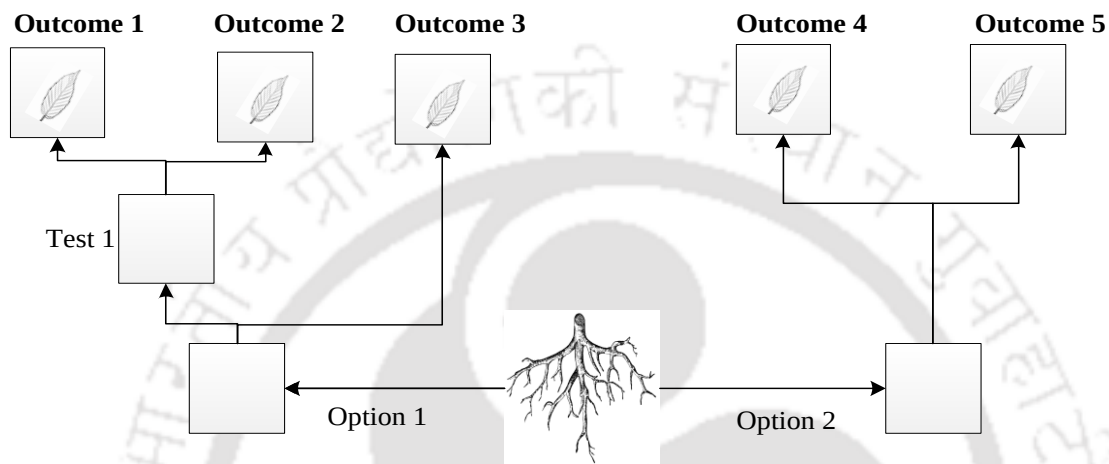


Figure 1.16: Structure of a decision tree algorithm

Sakthivel et al. (2010a) developed CP fault classification algorithm built on C4.5 decision tree algorithm. Faults, such as the bearing defects, seal faults, impeller faults, the combination of an impeller and bearing faults, cavitation was considered on the CP. Features, such as the standard error, standard deviation, sample variance, kurtosis, skewness, range, minimum, maximum and sum were used to define the nodes. The development of the decision tree comprised of four stages, i.e., the building phase, pruning phase, discretization of the continuous-valued attribute, and application of decision tree for feature selection. The algorithm was able to classify CP faults with 100% accuracy.

Muralidharan and Sugumaran used energy features extracted from various wavelet families from the discrete wavelet transform (DWT) (Muralidharan and Sugumaran, 2013c) to

construct the J48 classification algorithm. CP faults, like the healthy CP, cavitation, CP with bearing faults, CP with impeller defects, and CP with both bearing faults and impeller defects were considered. Classification accuracy of 99.84% was obtained using rbio1.5 wavelets.

Another rule-based classification technique is rough sets. Pawlak first developed the rough set theory as a mathematical tool to characterize imprecise knowledge (Pawlak, 2004; Pawlak and Skowron, 2007). The fundamental dissimilarity between a rough set and a classic set is the appearance of a ‘boundary region,’ as shown in Figure 1.17, where the uncertain elements exist (Chia-Chi and Kuo-Wei, 2012).

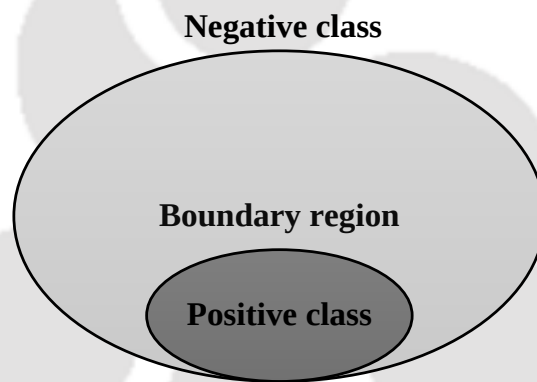


Figure 1.17: Rough set representation

Another tool to take care of imprecise knowledge is fuzzy-set. The chief difference between rough set and the fuzzy set is that the latter requires a membership function to account for a grade of elements contribution in it, whereas the rough set does not need that. An overview of fuzzy-logic-based classifiers is given in ref. (Bonarini, 1999). Fuzzy logic helps improve classification by permitting the use of overlying class definitions. Some researches that used fuzzy logic for the CP fault identification are (Perovic et al., 2001; Azadeh et al., 2010). Some

of the researches that used the rough set fuzzy theory for the CP fault classification are (Sakthivel et al., 2010b, 2012b; Muralidharan and Sugumaran, 2013b).

Perovic et al. (2001) classified complex combinations of mechanical and fluid faults in a CP using the fuzzy logic in conjunction with motor current spectra. The faults considered were discharge blockages, impeller faults, and the combination of discharge blockage and impeller defects. Features considered were the current amplitude, slip, accumulative noise, and amplitude of the shaft speed component. Membership functions were to take values based on their participation (i.e., the low, medium low, medium, medium-high, high). The combination of discharge blockage faults, the impeller faults and cavitation were classified with 83.7% accuracy, and the blockage and the cavitation with an accuracy of 61%.

Azadeh et al. (2010) used the fuzzy logic technique to classify various complex hydraulic and mechanical faults on the CP. The faults considered were the seal damage, misalignment, bent shaft, cavitation, bearing damage, plugged seal, dirty seal, rubbing, and the eccentric shaft. Various operating parameters, like the discharge pressure, $NPSH_R$, flow rate, brake horse power (BHP), vibration, efficiency, and temperature were used to generate fuzzy rules. The approach, when applied to a petrochemical industry, showed accurate fault identification.

In continuation to work in (Sakthivel et al., 2010a), Sakthivel et al. (2010b) compared the CP fault classification performance of two algorithms, viz. decision tree-fuzzy and rough-set fuzzy. The highest classification accuracy of 99.33% was obtained using the decision tree algorithm. Also in continuation to work in (Muralidharan and Sugumaran, 2013c), Muralidharan and Sugumaran (2013b) used the roughest-fuzzy method to classify CP faults.

Here also the features from the discrete wavelet transform (DWT) coefficients of various wavelet families were used. Classification accuracy of 99.84% was obtained.

1.15.2 Neural Network Based Classifier

Artificial neural network (ANN) based classifiers can classify data which are not linearly separable. A review of the work done on ANNs is given in (Zhang, 2000). The construction of a multi-layered ANN is given in Figure 1.18. It comprises some neurons (units) joined together with systematic pattern connections. These neurons are broadly classified into three layers, viz. the input layer, hidden layer(s) and the output layer. Note that there will be only one input and output layer, but there can be more than one hidden layers. Feed-forward ANNs allow signals to travel only in one direction, i.e., from input to the output layer. In general, it is challenging to decide the hidden layer numbers or the neurons numbers. Too few neurons may lead to poor generalization, and too many of them may lead to an overfit condition and ultimately make it difficult to search for the global optimum. Camargo and Yoneyama gave a good argument about this in ref. (Camargo and Yoneyama, 2001).

ANN construction relies on three significant aspects, (1) input and activation functions of the unit, (2) the architecture of the network and (3) the weight of each connection (w_{ij}, w_{jk}) (Kotsiantis et al., 2006). In case, the (1) and (2) aspects are fixed, the values of existing weights can determine the performance of the neural network. These weights are set initially to some random values, and then the training set instances are recurrently exposed to the ANN. The outcome is compared to the anticipated output for the considered instance. After that, the weights are attuned marginally in a way that would bring the outcome closer to that of the anticipated output.

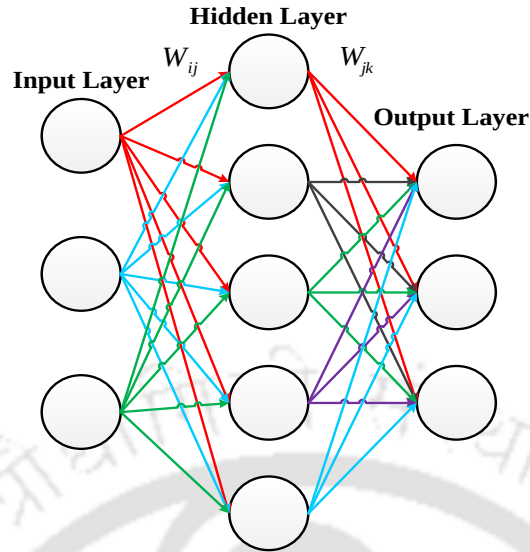


Figure 1.18: ANN architecture

Many pieces of research have successfully employed ANNs and their modified versions to classify CP faults (Zouari et al., 2004; Wang and Chen, 2007b, 2007a, 2009; Zhao et al., 2010; Nasiri et al., 2011; Saberi et al., 2011; Farokhzad et al., 2012; Zhao et al., 2016; Azizi et al., 2017; Shervani-Tabar et al., 2017).

Zouari et al. (2004) developed a CP fault identification system based on neural network (NN) and neuro-fuzzy techniques. CP faults considered were the cavitation, partial flow, loosening of front or rear CP attachments, misalignment, and air injection. The partial flow condition was considered with varying severities. Also, different percentages of air-injection were considered. To handle the problem of dealing with unlearned classes neuro-fuzzy networks were considered.

Wang and Chen (2007a) developed a diagnosis method to classify CP faults at early stages, with frequency based symptom parameters using wavelet transform (WT), rough sets, and

partially-linearized – NN (PNN). The WT was used to obtain features from the optimal frequency range. The rough sets were employed to acquire the identification knowledge. The PNNs learned from the thus acquired knowledge. The considered faults were cavitation, impeller damage, and misalignment. Each of these faults classified with high accuracy (~98%) when the features were extracted from their specific optimum frequency range. In another work, Wang and Chen (2007b) proposed a sequential fault diagnosis method using the PNN. In this technique, non-dimensional symptom parameters (NSPs) were defined from the time-domain data. To identify the sensitivity of non-dimensional symptom parameters to detect the faults, a synthetic detection index was proposed. The proposed method was then tested to identify the aforementioned CP fault conditions. In continuation, Wang and Chen (2009) developed a fault diagnosis approach based on the DWT, rough sets and PNN. Rbio 2.8 wavelet was used to extract fault information from each of the faulty conditions on the CP. NSPs were calculated for the recomposed signal at each level. The PNN could then denote intricate relationships between fault types and symptoms.

Nasiri et al. (2011) developed a method to classify different severities of cavitation using vibration signature and NN. Also, a study was presented regarding the number of sensors to be used and their optimal position. The study reveals that in case of availability of a single sensor, it may be placed in the radial location on the CP if two are available then they may be placed at the radial and back locations of the CP. To minimize the error of fault detection further a vibration sensor at the radial, back and front locations may be used. Farokhzad et al. (2012) diagnosed seal defects, impeller faults, and cavitation in CP with 100% accuracy using features extracted from the vibration spectra and the ANN algorithm.

Sakthivel et al. (2012b) developed a rough set – fuzzy system to generate the rules based on statistical features and to classify the mentioned CP faults. The performance of the developed classifier was also compared with the fuzzy-antminer classifier and the multi-layer perceptron classifier. Classification accuracy of 97.5% was obtained using the developed methodology.

1.15.3 Statistical Learning Based Classifier

These techniques are identified by their characteristic probability model. According to this model, the probability of a sample belonging to a particular class is defined, unlike the standard classification problems. Bayesian networks (Jensen, 1996) and k nearest neighbors (instance-based learning technique) fall under this category of machine learning techniques. Usually, the learning of a Bayesian network is split into two tasks: (1) the learning of the direct acyclic graph (DAG) structure of the network, and (2) the estimation of its parameters. The probabilistic parameters in the form of local conditional distributions of the variable given its parents are encrypted into a set of tables. The joint distribution is calculated by multiplying these tables (Kotsiantis et al., 2006). In case there are too many features defining the datasets then it is highly involved to construct complex network structures (Bell et al., 2002).

Naïve Bayes network is a simplified Bayesian algorithm in which the DAGs are constructed based on a single parent, and there are several child nodes. Here there is a powerful underlying assumption of independence of all the child nodes with respect to their parent (Kotsiantis et al., 2006). The advantage of this algorithm is that it is computationally very less time consuming to train. However, not always are the features independent of each other, and that is when Naïve Bayesian networks may not work as good as other high-level machine learning

algorithms. Some of the researches that used Bayes Net and Naïve Bayes algorithms for the CP fault diagnosis are (Muralidharan and Sugumaran, 2012; Sakthivel et al., 2014).

Muralidharan and Sugumaran (2012) compared the CP fault classification performance using algorithms based on Naïve Bayes and Bayes Net. They used DWT based features to classify various CP faults, like the cavitation, bearing faults, impeller defects, and CP with coexisting bearing faults and impeller defects. In most cases, the Bayes net outperformed Naïve Bayes. They obtained 100% classification accuracy using the sym3 and rbio2.6 wavelets and the Bayes net algorithm.

Another machine learning algorithm in this category is a k nearest neighbors (kNN) technique, which is an example based learning technique. It is based on the idea that an example belonging to a particular dataset lies close to other examples having similar attributes. If each example is marked with its respective class label, then an unclassified example could be assigned the label of its nearest neighbor. kNN identifies k nearest neighbors to the unclassified example, and that class that occurs the most frequently is assigned.

Sakthivel et al. (2014) classified CP faults, such as the bearing faults, seal defects, impeller defects, a combination of the impeller and bearing faults, cavitation, using the decision trees, Bayes Net, Naïve Bayes and kNN algorithms. They later compared the performance of these algorithms. Various dimensionality reduction techniques (both linear and non-linear) were used to consolidate the features. The PCA-decision tree combination was found to give the best classification performance.

1.15.4 Support Vector Machine Classifier

Support vector machine (SVM) is one of the latest machine learning techniques based on statistical learning theory (Vapnik, 1995; Vapnik, 1998). An excellent review of applications of SVMs for machine fault diagnosis was given by Widodo et al. (2007). SVMs are the non-probabilistic, binary, and linear classifiers. They revolve around the concept of construction of a linear *hyperplane* to separate two classes of data as shown in Figure 1.19. The hyperplane's location and orientation are found out by solving an optimization problem to maximize the *margin*.

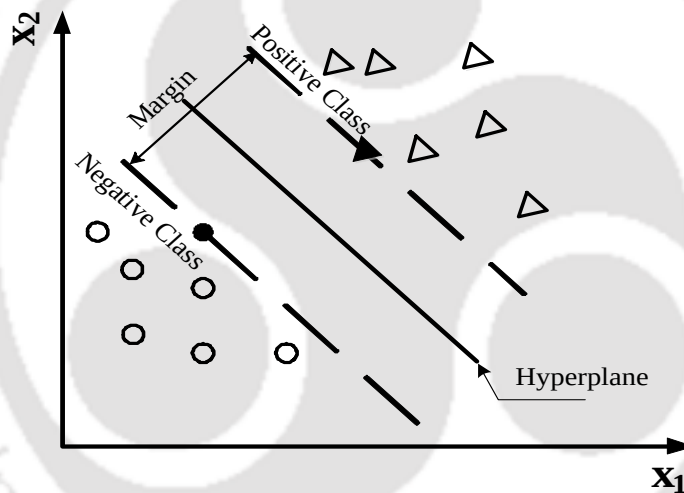


Figure 1.19: SVM classifier

SVM works well for the linearly separable data, however, when the data is not linearly classifiable, a kernel transformation to a greater dimensional space is carried out. In such a space the data can be linearly classifiable. The advantages of SVMs are that, (1) they require very fewer data to construct a reliable classifier (2) their training time is less and (3) because of the convex function optimization they find the global minimum. However, the disadvantage

is that classifier's performance majorly depends upon the values of SVM parameters and kernel parameters.

Many optimizations and cross-validation techniques have been used by researchers to construct an optimum SVM classifier. Various CP fault classification algorithms developed are given in refs. (Yuan and Chu, 2006; Zhao et al., 2010; Saberi et al., 2011; Sakthivel et al., 2012a; Yunlong and Peng, 2012; Azadeh et al., 2013; Muralidharan et al., 2014; Kumar and Kumar, 2017; Shervani-Tabar et al., 2017).

Yuan et al. (2006) used the multi-class SVM algorithm to classify various pump-rotor faults, such as the gear damage, structure resonance, vane rupture, shaft crack, bearing damage, bearing looseness, rotor part looseness, rotor eccentricity, rotor radial touch friction, body joint looseness, pressure pulse, cavitation, and shaft bend. They used the feature dimensionality reduction technique the PCA. They compared various multiclass SVM strategies and found that 'one to others' gave the maximum precession ratio and took the least training time.

Zhao et al. (2010) used neighborhood rough set models to select the features for the slurry CP fault classifications. Where the faults considered were leading and trailing edge damages of the CP impeller. The authors illustrated that using a shared neighborhood size for every feature may miscalculate a feature's degree of dependency. Therefore, the neighborhood rough set model was revised by setting different neighborhood sizes for various features. To check the effectiveness of feature selection, probabilistic neural networks, kNN and SVM algorithms were used. It was found that the average classification accuracy improved from 84.06%

(without feature selection) to 99.56% (with the aforementioned feature selection). The SVM and the kNN performed on par with each other.

Saberi et al. (2011) compared the performance of ANN and SVM to classify CP faults with noisy data. It was found that the SVM outperformed ANN in the fault classification. To extract suitable features from the vibration data Yunlong and Peng (2012) used the empirical mode decomposition (EMD) and intrinsic mode functions (IMFs). They classified CP faults, such as the unbalance, misalignment and looseness. The classification algorithm adapted was the least square support vector machine (LSSVM). They obtained a classification accuracy of 92.6%. Sakthivel et al. (2012a) compared the classification accuracy of gene expression programming (GEP) technique with other pattern recognition techniques, like the SVM, wavelet-GEP, and PSVM in classifying the same faults. It has been observed that both SVM and GEP outdo the rest of the methods, giving a classification accuracy of 99.93%.

Muralidharan et al. (2014) used CWT analysis in conjugation with the SVM to diagnose the aforementioned CP faults. They compared the performances of various wavelet families at different levels. The best classification accuracy was found with db8 wavelet. The accuracy presented was 99.84%. Azadeh et al. (2013) proposed the SVM based flexible algorithm with a hyperparameter optimization (optimization techniques used were the genetic algorithm (GA) and particle swarm optimization (PSO)) and ANN. A binary CP fault classification was attempted in standard as well as noisy states of the signals. Classification accuracy of 93.3% was obtained in both the considered states. The SVM with hyperparameter optimization gave better prediction performance than ANN.

Xue et al. (2014) diagnosed different severities of CP faults, such as the cavitation, unbalance and misalignment using an intelligent fault diagnosis technique. A statistic filter was employed to obtain the feature signals from the acquired vibration signals in an optimum frequency region. Also, NSPs were identified to represent the feature signals for differentiating CP fault types. The optimal classification hyperplane function obtained using SVM and NSPs was called the synthetic symptom parameter (SSP). Based on SSPs, the CP fault detection and the fault class identification was attempted based on the possibility theory and the Dempster-Shafer theory. The developed methodology identified high severity faults with good precession; however, the low and medium severity of faults could not be adequately identified.

Azizi et al. (2017) built an algorithm based on the generalized regression neural network (GRNN) estimated the severity of cavitation. Three states of the cavitation, namely, the developed cavitation, limited cavitation and no cavitation were considered. The EMD method was used to decompose original signals into IMFs. A hybrid feature selection algorithm based on artificial-bees algorithm was used to improve the accuracy of classification from 97.5 to 100%. Bordoloi and Tiwari (2017) attempted classification of different severities of suction blockages and cavitation faults in a CP at varying operating speeds of it. They found that the classification accuracy improved from 55.3 to 94.6% from the low to high speeds. The algorithm they adopted was SVM with the hyper-parameter optimization using the genetic algorithm (GA) and the artificial bee colony algorithm (ABCA).

Ebrahimi and Javidan (2017) presented an SVM based approach to classify CP faults including impeller defects and seal faults. The vibration signals acquired were decomposed

using discrete wavelet transform (DWT) to three levels using Daubechies wavelets. Classification accuracy of 96.67% was obtained. Tabar et al.(2017) developed an algorithm based on SVMs to identify different intensities of cavitation in a CP. Features used for the classification were selected based on their sensitivity towards the fault using compensation distance evaluation approach. Kumar and Kumar (2017) used the GA optimized SVM algorithm for the fault detection in CPs. The faults considered were the clogged impeller, broken impeller, bearing inner race and outer race defects. The features that are sensitive to the faults are acquired from the raw vibration signals and scale marginal integration graph (from CWT). An accuracy of 96.66% was found in the identification of various CP faults.

1.15.5 Conclusions – Machine Learning for CP Fault Detection

Machine learning techniques are found to be very useful in the fault diagnosis of CPs. They use minimum human intervention to identify the CP faults accurately. As it was presented, there are numerous machine learning algorithms to choose from. The choice of the algorithm depends upon the quantity of data available, the learning speed required, the amount of accuracy expected, the types of faults dealt with and last but not the least the comfort of the engineer to make use of the technique. A summary and comparison of different works done by researchers in the area of CP fault diagnosis using machine learning are tabulated in Table 1.8.

Table 1.8: Comparison of researches on application of machine learning to fault diagnosis of CPs

References	Faults considered	Signal type	Domain of signal	Machine learning algorithm used	Data sampling rate/sample length	Features considered	Multiclass	Accuracy obtained
Perovic et al., 2001	Discharge blockage faults, impeller defects and both discharge blockage and impeller defects	Motor current	FFT	Fuzzy logic	n/a	Slip, current amplitude, amplitude of the shaft speed component, accumulative noise	Yes	83.7%
Zouari et al., 2004	Loosening of front or rear CP attachments, misalignment, cavitation, partial flow, and air injection	Vibration	Time-domain	NN, neuro-fuzzy network	50kHz	Fischer, PCA	Yes	n/a
Yuan and Chu, 2006	Gear damage, structure resonance, rotor radial touch friction, shaft crack, bearing damage, body joint looseness, bearing looseness, rotor part looseness, pressure pulse, cavitation, vane rupture, rotor eccentricity, shaft bend	Vibration	FFT	SVM	n/a	Relative spectral amplitude, PCA	Yes	Precession ratio >95%
Azadeh et al., 2010	Seal damage, misalignment, bent shaft, cavitation, bearing damage, plugged seal, dirty seal, rubbing, eccentric shaft	Flow rate, discharge pressure, NPSH _R , BHP (Brake Horse Power), efficiency, vibration and temperature	n/a	Fuzzy logic	n/a	n/a	Yes	n/a

Table 1.8: Comparison of researches on application of machine learning to fault diagnosis of CPs (contd.)

References	Faults considered	Signal type	Domain of signal	Machine learning algorithm used	Data sampling rate/sample length	Features considered	Multiclass	Accuracy obtained
Sakthivel et al., 2010b	Bearing faults, seal defects, impeller defects, impeller and bearing faults, cavitation	Vibration	Time-domain	decision tree fuzzy, rough set fuzzy	24kHz/ 1024	Mean, standard error, median, standard deviation, sample variance, kurtosis, skewness, range, minimum, maximum, and sum	Yes	99.33%
Saberi et al., 2011 and Zhao et al., 2010	n/a Impeller leading edge and trailing edge damage	Vibration Vibration	Time-domain FFT	SVM and ANN Probabilistic neural network, k nearest neighbours, SVM	n/a 9kHz	n/a Noise based features	No Yes	100% 99.56%
Sakthivel et al., 2010a	Bearing faults, seal defects, impeller defects, impeller and bearing faults, cavitation	Vibration	Time domain	C4.5 decision tree	24kHz/ 1024	Standard error, standard deviation, variance, kurtosis, skewness, range, minimum, maximum, sum	Yes	100%
Nasiri et al., 2011	Cavitation with varying severities	Vibration	Discrete Fourier transform (DFT)	NN	10kHz	Kurtosis, crest factor	Yes	n/a
Muralidharan and Sugumaran, 2012	Bearing faults, Impeller faults, both bearing and impeller faults, cavitation	Vibration	Wavelets (DWT)	Naïve Bayes, Bayes Net	24kHz/ 1024	DWT coefficients	Yes	100%

Table 1.8: Comparison of researches on application of machine learning to fault diagnosis of CPs (contd.)

References	Faults considered	Signal type	Domain of signal	Machine learning algorithm used	Data sampling rate/sample length	Features considered	Multiclass	Accuracy obtained
Yunlong and Peng, 2012	Mass unbalance, rotor misalignment, foundation looseness	Vibration	FFT	LSSVM	800Hz/4096	EMD, IMF	Yes	92.6%
Farokhzad et al., 2012	Impeller defects, seal defects and cavitation	Vibration	FFT	ANN	n/a	Mean, standard deviation, variance, skewness, kurtosis, crest factor, slippage, root mean square mean, standard error, median, standard deviation, sample variance, kurtosis, skewness, range, minimum, maximum, and sum	Yes	100%
Sakthivel et al., 2012b	Bearing faults, seal defects, impeller defects, impeller and bearing faults, cavitation	Vibration	Time-domain	Roughset – fuzzy	24kHz/1024	mean, standard error, median, standard deviation, sample variance, kurtosis, skewness, range, minimum, maximum, and sum	Yes	97.5%
Sakthivel et al., 2012a	Bearing faults, seal defects, impeller defects, impeller and bearing faults, cavitation	Vibration	Time domain, Wavelets	Gene expression programming, SVM, wavelets – gene expression programming	24kHz/1024	mean, standard error, median, standard deviation, sample variance, kurtosis, skewness, range, minimum, maximum, and sum	Yes	99.93%
Azadeh et al., 2013	n/a	Flow rate, discharge pressure, suction pressure, vibration, velocity	n/a	SVM-GA, SVM-PSO, ANN	n/a	n/a	No	93.3%

Table 1.8: Comparison of researches on application of machine learning to fault diagnosis of CPs (contd.)

References	Faults considered	Signal type	Domain of signal	Machine learning algorithm used	Data sampling rate/sample length	Features considered	Multiclass	Accuracy obtained
Muralidharan and Sugumaran, 2013c, 2013a	Bearing faults, Impeller faults, both bearing and impeller faults, cavitation	Vibration	Wavelets (DWT)	Decision tree, J48	24kHz/1024	DWT coefficients	Yes	99.84%
Muralidharan and Sugumaran, 2013b	Bearing faults, Impeller faults, both bearing and impeller faults, cavitation	Vibration	Wavelets (DWT)	Roughset fuzzy	24kHz/1024	DWT coefficients	Yes	99.84%
Muralidharan et al., 2014	Same as above	Vibration	Wavelets (CWT)	SVM	24kHz/1024	CWT coefficients	Yes	99.84%
Xue et al., 2014	Varying severities of cavitation, unbalance and misalignment	Vibration	Frequency – domain	SVM	50kHz	NSP	Yes	0.9649/1
Sakthivel et al., 2014	Bearing faults, seal defects, impeller defects, impeller and bearing faults, cavitation	Vibration	Time-domain	decision tree, Bayes Net, Naïve Bayes and kNN classifiers		PCA, LLE, LTSA, HLLE, Isomap, kPCA, Laplacian Eigenmap, Manifold chart, MVU	Yes	100%
Zhao et al., 2016	Inner race wear, outer race wear, roller wear, and impeller wear	Vibration	Time-domain	Deep learning, softmax regression	10.24 kHz	n/a	Yes	>99%

Table 1.8: Comparison of researches on application of machine learning to fault diagnosis of CPs (contd.)

References	Faults considered	Signal type	Domain of signal	Machine learning algorithm used	Data sampling rate/ sample length	Features considered	Multiclass	Accuracy obtained
Azizi et al., 2017	No cavitation, limited cavitation and developed cavitation	Vibration	Time-domain	GRNN	16 kHz/ n/a	RMS, mean, variance, standard deviation, skewness, kurtosis, crest factor, upper bound, lower bound, range, shape factor, margin factor, impulse factor, entropy, zero crossing rate	Yes	100%
Bordoloi and Tiwari, 2017	Varying severities of suction blockages and cavitation	Vibration	Time-domain	SVM-GA and SVM-ABCA	20kHz/ 2000	Standard deviation, skewness, kurtosis	Yes	94.6%
Ebrahimi and Javidan, 2017	Impeller faults and seal defects	Vibration	Wavelets (DWT)	SVM	n/a	Mean, standard deviation, sample variance, RMS, crest factor, skewness, slippage, kurtosis, fifth and sixth central moments	Yes	96.67%
Shervani-Tabar et al., 2017	5 levels of cavitation	Vibration	Time-domain	SVM and ANN	n/a	Mean, peak, root mean square, standard deviation, kurtosis, skewness, kurtosis, crest factor, clearance factor, shape factor, impulse factor	Yes	97%
Kumar and Kumar, 2017	Broken impeller, clogged impeller, bearing inner race defects and bearing outer race defects	Vibration	Time-domain and CWT	SVM	70kHz/ 7000	Kurtosis, skewness, mean, RMS, variance, peak, impulse factor, shape factor, crest factor, scale corresponding to highest energy, ration of peak energy to total energy	Yes	96.66%

Despite so many advantages offered by machine learning approaches, there are a few drawbacks, and these are,

- In the case of supervised learning strategies, the pre-requisite is the availability of labeled fault data. However, in the most of industries many times either they do not have any data available, or even if they have the data, they do not know which fault it belongs to. In such circumstances, visual inspection techniques serve to be better options than supervised machine learning. Otherwise, unsupervised machine learning techniques may be used, like, Zhao et al. (2016) presented an unsupervised self-learning fault diagnosis method using deep learning to classify various CP faults including bearing faults (i.e., the inner and outer race wear, roller wear) and impeller wear. To demonstrate the robustness of the feature recalculation and effective high-level feature extraction, the noise was introduced in the input data of every auto-encoder. Classification accuracy of over 99% was found.
- The classifier designed for one system may not apply to another system, which is of different geometry.
- When designing a highly sophisticated machine learning algorithm, it may be difficult to maintain and debug it.
- The developed algorithm may be sensitive to the noise.
- The algorithm may be operating speed dependent
- The data sampling resolution seems to have a significant effect on the algorithms classification performance. From Table 1.8, it can be seen that there has not been a systematic choice of the sampling rate of the sensor signals.

1.16 Outcomes of the Literature Review

The literature on the CP fault diagnosis using various techniques has been reviewed. The CBM based maintenance techniques rely on collecting the signals from the equipment to be monitored and responding to the deviations in the signal values from the baseline conditions, thereby suggesting maintenance based on the fault present in the system. The importance of CBM based techniques to identify various faults in CPs is presented in detail. Following are the outcomes of the literature review.

- Rotating machines are prone to failures due to various reasons, including, the unbalance, misalignment, mechanical component damages, electrical faults, hydraulic faults, etc.
- To increase the longevity/ useful life of the system (rotating equipment), it is important to identify the system faults at their inception. Therefore, the condition based monitoring techniques are gaining popularity in industrial applications. They reduce the maintenance costs by using the information collected from the equipment to monitor their state.
- Centrifugal pumps are very important fluid energizers used in many industrial and household applications. They maintain the processes in the plant and any breakdown of them directly affects the process flow. Therefore, researchers and industry practitioners are trying to understand their various failure mechanisms and are working out strategies to curb the progression of faults in them. This helps improve their durability and avoids any unscheduled breakdown or maintenance.
- CP generally fails due to fluid flow instabilities or mechanical component faults or a mix of both. Fluid-related anomalies in a CP are due to the fluid-recirculation, cavitation and dry-run condition. Mechanical faults in a CP may be attributed to the seal defects,

bearing faults, rotor misalignment, rotor unbalance, CP internal surface damages and impeller defects.

- Operating parameters that are generally acquired for the CP condition monitoring are the vibration, pressure, acoustic emissions, noise spectra, and motor line current signatures. Vibration signals carry intricate fault information and are popularly used in the CP condition monitoring.
- Most of the research work on the CP fault diagnosis considered a maximum of three or four faults on the CP. Also, the faults considered were either purely mechanical or purely flow related. However, it is important to consider the combined existence of mechanical and hydraulic faults as they can cause a catastrophic CP failure.
- Identification of severity of the fault is also essential as it can help initiate a more specific corrective action. A single fault detection scheme for the identification of independently existing and co-existing faults in a CP system along with the estimation of the fault severity is still not very common in the literature.
- CBM techniques have moved from the traditional visual inspection methods to much more versatile machine learning techniques. Therefore, there is much scope in this area for the research. The advantages of using a machine learning algorithms are (1) garbled big-data of system operating parameters available with the industries could be effectively used to train the algorithm, thereby, predict the fault, (2) a reliable classifier could accurately identify faults, (3) realize maintenance process automation strategies, and (4) have quick on-line monitoring systems. However, there is a need to develop a reliable learning algorithm that gives a quick decision with fewer databases, and which is robust to the CP operating conditions and surrounding noise.

- Some of the expansively used machine learning algorithms for fault classifications are the neural network, fuzzy logic, rough sets and partially linearized neural networks, and support vector machines (SVM). The SVM algorithm is a relatively new classifier, and there has been recently extensive research on this. Some studies compared the efficiencies of the SVM algorithm with neural networks. It was found that the former has been efficient and less time consuming compared to the latter. The diagnosis of multi independent and co-existing faults in CPs is still uncommon using SVMs.
- In all the works mentioned in the literature, there is no guideline given on the data acquisition sampling rate. However, the choice of the sampling rate (and the number of data points) is critical and plays a deciding role on in identifying the fault in the CP system.
- The machine learning techniques basically comprise of the following steps, (1) acquiring the data in a prescribed domain, (2) pre-processing and acquiring suitable features from the data, and (3) using this feature data to train the machine learning algorithm. The effectiveness of a machine learning algorithm in classifying various CP faults depends on the sensitivity of features used and the domain from which the features are extracted. The choices of domains are the time, frequency and time-frequency (wavelets). The choice of the domain also depends upon the type of faults present in the system.
- Wavelets are very lucrative signal processing techniques and a very rarely used in conjunction with machine learning methods to develop intelligent diagnosis tools.
- From the literature, it can be understood that there are several choices for the mother wavelet. Despite many studies on the wavelets, the choice of the mother wavelet is still an open topic for research.

- Also, the methodologies designed for the CP fault identification and diagnosis were only applied to a unique CP operating condition (speed). However, in many applications, the CP speed may not be unique. Also, in complicated non-linear systems, it is challenging to build precise methodologies describing system operations at many different speeds.
- The literature lacks study of methods to deal with the unavailability of data at a particular operating speed of the CP.
- Also, the effect of corruption of signal data on the efficiency of the machine learning algorithm is rarely touched upon. However, this is an important area as it is close to impossible to avoid data corruption, especially in industrial applications.

From the outcomes mentioned above of literature, the aims and objectives of the thesis work are derived and presented in the subsequent section.

1.17 Aims and Objectives of the Current Research

After an extensive literature survey, it can be understood that CPs are critical and popular hydraulic machines used in versatile applications. From the literature, it is clear that SVM is a straightforward non-probabilistic technique for the fault diagnosis, and has been very rarely used for the CP fault diagnosis, however, it has great potential. Taking into account the advantages of SVM and its benefits over the other machine learning techniques to identify faults it may be a useful technique to identify CP faults. From the literature and the aforementioned outcomes of literature, the following are the derived objectives of the current research:

1. To formulate and explore machine learning algorithm based on the SVM to detect and diagnose mechanical and hydraulic faults on the CPs.

2. To develop a fault diagnostic approach to identify severity level of various flow related faults considered on the CP.
3. To experimentally study the interdependency of the mechanical and hydraulic CP faults.
4. To study the effectiveness of simultaneously using vibration signals and motor current signals to identify the CP faults. Also, to estimate if they have any complementary effect.
5. To choose the adjusted sampling rate and data points (or frequency resolution) for the fault diagnosis of the CPs that can improve the fault detection time and reduce the cost of implementation.
6. To extract and select the best features from the acquired data from experimentation for effective fault diagnosis of the CPs. Signals will be considered in time-domain, frequency-domain and time-frequency (wavelets) domain. In case of wavelet analysis, two techniques namely continuous wavelet transform and wavelet packet transform will be employed. Also, in these techniques, different mother wavelets would be investigated, and the best would be selected.
7. To investigate the robustness of developed methodology to handle the addition of noise to data in various domains. Also, the estimation of the amount of noise addition that can be adequately handled.
8. To check if the developed algorithms have any CP operating speed dependencies. Hence, the algorithm is checked at a wide range of CP operating speeds.
9. To perform the fault diagnosis at the same speed (same training and testing conditions) and intermediate test speed.

10. To compare the results of domain-based methodologies developed, thereby concluding which method suits best for what conditions.

1.18 Organization of Thesis

The thesis is organized into ten chapters. The introduction and literature review are presented in Chapter 1 (the current chapter). The experimental set-up and the laboratory test rig along with the equipment used, and the phases of experimentation are described in Chapter 2. The introduction to support vector machine based fault diagnosis methodology and the basic flowchart of the diagnostic methods are described in Chapter 3. In Chapter 4, the preliminary time and frequency domain based analysis and their results are presented. The learning from this Chapter is taken forward to develop the rest of the chapters. Chapter 5 presents the multi-fault diagnosis of CP faults using the corrected sampling rate (low sampling rate) in the time domain. In Chapter 6, the fault diagnosis is performed with the high-frequency resolution data in the frequency domain. In Chapter 7, the multi-fault diagnosis based methodologies using CWT data are presented. Chapter 8 describes the multi-fault diagnosis methodologies using WPT data. Chapter 9 gives the comparative analysis of the methodologies developed in various domains (time, frequency and wavelet-based). The major conclusions and the scope for future work are presented in Chapter 10.

CHAPTER 2: Laboratory Rig and Experimental Procedure for Centrifugal Pump Faults

2.1 Introduction

In Chapter 1, the main objectives of the present work were derived. From the objectives it is clear that a proper experimental rig is required to acquire various signals from the CP system. In the present chapter, the experimental set-up for the CP fault diagnosis is presented and the instrumentation used for measurement and acquisition of the signals is discussed. The procedure for the data collection from CPs with different faults is also presented in the current chapter. The experimentation involved in the current study has been carried out in the ‘Vibrations and acoustics laboratory’ at IIT Guwahati. The research team comprising of Prof. Rajiv Tiwari, Ms. Janani Shruti Rapur, Dr. DJ Bordoloi, and Mr. Aashish Panda have designed the experiments. The experiments were conducted in May, 2015 (Phase – I), April, 2016 (Phase – II), August, 2016 (Phase – III). The equipment such as MFS, motor, CP, sensors, tank were already present in the lab. Also, the preliminary investigation including Phase-I experiments were majorly carried out by Mr. Ashish Panda. The rest of the experiments design were carried out by the author.

2.2 Laboratory Rig and Experimentation

For the purpose of experimentation, a Machine Fault SimulatorTM (MFSTM) setup was used. This set-up is capable of testing of a range of machine conditions including pump faults, gearbox faults, shaft misalignments, unbalance, rolling bearing damages, induction motor faults, etc. The actual set-up and a schematic diagram of the MFSTM are shown in Figure 2.1 and Figure 2.2, respectively. The following are the list of components used and their original

equipment manufacturer (OEM) are mentioned in the brackets. The details of these components are described in the subsequent sections. The following are the list of components used in the test rig,

1. Machine fault simulator (SpectraQuest, USA)
2. Tri-axial accelerometers (Meta Mess-und Frequenztechnik, Germany)
3. AC line current probes (Keysight Technologies, India)
4. Constant DC power source unit (MES-Technik Pvt. Ltd., India)
5. Data acquisition system, DAQ (National Instruments, USA)
6. CPs (OberdorferTM, USA)

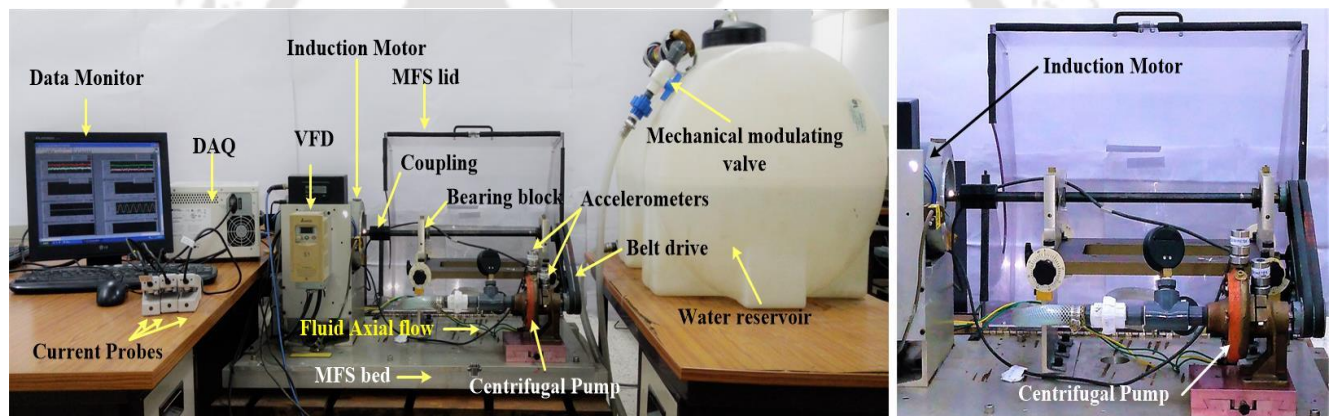


Figure 2.1: CP fault diagnosis experimental set-up

2.2.1 Machine Fault Simulator

MFS is an innovative tool used for studying signatures generated by various machinery faults. This system weighs close to 59 kilograms and is compact in dimensions. All the components in a MFS are machined to high precision so that it can operate in a wide range of speeds without any conflicting vibrations. Depending upon the requirement various individual or combined faults can be simulated on this set-up in a completely controlled environment. The technical details of MFS and its allied components are given in Table 2.1.

Table 2.1: Technical specification of various MFS and allied components

S.No	Component name	Technical specification
Electrical equipment:		
1	Induction motor	Three phase, 0.5 HP, pre-wired, self aligning mounting system for easy installation
2	VFD	Variable frequency drive with multi-featured front panel programmable controller, rpm range 0 to 10,000 rpm (short duration) variable speed
3	Voltage	Supply voltage 115/230 V AC, single phase, 50 Hz
4	Tachometer	Built-in tachometer with LCD display and one pulse per revolution analog output for DAQ purposes. Requires DC power supply (± 30 V/ 2 A)
5	DAQ	NI PXI-8108 Core 2 Duo 2.53 GHz controller with NI PXI-4472, 8 channels, 24 bits, 102.4 kS/s sampling rate for accelerometer module
Mechanical equipment:		
1	Centrifugal Pump	Oberdorfer™ 60P, pump body: bronze, inlet: 0.019 m, outlet: 0.127 m, flow: 5.45 cubic meter per hour, pressure: 0.16 MPa max, 16.15 m of head
2	Shaft	19.05 mm diameter, turned ground and polished steel
3	Bearings	Two rolling element bearings mounted in aluminum horizontal split housing
4	Rotor base	360 mm long, completely movable using jack bolts
5	Belt mechanism	Two double groove ‘V’ belt drives with one set screw mounting and one set bush/ key mounting
6	Foundation	12.7 mm die cast aluminium base, base stiffeners and 8 rubber isolators
Measuring equipment		
1	Accelerometers	Details in Table 2.2
2	AC line current probes	Frequency range: DC to 100 kHz (-3 dB with current de-rating) Current range: 100 mV/A: 100mA to 10A, 10 mV/A: 10A to 100A

The construction of the MFS comprises of an induction motor, a split bracket bearing housing, a sliding shaft, a rotor with split collar ends, a flexible coupling, ‘V’ belt pulleys, belt tension

devices, a variable frequency drive with multi-featured front panel programmable controller, a photovoltaic sensor and a sliding platform with a central jack screw. All these components can be mounted and dismounted with ease, based on the requirement. The AC variable frequency drive is used to adjust the speed of the motor. The photovoltaic sensor is mounted near the coupling to measure the actual speed of the shaft via a reflecting tape on the shaft.

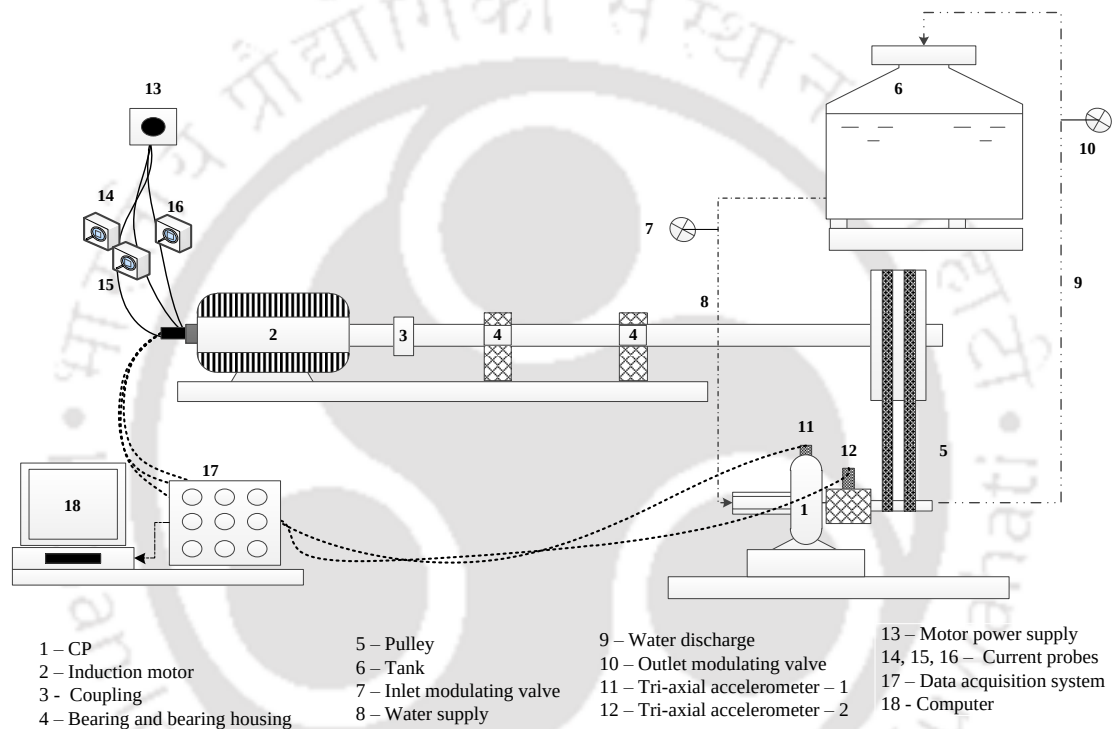


Figure 2.2: A schematic diagram of the experimental set-up

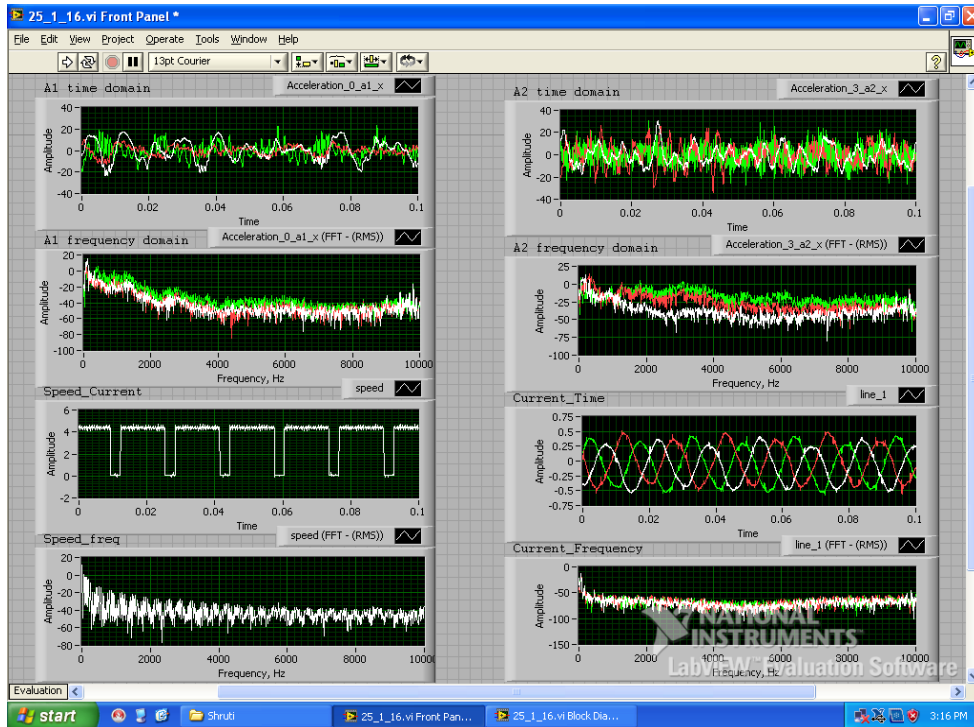
Various sensors can be used for acquiring the signals from the operating equipment. Whenever a fault appears in a CP system, there are changes in its operating conditions, which reflects as a change in load on the induction motor that is driving it. Therefore, it is expected that the line current of the motor may serve as a good metric to understand the CP fault state. In continuation, vibration signals are versatile and are sensitive to the dynamics of the rotating machinery. Hence, in the present study, the vibration and motor line current signals acquired

from the CP and the motor, respectively are used to monitor the CP. The accelerometers used, measure vibration dynamics in three orthogonal directions (axial, radial and vertical transverse). Two tri-axial accelerometers are mounted on the CP housing and the CP bearing housing locations. Three AC current probes are clamped around the power leads of the motor, which are very easily reachable. These sensors are connected to a central DAQ via proper cable connections. LabVIEW software supplied by the National Instruments, USA is used to configure the data collection. Figure 2.3(a) shows the LabVIEW architecture and Figure 2.3(b) shows the sampled data. Details about some of the components used in the current experimentation are given in subsequent section. This is followed by the details of experimentation in Section 2.3.

2.2.1.1 Variable frequency drive

A motor speed controller or a VFD is used to alter the speed of the motor by varying the frequency and voltage supplied to it. This is a power electronics based device, which converts a fixed frequency and fixed voltage sine power wave to a variable frequency and variable voltage output, which is used to control the speed of the motor. It comprises of a rectifier bridge converter, a direct current link and an inverter which allows it to increase or decrease the speed as per the user's application requirement. In the present work a variable speed AC drive (0.746 kW, single phase, 9.2 A/3-phase, 5.1 A, 200-240 V and 50/60 Hz, Output: 3-phase, 4.2 A, and 0-240 V) is used. Using a knob, the speed can be manually adjusted between 0-400 Hz.

(a)



(b)

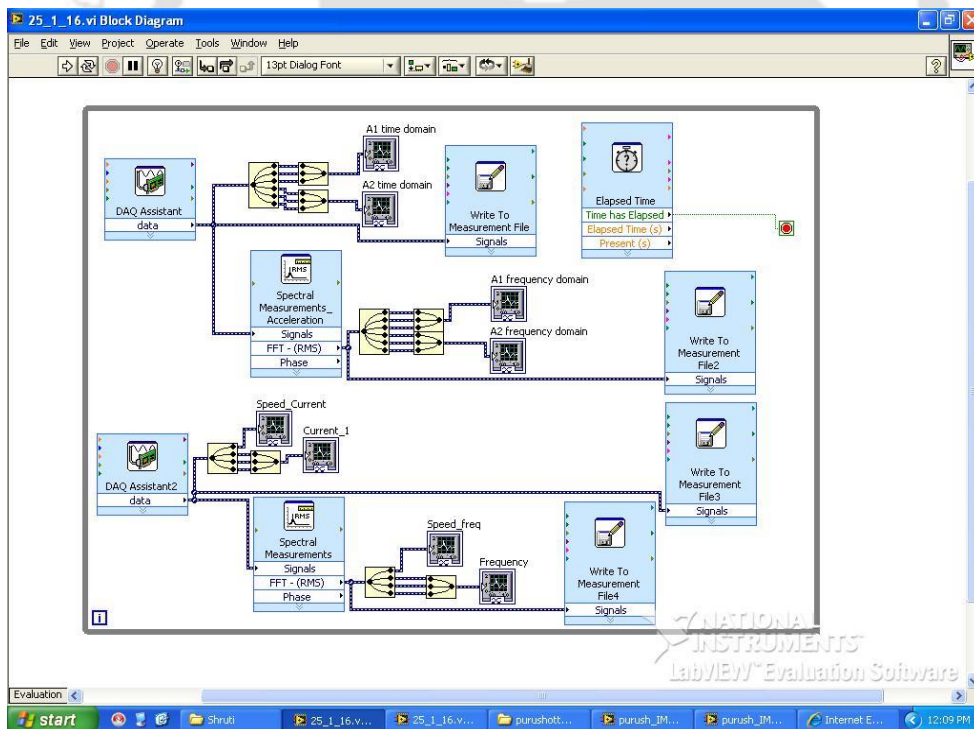


Figure 2.3: (a) LabVIEW environment, (b) samples raw data plot

2.2.1.2 Photovoltaic sensor and a constant DC power source

A photovoltaic sensor, which acts as a tachometer, is mounted near the MFS coupling. It measures the rotational speed of the shaft. This sensor needs a constant DC power source ± 30 V/2A to function.

This equipment converts the AC power supplied by the mains to constant DC voltage. Thus, it protects the probe from overload or short circuit. It has an adjustable current limiter and a digital display for the voltage and current. Thus, the high resolution voltage and current outputs can be acquired. The purpose of the tachometer is to measure accurately the rpm of the shaft, which is recorded as one pulse per revolution in the DAQ. The probe is assembled near the shaft end of the motor. It strikes a laser beam onto the shaft, which is reflected by a small reflecting strip attached to it.

2.2.2 Measurement Sensors

The CP condition is set to be monitored with the help of vibration measurements and current variation measurements.

2.2.2.1 Tri-Axial Accelerometers

A real-time data in time domain was measured using two tri-axial accelerometers mounted on the CP casing and the CP bearing housing locations, as shown in Figure 2.4. Accelerometer-1 was mounted on the pump casing. Accelerometer-2 was mounted on the bearing housing location. The pump casing location is chosen so as to capture the dynamics of bubble implosions. In addition, it is generally expected that vibration characteristics are best captured at the bearing location, where force transfer takes place, hence the bearing

housing location is chosen for mounting the second accelerometer. These accelerometers have a strong magnetic base and using a metal strip and strong adhesive they were mounted in the aforementioned locations. These accelerometers work based on the principle of piezoelectricity. An electric charge signal proportional to the amplitude of vibration is generated. The advantages of these types of accelerometers are: they have wide dynamic range, they are highly sensitive, they have no moving parts, and they are self-generating hence require no external power source. These accelerometers measure the vibration data in three orthogonal directions. Three directions are important because as mentioned in Chapter 1, various faults have specific characteristic frequencies in different directions. The output from the accelerometers and current probes was then connected to the data-acquisition hardware using BNC cables. Various properties of the accelerometers used are given in Table 2.2.

Table 2.2: Details of the accelerometers

Accelerometer location	CP casing (Accelerometer 1)	Bearing housing (Accelerometer 2)
Type	KS 813 B	KS 813 B
Serial number	07013	07012
Voltage sensitivity	100.3 mV/g-x axis, 100.7 mV/g-y axis, 101.4 mV/g-z axis	101 mV/g-x axis, 101.1 mV/g-y axis, 101.4 mV/g-z axis

2.2.2.2 AC current probes

The motor line-current was measured using AC line current probes is shown in Figure 2.. Three AC current sensors were clamped around three motor cables to acquire stator current signals in all the three phases. The ‘The Keysight 1146B’ AC/DC current probe provides an accurate display and measurement of currents from 100mA – 10 A rms, DC to 100 kHz,

without breaking the circuit. The sensors make use of Hall effect sensor technology to measure the AC and DC signals. The load variation on the CP due to various faults also varies the amount of load required by the motor. Therefore, it was expected that current signals carry useful fault information. Also, the current variation in all the three phases was measured so that adequate fault information was obtained. Figure 2. shows the probes measuring the line current. These sensors were also connected to the DAQ using BNC cables.

2.2.2.3 Data acquisition system

A data acquisition system (DAQ) is an electronic data collector system comprising sensors, DAQ measurement hardware, and a computer or a monitor with programmable software. This DAQ is an analog to digital converter which converts the conditioned sensor signals to digital values. The analog signal is a varying voltage proportional to the measured current and vibration value. The DAQ hardware includes a controller (specification: NI PXI-8108 Core 2 Duo 2.53 GHz) and a dynamic signal analyser (specification: NI PXI-4472, 8 channel, 24 bits, sampling rate 102.4kS/s). This controller unit runs on Windows XP and it has inputs for devices like mouse, keyboard, LAN connector jack. Outputs of the device include, monitor, hard drive and the USB ports. Amongst the many modules available, the vibration and the current signals are acquired using the time capture module (signals in time domain) and the frequency capture module (signals in frequency domain). The output from different sensors including the accelerometers, the current probes and the photoelectric probe are connected to different channels of PXI-4472. The data configuration software LabVIEW is installed on to the PXI-8108 system, thus enabling collection of real-time time and frequency domain data. Thus collected data is stored in “.lvm” format in the working directory with the user defined name.

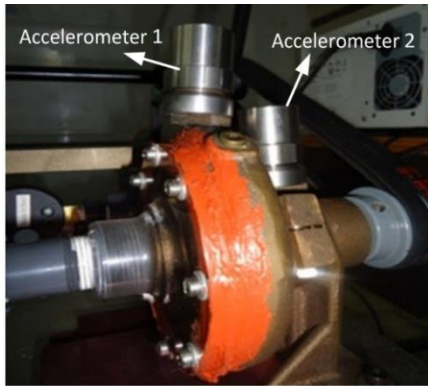


Figure 2.4: Accelerometers mounting position



Figure 2.5: Current probes measuring motor line current

2.2.3 Faults Considered on the CP

As discussed in Chapter 1, CPs may fail due to mechanical component failures or due to fluid flow instabilities or the combination of both these factors. Therefore, to take these factors into consideration both mechanical and hydraulic CP faults are considered.

2.2.3.1 Healthy CP Condition

A healthy CP condition basically means that there are no faults that have been artificially given to the CP system. All the mechanical components including the impeller and the cover plate are theoretically in perfect condition or the required benchmark condition. However, owing to the manufacturing imperfections or assembly errors, there might be slight unbalances or disturbances in the operation of this CP. Yet, this is considered to be a ‘healthy’ condition as this is the state in which the industries would consider the CP working as acceptable.

2.2.3.2 Hydraulic CP faults

The details of various hydraulic CP faults, their causes and effects have been discussed in details in Chapter 1. From the descriptions given in Chapter 1, it is evident that there might

be many reasons why hydraulic CP faults may arise. However, some of those causes are considered in the present work.

I. Suction blockage

The low pressure side or the piping carrying the fluid to the CP from the sump is called as the 'suction side' of the CP. The suction pressure is generally low so as to draw fluid to the CP. Whenever the suction side is clogged (or blocked or flow is restricted), the pressure further drops at the eye of the impeller to overcome the resistance. In addition, the fluid entry angle at the inlet changes, resulting in the flow recirculation. This type of flow recirculation results in low pressure pulsations and thus leads to the bubble formation, formation of it inside a CP is dangerous as it results in a cavitation like damage and head loss resulting from the energy expended in compressing the air bubbles.

To simulate the suction blockage (SB) condition, a mechanical modulating valve which can be seen in Figure 2.6(a) is placed on the inlet hose pipe connecting the CP. As the restriction to the flow increases the associated recirculation and bubble formation also increases. This valve is calibrated so as to give six different levels of flow restrictions between 0% blockage and 100% blockage. The line diagram of the calibration is shown in Figure 2.. The different flow restrictions are depicted by SB k . where, 'k' stands for $k/6^{\text{th}}$ suction blockage restriction. Though six different blockages may be administered on the CP using the modulating valve, only SB1, SB2, SB3 and SB4 are considered to be the suction blockage conditions.

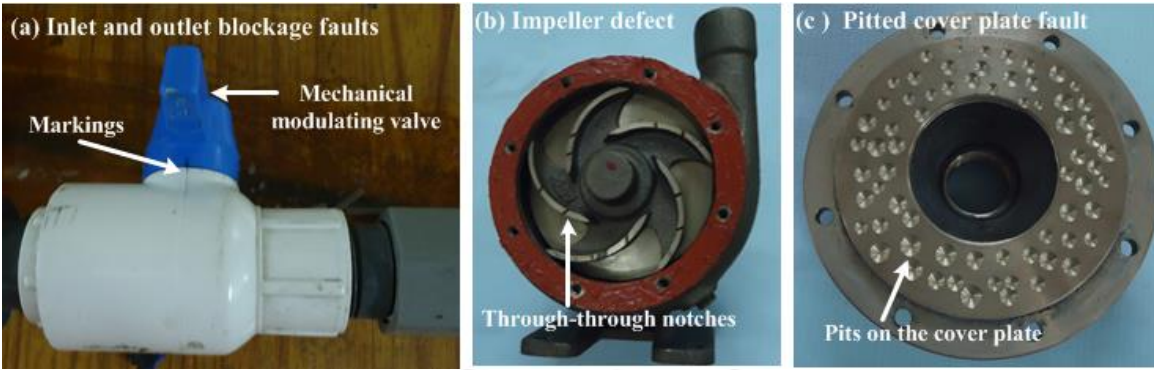


Figure 2.6: (a) Mechanical modulating valve with markings; (b) cracks on the impeller; (c) pitted cover plate

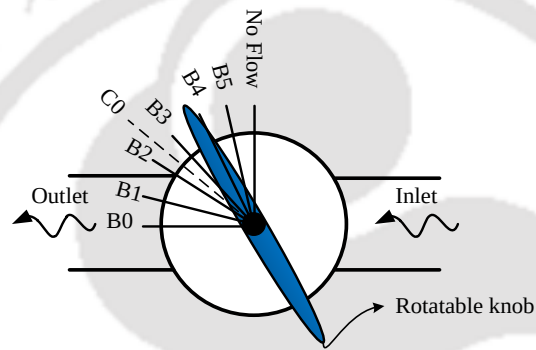


Figure 2.7: Line diagram showing different blockage levels

II. Discharge blockage

The second type of fluid-flow abnormality is induced by clogging the ‘discharge side’ of the CP. Discharge side basically refers to the high pressure side or outlet of the CP. Whenever the CP outlet is restricted, the discharge flow rate drops and thus the load on the impeller vanes increases. This again results in a suction recirculation and bubble formation.

To obstruct the flow on the discharge side, a mechanical modulating valve placed on the discharge pipe is used. A calibration similar to that of the suction blockage modulating valve as shown in Figure 2.6(a) and Figure 2. is used. The discharge blockage fault is given by the

abbreviation DB. Here, DB k signifies $k/6^{\text{th}}$ discharge blockage restriction. DB1 to DB5 levels of discharge blockages are considered in this work.

III. Dry run faults

The absence of priming fluid in the CP results in a ‘dry run’ condition. To simulate this condition on the CP, the flow is restricted on the suction side of the CP. It is observed that when SB5 restriction has been given on the suction side, the CP falls short of fluid. Hence, SB5 condition has been considered as a dry run condition.

2.2.3.3 Mechanical CP faults

Mechanical faults refer to the damages on the stationary or rotating mechanical components of the CP. The different possible mechanical faults in the CP are described in details in Chapter 1. In this work two mechanical CP faults have been considered. They are: impeller defects and pitted cover plate faults.

I. Impeller defects

The impeller is the heart of the CP. The faults in the impeller could be due to manufacturing defects, handling corrosive fluids, erosion of the impeller material due to abrasives present in the fluid, and cavitation damages. Whenever there is damage on the impeller, there is a rotating unbalance that is created and also the head developed by the CP gets affected. Hence, impeller damage is considered as one of the mechanical faults in this work. The impeller faults (IF) have been artificially created by cutting through-through notches on the CP vanes. Two notches have been cut on each vane. The location and orientation of these notches have been kept arbitrary. The IF can be seen in Figure 2.6 (b).

II. Pitted cover faults

The mechanical faults in the CP can also be stationary faults. To account to this, a pitted cover plate fault (PC) is considered. Any deviation in the geometric configuration of the CP from its standard design may result in fluid flow abnormalities and thus affecting the performance of the CP. The PC fault is shown in Figure 2.6 (c). This is also an artificially created fault. On the cover plate, pits of arbitrary depth have been made at random locations. Approximately, 70 pits which are protruding into the CP have been made on the flat surface of cover plate having an outer diameter of 92 mm and inner diameter of 49.98 mm.

2.2.3.4 Combination of CP faults

In this work, apart from considering the independent existence of the hydraulic faults or the mechanical CP faults, the co-existence of them has also been considered. Three configurations of CPs are used for this purpose whose details are shown in Table 2.3. On each of these CP configurations, the suction and discharge blockages are given externally using a mechanical modulating valve at the inlet and outlet of the CP, respectively. The different fault combinations considered in this work are given in Table 2.4.

2.2.3.5 CP faults nomenclature and description

A total of 33 faults are considered on the CP. The nomenclature, description and characteristics of various faults the CP are given in Table 2.5. The faults considered are:

- (i) healthy pump (HP),
- (ii) suction blockage faults (SB1, SB2, SB3, SB4),
- (iii) discharge blockage faults (DB1, DB2, DB3, DB4, DB5),
- (iv) impeller defects (IF),
- (v) impeller defects in addition to suction blockages (IFSB1, IFSB2, IFSB3, IFSB4),

- (vi) impeller defects in addition to discharge blockages (IFDB1, IFDB2, IFDB3, IFDB4, IFDB5),
- (vii) pitted cover plate faults (PC),
- (viii) pitted cover plate faults in addition to suction blockages (PCSB1, PCSB2, PCSB3, PCSB4),
- (ix) pitted cover plate faults in addition to discharge blockages (PCDB1, PCDB2, PCDB3, PCDB4, PCDB5) and
- (x) dry run faults for all the CP configurations (SB5, IFSB5, PCSB5).

Table 2.3: Three CP configurations used in the current research

Configuration number	CP Type	Good impeller	Good cover plate
I	Healthy CP	Present	Present
II	CP with impeller faults	Absent	Present
III	CP with pitted cover plate faults	Present	Absent

Table 2.4: Fault combinations considered in the proposed work

Fault combination	SB	DB	IF	PC
Healthy CP	Absent	Absent	Absent	Absent
Only SB	Present	Absent	Absent	Absent
Only DB	Absent	Present	Absent	Absent
Only IF	Absent	Absent	Present	Absent
Only PC	Absent	Absent	Absent	Present
Both IF and SB	Present	Absent	Present	Absent
Both IF and DB	Absent	Present	Present	Absent
Both PC and SB	Present	Absent	Absent	Present
Both PC and DB	Absent	Present	Absent	Present
Dry run	Present	Absent	A	Absent
Dry run with IF	Present	Absent	Present	Absent
Dry run with PC	Present	Absent	Absent	Present

2.3 Experimental Procedure

In this section, the general description about the experimental procedure, nomenclature of various faults considered, data collection procedure, are described in detail.

2.3.1 CP Mounting and Assembly

Oberdorfer™ 60P CP is used for the purpose of experimentation and signal generation. It consists of a rugged bronze construction having a single rotating metal impeller, which is not in contact with other pump parts. Liquid enters at the centre and is thrown outward radially by centrifugal force. The pedestal CPs can be direct driven by electric motors at either 3450 rpm or 1725 rpm. It can also be pulley driven at any desired intermediate speed. The liquid flow is proportional to the CP speed, i.e. the higher the speed the greater the flow. Power required varies proportionally with the CP flow and head developed. Maximum power is required with a wide open discharge line.

With a throttled discharge, the flow decreases at a greater rate than the pressure increases thereby requiring less power at the higher pressures. This is known as an “unloading” feature of the CP. This CP is not self-priming. The assembly drawing with dimensions of the CP used is shown in Figure 2..

The chosen configuration of CP is mounted on the fixed base of the MFS located at its lower right corner and this is driven by the motor via a double belt pulley drive. The speed ratio obtained is 1:1.04. That is if the motor is driven at 30 Hz the CP is driven at 31.3 Hz. The VFD is employed to change the CP rotational speed. A photovoltaic sensor is mounted near the coupling location to measure the rotational speed of the shaft.

Table 2.5: Nomenclature, characteristics and description of various CP faults considered

No	Fault condition	Fault description	Remarks
1	Healthy CP (HP)	CP without defects	Healthy/ benchmark condition
2	Suction blockage (SB)	SB1	Low severity, close to HP
		SB2	Low medium severity
		SB3	Medium severity
		SB4	High severity
3	Discharge blockage (DB)	DB1	Low severity, close to HP
		DB2	Low medium severity
		DB3	Medium severity
		DB4	High severity
		DB5	Very high severity
4	Impeller faults (IF)	CP with defective impeller	Administered by cutting two through-through notches on each of the impeller vanes
5	Impeller faults with suction blockage (IFSB)	IFSB1	Low severe
		IFSB2	Low-medium severe
		IFSB3	Medium severe
		IFSB4	High severe
6	Impeller faults with discharge blockage (IFDB)	IFDB1	Low severe
		IFDB2	Low-medium severe
		IFDB3	Medium severe
		IFDB4	High severe
		IFDB5	Very high severe
7	Pitted cover faults (PC)	CP with defective cover plate	Administered by pitting the cover plate of the CP randomly
8	Pitted cover faults with suction blockage (PCSB)	PCSB1	Low severe
		PCSB2	Low-medium severe
		PCSB3	Medium severe
		PCSB4	High severe
9	Pitted cover faults with discharge blockage (PCDB)	PCDB1	Low severe
		PCDB2	Low-medium severe
		PCDB3	Medium severe
		PCDB4	High severe
		PCDB5	Very high severe
10	Dry run faults (SB5)	SB5	Dry run on a geometrically healthy CP
		IFSB5	Dry run on a CP with IF
		PCSB5	Dry run on a CP with PC

In this work, the motor is run at discreet speeds ranging from 30 Hz to 65 Hz in steps of 5 Hz. As the motor is found to get heated up beyond a speed of 67 Hz, the upper bound for the experimentation is fixed at 65 Hz. For the sake of simplicity, the motor speed is referred to as the CP speed in the thesis. The actual experimental setup and its schematic are shown in Figure 2.1 and Figure 2.2, respectively. The CP takes water from a tank, which is positioned in such

a way that there is sufficient head at the inlet in order to avoid any inherent cavitation. The water is circulated in a closed loop. The inlet hose pipe has an inside diameter of 19.05 mm and the discharge hose pipe has an inside diameter of 12.70 mm. Both these pipes are transparent so that any bubble formation inside the pipe can be clearly seen.

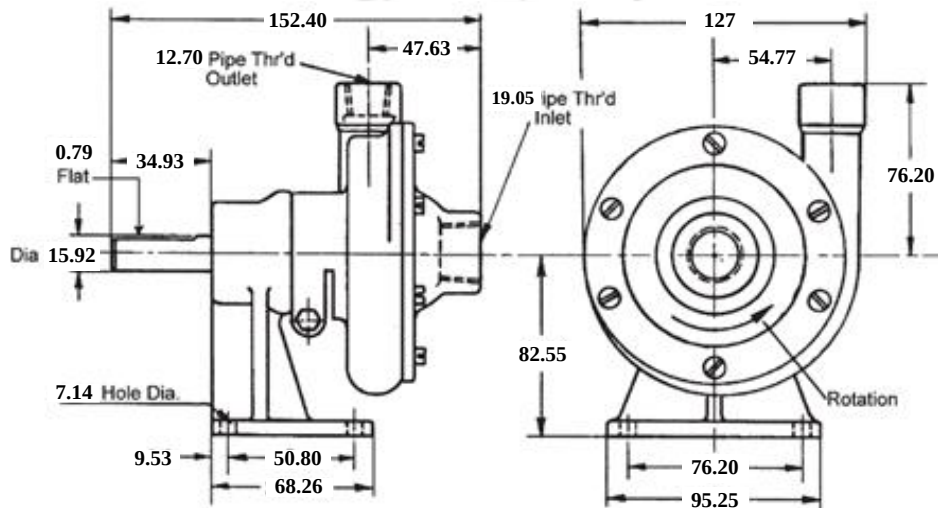


Figure 2.8: Assembly drawing with dimensions (all dimensions in mm)

For I and II configurations of CP faults (as referred in Table 2.3), a transparent polycarbonate cover plate is used on the CP. This aids visualization of the flow instabilities. The CP mounted on the MFS with a polycarbonate cover is shown in Figure 2.. But, in configuration III, when the faulty cover plate is fitted on the CP, the flow instabilities inside the CP may not be visualized.

2.3.2 Data Collection

LabVIEW software was used to build an automatic data collection system, as shown in Figure 2.3. The acceleration (two tri-axial accelerometers) and stator line currents (three line current probes) is collected in time domain and frequency domain. A separate measurement for each

fault at each operating speed is collected. Thus, collected data is then stored in distinct files for further processing.



Figure 2.9: Pump mounted on the MFS with a poly-carbonate transparent cover

To study the effect of sampling rate on the intricacy of the signal obtained and thereby the effect of it on the fault identification algorithm, two different sampling rates are tried out. The data is sampled at— 20,000 samples/s (2,000 samples/dataset) and 5,000 samples/s (5,000 samples/ dataset). The details of the amount of data collected for each fault and for all the faults are specified in Table 2.6. Table 2.6 shows the data resolution obtained in the frequency domain with each of these sampling rates. It can be observed that both the sampling rates chosen show different Nyquist frequencies and data resolutions. This is so chosen to check the reliance of algorithm's classification performance on the resolution of data acquisition.

The time domain and power spectral density of the raw data in linear scale, obtained from both high and low frequency resolution data sampling is plotted in Figure 2.(a), (b), (c) and (d), respectively. From the figure, it can be seen that the frequency activity for this system is typically within 1000 Hz. Over that, the high frequency resolution sample (Figure 2.(c)) tends to carry more details (intricate frequency amplitude information) of the characteristic fault

frequencies. Whereas, with the low frequency resolution data sampling (Figure 2.(d)) data of higher frequencies is obtained but with a loss of amplitudes at refined characteristic frequencies. Clear peaks can be seen at 3x, 3BPF and 8LCF frequencies.

Table 2.6: Details of data resolution and frequency content with different sampling rates

Samples to read per dataset (S)	2000	5000
Sampling rate	20 000 Hz	5,000 Hz
The time taking for acquiring one dataset (T)	2000/20000 = 0.1 s	5000/5000 = 1 s
Time taken to acquire single data point (ΔT)	$\frac{T}{S} = \frac{0.1}{2000} = \frac{1}{20000}$	$\Delta T = \frac{T}{S} = \frac{1}{5000}$
Resolution in frequency domain	$\frac{1}{T} = \frac{1}{0.1} = 10 \text{ Hz}$	$\frac{1}{T} = \frac{1}{1} = 1 \text{ Hz}$
Nyquist frequency (N, Hz)	$\frac{1}{2\Delta T} = 10000 \text{ Hz}$	$\frac{1}{2\Delta T} = 2500 \text{ Hz}$
Total time (s)	35	150
Total datasets	350	150
Data resolution in frequency	Low	High

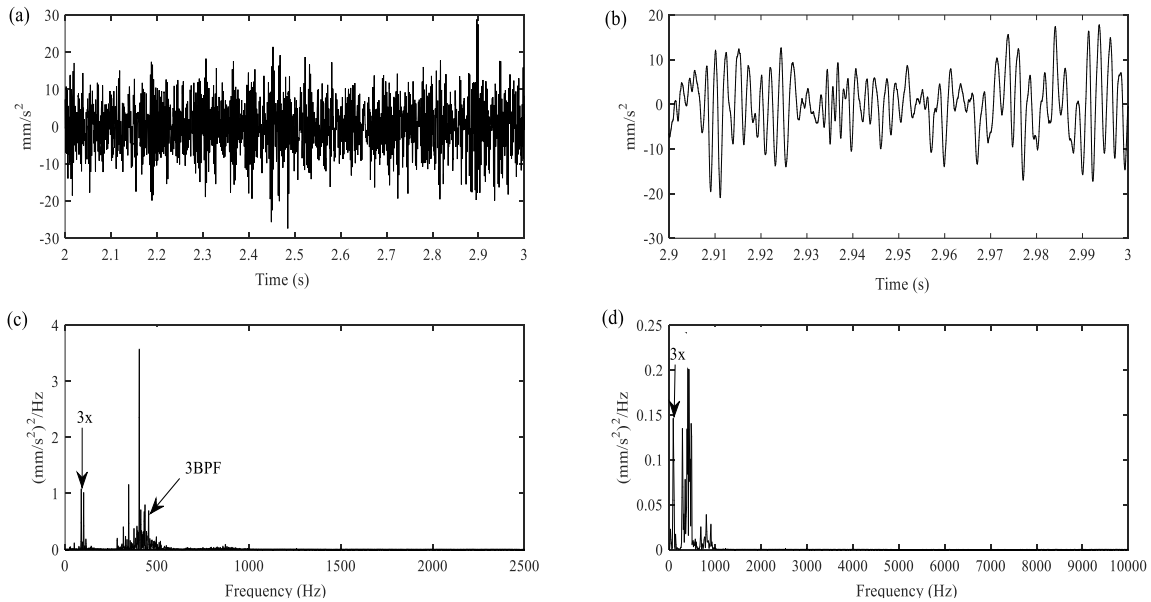


Figure 2.10: Raw time domain data and power spectral density data in linear scale from the accelerometer-1 for HP condition at 30 Hz with high frequency resolution data sampling in (a) time domain, (b) frequency domain, low frequency resolution data sampling in (c) time domain and (d) frequency domain.

Table 2.7: Total data collected using different sampling rates of data acquisition

Sampling rate (kHz)	No. of samples/dataset in time domain	No. of samples/dataset in frequency domain (a)	No. of data sets (b)	No. of measurements using two tri- axial accelerometers (c)	No. of measurements using three line current probes (d)	No. of faults (e)	Total data collected per fault $f=(a \times b \times (c+d))$	Total data collected for all faults $g=(f \times e)$
20	2000	1000	350	6	3	33	3150000	103950000
5	5000	2500	150	6	3	33	3375000	111375000

That means, it is safe to say that instead of capturing data at higher frequencies (where there is not much vibration activity), a better resolution may be chosen to capture intricate frequencies depending on the requirement and type of faults being studied. The fault signatures of various faults generated by the accelerometers and the line current probes in time domain and are given in Appendix B (high sampling rate) and Appendix C (low sampling rate), and in frequency domain are given in Appendix D (high sampling rate) and Appendix E (low sampling rate). From the raw data it can be observed that it is very difficult to visually inspect the data (especially in time-domain) and understand the absence or presence of a fault in the CP system. Hence statistical method (like machine learning) may be useful in fault identification.

2.4 Experimental Observations

The following observations were made during experimentation.

- HP configuration of the CP does not produce any visually observable flow anomalies. The vibration or the audible noise is also moderate even at high speeds of CP operation and therefore, there is no ambiguity in considering this state as the ideal CP operating state.
- The suction CP blockage up to SB2 level does not seem to change the flow pattern in CP. However, there is bubble formation observed between the SB2 and SB3 levels. The severity of bubble formation and the size of bubbles keep increasing with the increase in the suction blockage severity and also with the CP operating speed.
- The presence of impeller defects (even without any suction/ discharge blockage) causes bubble formation.

- Pits on the cover plate also seem to disturb the flow pattern. Tiny bubbles are observed with PC fault in the suction and discharge pipes at increased speeds of the CP operation.
- Dry run faults cause a lot of vibration and associated noise.
- The presence of discharge blockage reduces the pump discharge significantly and increases the vibrations.

2.4.1 Interpretation of the Experimental Observations

From the experimental observations, it is evident that even without forcing a flow abnormality on the CP (i.e., SB or DB), a CP with mechanical faults (i.e., IF or PC faults) generates flow instabilities by itself. This implies that the mechanical faults are capable of generating flow instabilities. In this section, a hypothesis for the experimental observations is presented.

As mentioned in the experimental observations, the CP with impeller cracks revealed bubble formation even in the absence of any blockage. This implies that the impeller cracks act as flow instability inducers. The CP as shown in Figure 2.4(a) runs in the backward curved vane configuration. Figure 2.4(a) shows the flow pattern inside a CP without impeller cracks and Figure 2.4(b) shows the flow pattern inside a CP with impeller cracks. The convex side of the impeller blade is on a higher pressure than the concave side. In Figure 2.4(b), the HPZ stands for high-pressure zone and the LPZ stands for low-pressure zone. This pressure difference provides for the fluid to flow through the cracks as shown in Figure 2.4(b). Hence, some fluid, which is supposed to flow along the convex vane face, leaks into the LPZ of the following vane as shown in Figure 2.4(b). This flow disrupts the flow pattern of the following impeller

blade and may cause a *pseudo re-circulation*. Re-circulation is always accompanied by low-pressure pulsation zones and hence this might be the reason for bubble-formation.

Any bubble formation inside the CP system results in a decreased CP performance and material erosion on internal CP surfaces. The collapse of bubble cloud creates a *shock wave* which travels in the fluid and whose magnitude attenuates as it gets closer to the solid surface. Near the solid surface solitary bubbles are present and when the shock wave reaches the bubbles, they oscillate and a micro-jet phenomenon may occur, as shown in Figure 2.5 . When this high velocity micro-jet impinges on the solid surface pitting takes place. Material erosion on the impeller blades or any internal surface of the CP causes or enhances the mechanical faults. Therefore, the mechanical and hydraulic faults form a vicious fault cycle, and hence need to be curbed at their formative stages itself. A probable vicious fault cycle is shown in Figure 2.6.

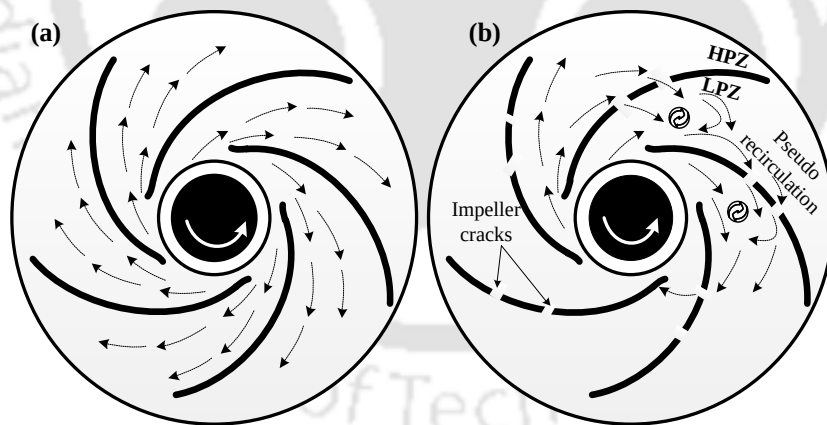


Figure 2.4: (a) Flow pattern over healthy impeller blades; (b) flow pattern over cracked impeller blades

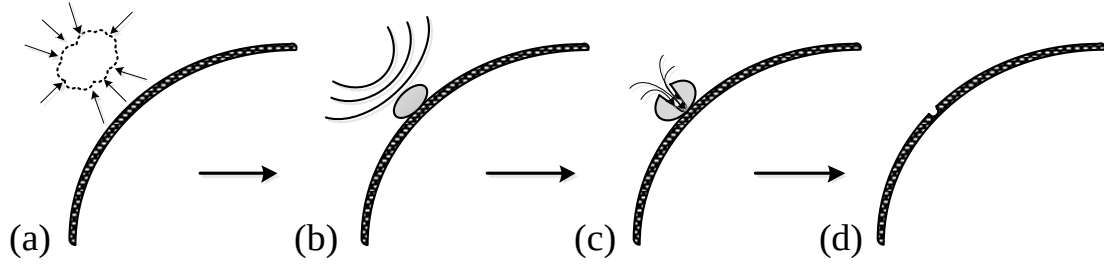


Figure 2.5: Stages of material erosion due to bubble formation (a) Collapse of bubble cloud, (b) pressure wave emission, (c) micro-jet formation, (d) pit formation and material erosion (Dular et al., 2006)

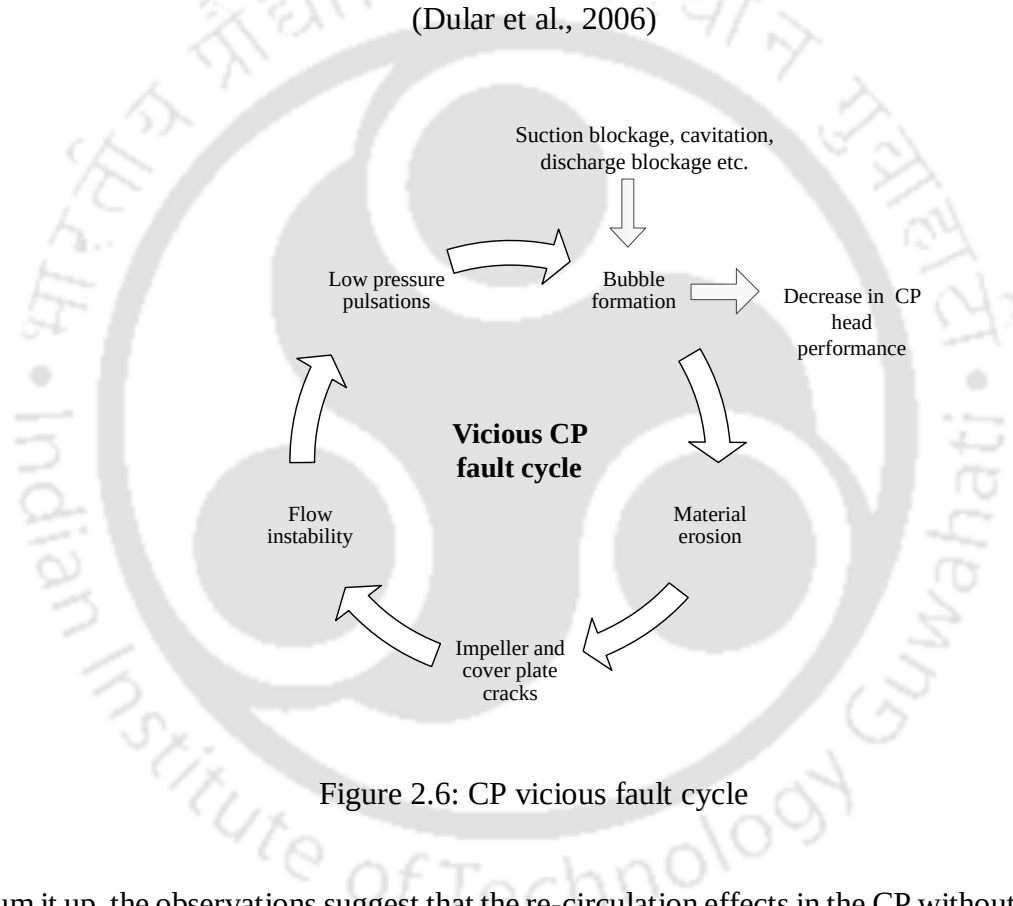


Figure 2.6: CP vicious fault cycle

To sum it up, the observations suggest that the re-circulation effects in the CP without impeller defects/ pitted cover plate faults is due to the suction flow restriction and discharge flow restriction governed by the blockage. While in a pump with impeller defects, the cracks govern a pseudo re-circulation. It is also expected that the pits on the cover plate cause a flow pattern disruption in the CP.

2.5 Experimentation Phases

Since the understanding of the faults and their behavior matured day by day during the course of this thesis work, the experimentation and analysis of the results also evolved likewise. Hence, the experimentation pertaining to this thesis work was carried out in three phases.

2.5.1 Experimentation - Phase One

The objective of the experiments in phase one was to identify the start of bubble formation and different severities of suction blockages on the CP. Hence, only HP, SB1, SB2, C0, SB3, SB4 blockages were considered. Here, C0 depicts the beginning of the bubble formation in the CP. As explained previously the bubble formation started between SB2 and SB3 suction restriction levels and the amount of flow restriction for the beginning of bubbles was dependent on the speed of operation of the CP.

Vibration data was acquired for different faults using 20 kHz sampling rate. From the data acquired using this study, fault diagnosis was done in time domain and the frequency domain and the results and observations of this are presented in Chapter 4. In phase I experiments, the CP is run at speeds from 40 Hz to 65 Hz in steps of 5 Hz.

2.5.2 Experimentation - Phase Two

After experimentations in phase one, a study on classification of mechanical faults was attempted. Therefore, impeller defects were considered on the CP. During experimentation it was observed that the mechanical and hydraulic faults are not always independent of each other. This triggered the thought of considering a combination of mechanical and fluid flow faults. Thus, in phase two, HP, SB1, SB2, SB3, SB4, IFSB1, IFSB2, IFSB3, IFSB4 faults

were considered. Here also vibration data at 20 kHz sampling rate was acquired and the results and observations of this are presented in Chapter 4.

2.5.3 Experimentation - Phase Three

The fault diagnosis results obtained from phase one and phase two experiments revealed that the data sampling rate may play an important role in how well a classifier may identify the fault data. This further manifested to the phase three experimentation. In this one of the objectives was to figure out the classification accuracy dependency on the data resolution. Therefore, in these experiments the data was sampled at 20 kHz as well as 5 kHz (as explained in Section 2.3.2). Also, from the literature it was observed that most researchers and scientists confined their analysis to data acquired from any one type of the signals. It was intriguing to find out if usage of any other type of signal could help improve the algorithm's performance. Hence, in this study both vibration and motor line current data were acquired. Apart from this the faults that have been considered have also been increased. Here the objective was to not only classify various faults but also to identify the severity of faults successfully. The faults described in have been considered in this work. The results of fault classification using time-domain data, using frequency domain data and using the time-frequency domain data has been presented in Chapters 5 to 8. In phase two and three the CP is run from 30 Hz to 65 Hz in steps of 5 Hz. The CP is run from 30 Hz to 65 Hz because 30 Hz is found to be the minimum speed at which the CP could overcome the resistance of the belt drive, and other inertia to suck fluid into it. Over that the objective of this research is to find out if the developed methodology is indifferent to the running CP speed. Hence it was understood that taking eight speed levels was enough to solve the problem. The amount of data handled (in the order of 50+ GB for each data sampling rate considered) and the time taken for the data acquisition

(more than 2 months) is also high for the current study. However, this is not a limitation of the developed methodology and the study could be extended to any number of discrete speeds, provided the necessary computation facility is available. The study was carried out only till 65 Hz because it was intended to run the system below its first natural frequency (107 Hz). Also, the authors wanted to maintain the CP's geometric and flow compatibility all through the experiments. If the CP were to be run at high speeds, it would have become very difficult to restrict the fault propagation.

Here it is also important to note that the data is sampled as follows,

Sample set 1: Sampling rate: 20 kHz, Sample length: 2000, No. of datasets: 350, Time taken to acquire the data for a fault condition at a particular speed: 35 seconds, Nyquist frequency: 10 kHz.

Sample set 2: Sampling rate: 5 kHz, Sample length: 5000, No. of datasets: 150, Time taken to acquire the data for a fault condition at a particular speed: 150 seconds, Nyquist frequency: 2.5 kHz.

For Phase – I and Phase – II experiments sample set 1 was used. The sampling rate was chosen based on literature survey. However, in previous works on fault diagnosis of rotary equipment, independent mechanical/ hydraulic fault classifications were considered. Also, closely varying fault severities were rarely considered. But, the current study considers complex fault combinations of mechanical and hydraulic faults. On analysis of the fault frequency domain data it is found that a higher resolution in frequency domain would better capture the fault characteristics in it. In addition, for sampling set 1, the Nyquist frequency was 10 kHz. However, for the highest speed of CP operation considered (65 Hz), frequency activity of till

about 1500 to 2000 Hz is expected. Hence, Phase – III experiments are designed using sampling set 2, where the frequency resolution is 1 Hz and Nyquist frequency is 2500 Hz.

2.6 Summary

In this chapter, the experimental data collection procedure has been presented for the CP with various critical faults, such as, blockage faults, impeller faults, pitted cover plate faults, dry run faults and their combinations. A laboratory test-rig was identified and the experiments were performed. The suction and discharge blockages on different configurations of the CP were given in stages to identify their severity. In this chapter, an interpretation of the bubble formation due to coexistence of the mechanical and hydraulic CP faults has been presented, which was not described in the literature before. In literature fault diagnosis of the CPs was attempted using any one type of the signal (acoustic, vibration, pressure, current, etc.) However, in this work both vibration and motor line current data are proposed to be used for the CP fault diagnosis. Vibrations were measured using tri-axial accelerometers mounted on the CP. The line-current was measured using current probes. The vibration and current data in time and frequency domains were collected for different fault conditions at different rotational speeds, and stored for further analysis. Like it was previously described in Chapter 1, the sampling rate needs to be judiciously chosen while acquiring signals from the machine. It is expected that sampling rate plays an important role in the classifier accuracy obtained. Hence, in the current research, the data was acquired at two different frequency resolutions so as to estimate the effect of data resolution on the classifiers performance, which is presented in the subsequent chapters. In the next chapter the fault classification methodology adapted in the present work is described.



CHAPTER 3: Centrifugal Pump Fault Classification Methodology based on SVM

3.1 Introduction

In this chapter, the methodology developed for the fault diagnosis of CP is presented. The vibration and current data acquired for different CP fault cases at different speeds/loads, as discussed in Chapter 2, will be used for the fault classification. Owing to the advantages offered by the SVM as studied in the literature (Chapter 1), it will be used as the machine learning technique in the present work. The SVM classification methodology will be discussed in detail in this chapter. As it will be explained in the later part of this chapter, to obtain a good classification performance an optimum combination of SVM penalty parameters and kernel parameters needs to be selected. For this purpose, a grid search technique (GST) along with a cross-validation (CV) technique are adapted in this work. For the selection of the best feature(s) an accuracy based model (wrapper model) is used. To check the versatility of the proposed algorithm, as a separate study, the experimentally generated fault data (as explained in Chapter 2) is corrupted with additive white Gaussian noise and this corrupted data is used for testing the algorithm, which has already been trained with the original uncorrected data.

3.2 Support Vector Machine Classifier

The SVM is a relatively new computational learning method based on the statistical learning theory (SLT) (Vapnik, 1998, 1999), and can serve as an expert system (ES). The SLT deals with supervised learning problems. The original SVM algorithm was developed by Vapnik

(Vapnik, 1995) and the current standard incarnation (soft margin approach) was proposed by Cortes and Vapnik (Cortes and Vapnik, 1995; Vapnik, 1998).

Many real world problems deal with binary classifications or decisions. For example, whether or not maintenance of a machine system is required, does a patient have a disease or not, etc. Basically these decisions involve a true/ false or positive/ negative or yes/ no answers. One of the best ways of solving such binary problems is by using SVM classifier which can be introduced as a *non-probabilistic binary linear classifier*. The SVM is an abstract *learning machine* that learns from a training data set and attempts to make correct estimates on novel data sets. The training data comprises of input vectors denoted by $\mathbf{x}_i$ and each input vector has a number of elements called the *features*. Each input vector is *paired* with a *label* denoted by y_i and there is p such pairs ($i = 1, 2 \dots p$). In a binary classification, it is obvious that there are two qualities (or classes) of data and these classes can be called as *positive* class or *negative* class, with data labels of $y_i = +1$ or $y_i = -1$, respectively.

The training data sets can be viewed as labeled data points in an input space, which can be depicted as shown in Figure 3.1. The plane that separates both the classes of data is called the *hyperplane*. There could be infinite orientations of the hyperplane that can separate the two classes of data as shown in Figure 3.1(a). But, the directed hyperplane found by a SVM is very intuitive: i.e., it is that hyperplane, which is maximally distant from the two classes of labeled points located on each side. The closest data points on both sides of the hyperplane have maximum effect on its position and orientation, and are therefore called *support vectors*. They are denoted by the dark markers as shown in Figure 3.1(b). The classifying hyperplane is given by $\mathbf{w} \cdot \mathbf{x} + b = 0$ (where symbol $\cdot$ is the dot product) in a space H . Symbol b is the

bias or offset of the hyperplane from the origin in input space, $\mathbf{x}$ are points located within the hyperplane, normal to the hyperplane and the weights $\mathbf{w}$ determine its orientation.

In Figure 3.1, two data clusters can be very nicely classified by a linear hyperplane in a 2-D space. Unfortunately, not always are the clusters linearly separable and could be highly meshed with overlying data points as shown in Figure 3.2. This situation is an impetus for the introduction of *kernel tricks* later in this chapter.

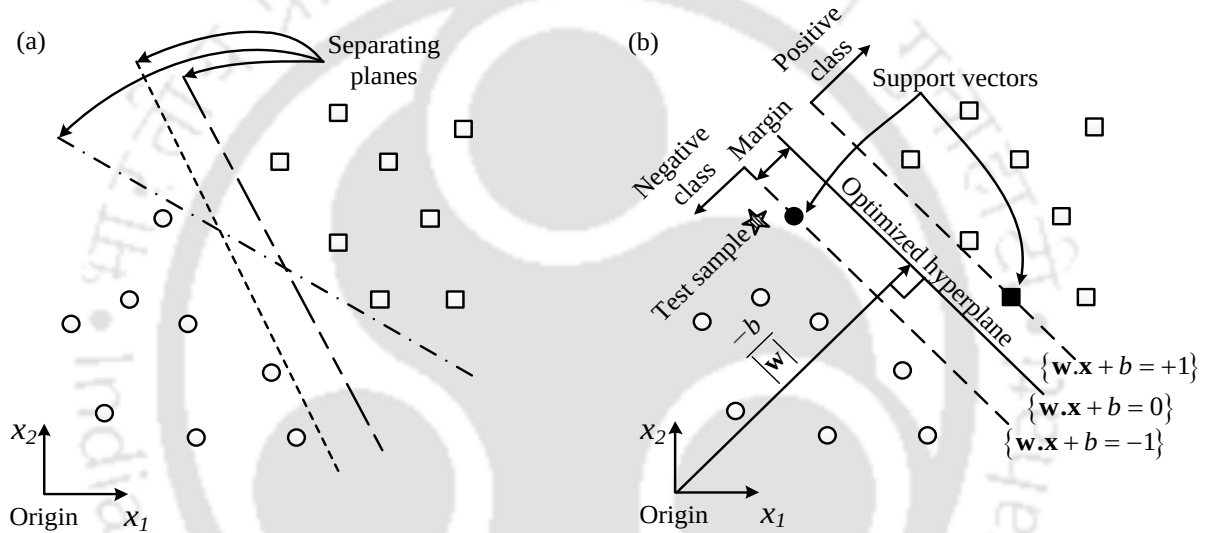


Figure 3.1 (a) Various orientations of the data separating planes; (b) the optimal orientation of the SVM hyperplane

The reason for considering SVM classifier comes from the theoretical upper bound on the *generalization error*, that is, the theoretical prediction error when applying the classifier to novel or unseen data points (training samples). This generalization error has two important attributes,

- The bound is minimized by maximizing the *margin*. The margin is defined as the Euclidean gap between the sorting hyperplane and the support vector of every class.
- The bound is independent of the dimensionality of the input space.

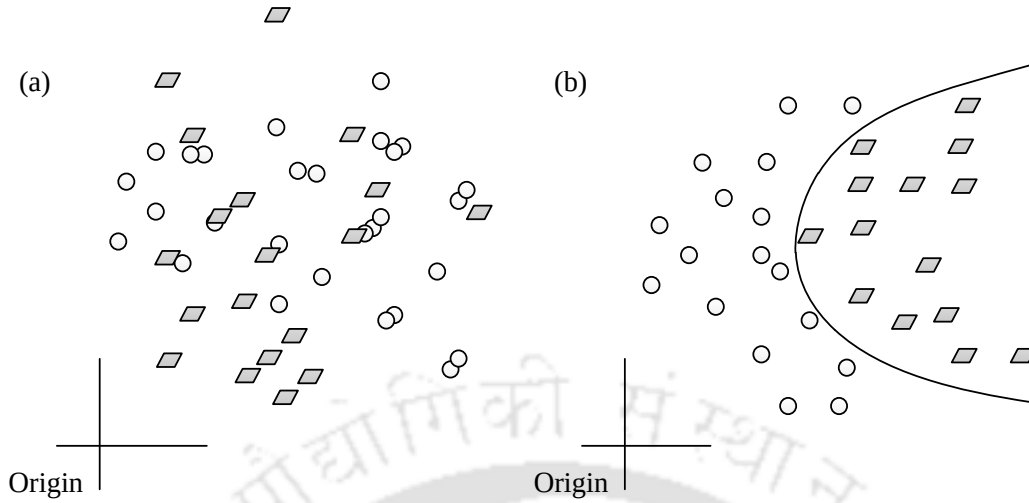


Figure 3.2 (a) Intermeshed non linearly separable data (b) using a Gaussian kernel the data gets separable using a non-linear boundary in the feature space

Consider a binary classification case to classify p sets of training data $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_p, y_p)$ into two classes, where, $\mathbf{x}_i \in R^N$, $i = 1, 2 \dots p$, and $y_i \in \{-1, +1\}$ specifies its class label. Now, let the decision function be,

$$f(\mathbf{x}) = \text{sign}(\mathbf{w} \cdot \mathbf{x} + b) \quad (3.1)$$

From the decision function given, it is clear that the data is correctly classified if it satisfies, $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) > 0 \quad \forall_i$ since $(\mathbf{w} \cdot \mathbf{x}_i + b)$ is positive for $y_i = +1$ and $(\mathbf{w} \cdot \mathbf{x}_i + b)$ is negative for $y_i = -1$. That is, the decision function is invariant of the positive rescaling of $(\mathbf{w} \cdot \mathbf{x}_i + b)$. Hence, an unspoken scale for $(\mathbf{w}, b)$ is defined by setting $\mathbf{w} \cdot \mathbf{x} + b = +1$ for closest data points on one side and $\mathbf{w} \cdot \mathbf{x} + b = -1$ for closest data points on the other side. The hyperplanes passing through $\mathbf{w} \cdot \mathbf{x} + b = +1$ and $\mathbf{w} \cdot \mathbf{x} + b = -1$ are called *canonical hyperplanes* and the region between these hyperplanes is called *margin band*.

Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be two points on the canonical hyperplanes on both sides. That is, if $\mathbf{w} \cdot \mathbf{x}_1 + b = +1$ and $\mathbf{w} \cdot \mathbf{x}_2 + b = -1$, it can be deduced that $\mathbf{w}(\mathbf{x}_1 - \mathbf{x}_2) = 2$. For the separating

hyperplane $\mathbf{w} \cdot \mathbf{x} + b = 0$, the orientation of the hyperplane can be given as $\mathbf{w}/\|\mathbf{w}\|_2$ (where $\|\mathbf{w}\|_2$ is square root of $\mathbf{w}^T \mathbf{w}$). The distance between the two canonical hyperplanes is equal to the projection of $(\mathbf{x}_1 - \mathbf{x}_2)$ onto $\mathbf{w}/\|\mathbf{w}\|_2$, that is, $(\mathbf{x}_1 - \mathbf{x}_2) \cdot \mathbf{w}/\|\mathbf{w}\|_2 = 2/\|\mathbf{w}\|_2$. As margin is half the distance between the canonical hyperplanes it can be given as $\gamma = 1/\|\mathbf{w}\|_2$. Therefore, maximizing the margin implies minimizing the function given by Eq. (3.2), subject to constraints given in Eq. (3.3) (Cortes and Vapnik, 1995; Vapnik, 1995).

$$\text{Minimize } \frac{1}{2} \|\mathbf{w}\|_2^2 \quad (3.2)$$

$$\text{Subject to constraints, } y_i [\mathbf{w} \cdot \mathbf{x}_i + b] \geq 1 \quad \forall_i \quad (3.3)$$

Eq.s (3.2) and (3.3) represent a constrained optimization problem. This can be reduced to a minimization problem using the following *Lagrange function* comprising a sum of the objective function and p constraints multiplied by their respective *Lagrange multipliers* (Everett III, 1963; Bertsekas, 2014). This function is called a *primal* and is given as,

$$L(\mathbf{w}, b) = \frac{1}{2} \mathbf{w} \cdot \mathbf{w} - \sum_{i=1}^p \beta_i [y_i \mathbf{w} \cdot \mathbf{x}_i + b - 1] \quad (3.4)$$

Where, β_i are the Lagrange multipliers and $\beta_i \geq 0$. To minimize Eq. (3.4), partial derivatives with respect to $\mathbf{w}$ and b can be equated to zero. That gives,

$$\frac{\partial L(\mathbf{w}, b)}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^p \beta_i y_i \mathbf{x}_i = 0 \quad (3.5)$$

$$\frac{\partial L(\mathbf{w}, b)}{\partial b} = -\sum_{i=1}^p \beta_i y_i = 0 \quad (3.6)$$

From Eq. (3.5), we have

$$\mathbf{w} = \sum_{i=1}^p \beta_i y_i \mathbf{x}_i \quad (3.7)$$

and from Eq. (3.6) we have $\sum_{i=1}^p \beta_i y_i \mathbf{x}_i = 0$. Substituting these two deductions back into

$L(\mathbf{w}, b)$, we get a *Wolfe dual formulation* (Wolfe, 1961). It can be given as,

$$W(\beta) = \sum_{i=1}^p \beta_i - \frac{1}{2} \sum_{i,j=1}^p \beta_i \beta_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \quad (3.8)$$

which must be maximized with respect to the following constraints,

$$\beta_i \geq 0 \quad \text{and} \quad \sum_{i=1}^p \beta_i y_i = 0 \quad \forall_i \quad (3.9)$$

This dual objective function is *quadratic* with respect to the Lagrange multipliers β_i , and it also has constraints in the form of Eq. (3.8). Hence this problem can be termed as a *constrained quadratic programming* problem.

But it can be seen given in the generalization theorem that: The bound is independent of the dimensionality of the input space, has not been used yet in the development of the formulation. From the dual objective function in Eq. (3.8), it can be seen that the input vectors $\mathbf{x}_i$ only appear in a scalar product. To get a different representation these data vectors can be mapped to another space of different dimensionality, called a *feature space*, by,

$$\mathbf{x}_i \cdot \mathbf{x}_j \rightarrow \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) \quad (3.10)$$

Where, $\phi(\cdot)$ represents the mapping function. Here, if we revisit Figure 3.2 (a), we can see that the two clusters of data are intermeshed and overlying. If there could be some

transformation possible to shift the data represented by the circles into a plane parallel to that of the paper, towards the reader and the data represented by the parallelograms into a plane parallel to the paper, away from the reader, the two clusters of data could be accurately classified. That is the beauty of changing the dimension of the input space. Also, as the bound is independent of the dimensionality of the space, this is a valid transformation. This mapping transformation is called as a *kernel* transformation and is given as,

$$K \mathbf{x}_i \cdot \mathbf{x}_j \rightarrow \phi \mathbf{x}_i \cdot \phi \mathbf{x}_j \quad (3.11)$$

There are limits on the possible choice of kernel out of which one obvious restriction is that the feature space has to be an *inner product space* called a *Hilbert space* and that is where the inner product is consistently defined. Therefore, a proper choice of kernel function is necessary to minimize the training error. The introduction of kernel with its implied mapping to feature space is called *kernel substitution*. Some of the common kernel choices are,

$$\left. \begin{array}{ll} K \mathbf{a}, \mathbf{b} = \mathbf{a} \cdot \mathbf{b} & \text{Linear kernel} \\ K \mathbf{a}, \mathbf{b} = \nu \mathbf{a} \cdot \mathbf{b} + c^d & \text{Polynomial kernel} \\ K \mathbf{a}, \mathbf{b} = e^{-\gamma \|\mathbf{a} - \mathbf{b}\|^2} & \text{Gaussian RBF kernel} \\ K \mathbf{a}, \mathbf{b} = \tanh \nu \mathbf{a} \cdot \mathbf{b} + c & \text{Sigmoid kernel} \end{array} \right\} \quad (3.12)$$

Where, $K \mathbf{a} \cdot \mathbf{b} = \phi \mathbf{a} \cdot \phi \mathbf{b}$ for the vectors $\mathbf{a}, \mathbf{b}$ and RBF stands for radial basis function.

For a binary classification with a given choice of kernel, the dual objective function in Eq. (3.8) becomes,

$$W \beta = \sum_{i=1}^p \beta_i - \frac{1}{2} \sum_{i,j=1}^p \beta_i \beta_j y_i y_j K \mathbf{x}_i \cdot \mathbf{x}_j \quad (3.13)$$

Subject to the constraints given in Eq. (3.9). For a data point with, $y_i = +1$ we note that,

$$\min_{i|y_i=+1} \mathbf{w} \cdot \mathbf{x} + b = \min_{i|y_i=+1} \left[\sum_{j=1}^p \beta_j y_j K(\mathbf{x}_i \cdot \mathbf{x}_j) \right] + b = 1 \quad (3.14)$$

Similarly, an expression can be written for data point with $y_i = -1$. On combining both the equations, we get,

$$b = -\frac{1}{2} \left[\min_{i|y_i=+1} \left(\sum_{j=1}^p \beta_j y_j K(\mathbf{x}_i \cdot \mathbf{x}_j) \right) + \max_{i|y_i=-1} \left(\sum_{j=1}^p \beta_j y_j K(\mathbf{x}_i \cdot \mathbf{x}_j) \right) \right] \quad (3.15)$$

Therefore, to construct a SVM binary classifier, the data $\mathbf{x}_i, y_i$ need to be substituted in Eq. (3.13), subject to the constraints given in Eq. (3.9). After finding the optimal value of the Lagrange multiplier (β_i^*), the bias can be calculated from Eq. (3.15). Thus, for a new input vector $\mathbf{n}$, the predicted class is then based on the sign of,

$$f(\mathbf{n}) = \sum_{i=1}^p \beta_i^* y_i K(\mathbf{x}_i, \mathbf{n}) + b^* \quad (3.16)$$

Where, b^* denotes the optimum value of the bias. With the hyperplane of maximal margin, only those points that lie closest to the hyperplane have $\beta_i^* > 0$, and these points are the *support vectors*. All other points have $\beta_i^* = 0$, and the decision function is independent of these samples. That is, even if some of these samples were removed, there would not be any change in the position or orientation of the current hyperplane.

3.3 Multiclass Support Vector Machine Classifier

From the mathematical description given in Section 3.2, it is clear that SVM is a potential binary data classifier. But, unfortunately not always is the data classification as simple as

classifying two classes of data. How does the SVM deal with multiclass classification? Many schemes are outlined to deal with multiclass SVM classification, they are,

- When small numbers of classes are involved a *directed acyclic graph* (DAG) approach can be used (Platt et al., 2000). The design is illustrated in Figure 3.3. Here each node defines a binary classification. For example, in a 3-class classification problem, the first node is a classifier making the binary decision, label 1 versus label 2, say. Depending on the outcome of this decision, the next steps are the decisions class 1 versus class 3 or class 2 versus class 3.
- The other method is to use *one versus all* (OVA) approach (Lee et al., 2004). The design is illustrated in Figure 3.3. Here each node defines a binary classification. For example, in a 3-class classification problem, the first node is a classifier making the binary decision, label 1 versus label 2, say. Depending on the outcome of this decision, the next steps are the decisions class 1 versus class 3 or class 2 versus class 3. If there are S classes of data, S separate SVMs are created. The k^{th} SVM is trained with data from class k as positively labelled samples and all others as negatively labelled. Associated with k^{th}

SVM the decision function becomes, $f^k(\mathbf{n}) = \sum_{i=1}^p \beta_i^k y_i K(\mathbf{x}_i, \mathbf{n}) + b^k$ and the new data

vector $\mathbf{n}$ is assigned to the class with maximum value for $f^k(\mathbf{n})$. Though this is a very popular method of the multiclass classification, it has a shortcoming. For instance, consider that there are S classes of data and there are equal numbers of data points in each of the S classes then while training the k^{th} SVM, 99% of the data is used with negative labels and only 1% of it is trained with positive labels which creates a data imbalance and so a skewed classifier.

- The next method is a *one versus one* (OVO) approach (Hastie and Tibshirani, 1998) If there are S classes of data, $S(S-1)/2$ binary class SVMs are created. Each one trains data from two different classes and after training the data from the i^{th} and j^{th} classes, a voting strategy is used to classify the data while testing. If a binary classifier, while testing sample x , says that the sample belongs to i class, then the vote for class i is incremented by 1. Lastly, sample x is classified into that particular class, which carries the maximum votes for it. To illustrate the working of OVO approach, consider that there are 3 classes of data. This implies that there will be $3(3-1)/2 = 3$ binary classifications. The new data is tested on each of these three binary SVM classifiers, as shown in Figure 3.4.

Hsu and Lin (Chih-Wei and Chih-Jen, 2002; Hsu et al., 2003) compared the aforementioned three approaches to tackle a multiclass classification problem using the SVM. Experimental studies followed by analysis revealed that OVO was the best for practical application because of its advantages like the high generalization accuracy and minimum training time. Hence, in this work the OVO approach has been adapted for the multi-class classification.

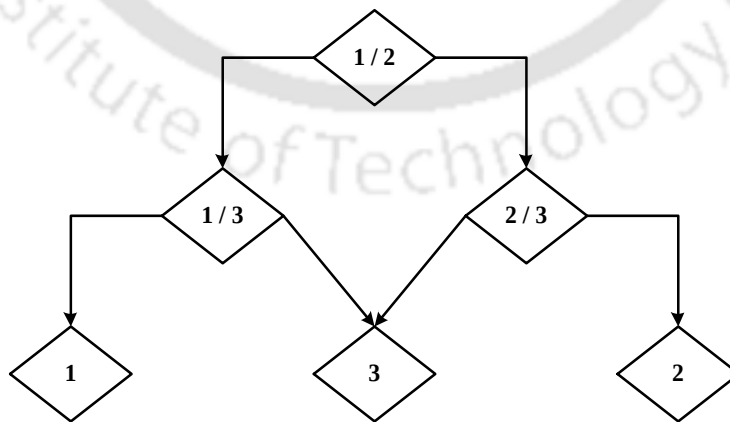


Figure 3.3 Working of the DAG

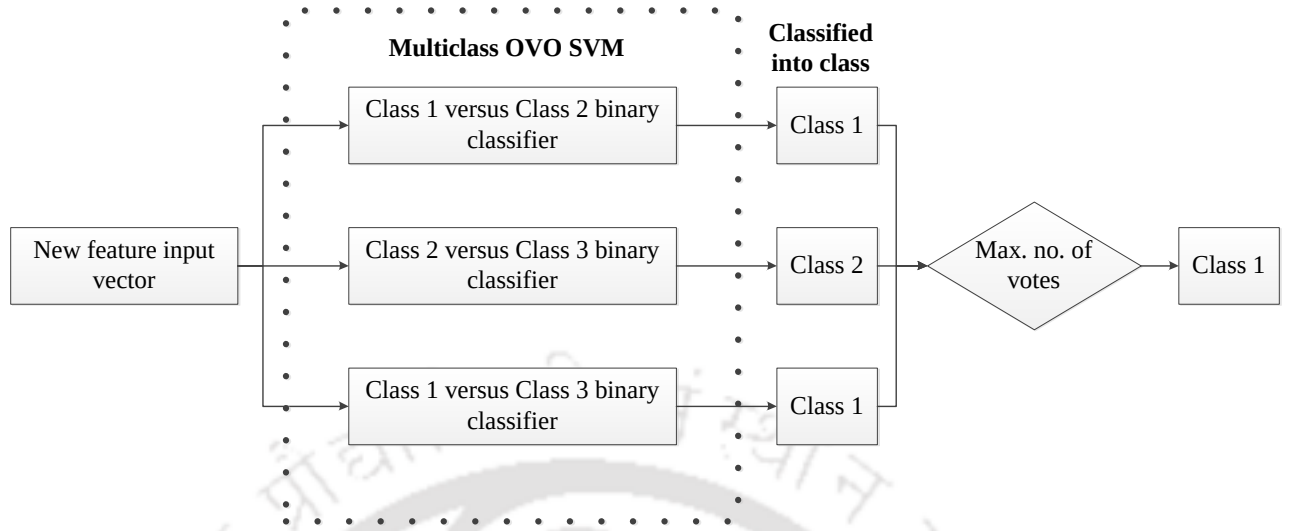


Figure 3.4: The OVO classifier

3.4 C-SVM: Soft Margin Approach

Data points with big values of β_i^* have a large effect on the orientation of the hyperplane and a substantial influence in the decision function. These data points may be correct but unfamiliar or may relate to incorrectly labelled or abnormal instances and so when training is complete, these abnormal data points should be rejected, if necessary; this is called *data cleaning*. Even a correctly labelled outlier has excessive impact and so the best method to reduce its impact is through the use of *soft margin* approach.

The influences of the outliers and the noise can be reduced by introducing a soft margin, and the scheme commonly used is,

- Using a U_1 error norm, a box constraint is introduced on the Lagrange multipliers and the same learning task (Eq. (3.13) and Eq. (3.9)) is performed,

$$0 \leq \beta_i \leq C \quad (3.17)$$

Proper values of these parameters can be found by doing validation studies. With limited data, cross-validation, or rotation estimation, may be used, in which the data is randomly partitioned into subsets and rotated successively as training and validation data and this would lead to less bias, if the validation set is small.

Data points inside the margin band or lying on the wrong side of the hyperplane are called *margin errors*. A negative slack variable is introduced in Eq. (3.3) to take care of the margin errors, as,

$$y_i [\mathbf{w} \cdot \mathbf{x}_i + b] \geq 1 - \xi_i \quad \forall_i \quad (3.18)$$

The objective is to minimize the sum of errors $\sum_{i=1}^p \xi_i$ along with $\|\mathbf{w}\|^2$. That is,

$$\min \left[\frac{1}{2} (\mathbf{w} \cdot \mathbf{w}) + C \sum_{i=1}^p \xi_i \right] \quad (3.19)$$

If $\xi_i > 0$, it means that there is a margin error. These equations can be put together to formulate a primal Lagrange function,

$$L(\mathbf{w}, b, \beta, \xi) = \frac{1}{2} (\mathbf{w} \cdot \mathbf{w}) + C \sum_{i=1}^p \xi_i - \sum_{i=1}^p \beta_i [y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1 + \xi_i] - \sum_{i=1}^p r_i \xi_i \quad (3.20)$$

With Lagrange multipliers $\beta_i \geq 0$ to handle Eq. (3.18) and $r_i \geq 0$ to handle $\xi_i \geq 0$. At optimum the partial derivatives with respect to $\mathbf{w}$, b , and ξ_i give,

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^p \beta_i y_i \mathbf{x}_i = 0 \quad (3.21)$$

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^p \beta_i y_i = 0 \quad (3.22)$$

$$\frac{\partial L}{\partial \xi_i} = C - \beta_i - r_i = 0 \quad (3.23)$$

Using Eq. (3.21) the dual objective function can be found eliminating $\mathbf{w}$, and the equation is same as Eq. (3.13). From Eq. (3.23) and $r_i \geq 0$ constraints, we get, $0 \leq \beta_i \leq C$. The bias can be found using,

$$r_i \xi_i = 0 \quad (3.24)$$

$$\beta_i \{y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1 + \xi_i\} = 0 \quad (3.25)$$

From deductions, at *non-bound* point k (the points at which $0 \leq \beta_i \leq C$) the bias can be given as (Campbell and Ying, 2011),

$$b = y_k - \sum_{i=1}^p \beta_i y_i (\mathbf{x}_i \cdot \mathbf{x}_k) \quad (3.26)$$

Using the OVO approach for training a C-SVM with the p^{th} and q^{th} classes of data among t classes of data, the classification function is obtained by transforming Eq. (3.19) to,

$$\begin{aligned} \min_{\mathbf{w}^{pq}, b^{pq}, \xi^{pq}} & \left\{ \frac{1}{2} \mathbf{w}^{pq T} \mathbf{w}^{pq} + C \sum_t \xi^{pq} \right\} \\ \text{Subject to} & \quad \mathbf{w}^{pq T} \phi(\mathbf{x}_t) + b^{pq} \geq 1 - \xi_t^{pq}, \quad \text{if } \mathbf{x}_t \text{ in the } p^{\text{th}} \text{ class} \\ & \quad \mathbf{w}^{pq T} \phi(\mathbf{x}_t) + b^{pq} \leq -1 + \xi_t^{pq}, \quad \text{if } \mathbf{x}_t \text{ in the } q^{\text{th}} \text{ class} \\ & \quad \xi_{pq} \geq 0 \end{aligned} \quad (3.27)$$

It is very important step to identify the optimum SVM and the kernel parameter value combination to develop an effective classifier with a good generalization. In the next section, the cross validation and grid search techniques are explained, which have been used to select these parameters.

3.5 Cross Validation Technique and GST for Optimal Parameter Selection

In the present work the multi-class classification of faults is attempted using OVO C-SVM and Gaussian RBF kernel. The SVM algorithm using the RBF kernel is very sensitive to both the kernel parameter, γ , (as defined in Eq. (3.12)) and the SVM Lagrange penalty parameter or soft margin constant, C (as defined in Eq. (3.17) and Eq. (3.19). The separating hyper plane, while using the Gaussian RBF parameter, is a combination of bell shaped curves centered at support vectors. The width of each bell shaped curve is inversely proportional to the value of γ (given in definition of Gaussian RBF kernel in Eq. (3.12)). If the width of the curve is smaller than the minimum distance between the support vectors, the classification leads to over-fitting, and does not capture the tendency of the data. In case the width of the curve is greater than the maximum distance between the support vectors, it leads to under-fitting. C marks a trade-off between the simplicity of the classification curve and the misclassification of the training sample. If γ is very high, any amount of fine-tuning of C would not help in improving the classification of the algorithm. Figure 3.5 shows an example of under-fit, just-fit and over-fit data.

As explained in the previous section the kernel parameter, γ , and the penalty parameter, C , play a vital role in the accuracy of classification obtained and hence, the objective is to choose an optimum combination of (C, γ) to give the best classifier. To do this, a cross validation (CV) method along with a grid search technique (GST) is employed. In an n fold CV technique, the training set is divided into n subsets. Each of the $(n - 1)$ subsets of these n is used for training the SVM algorithm and remaining one is used for testing. By completing the testing with each of the n sets the SVM parameters are estimated. Thus, trained algorithm with the best SVM parameters is further tested with totally new data. The advantage of cross

validation technique is that all the test data sets are independent, and thus the reliability of the results can be improved. In the present work a 5-fold cross validation technique is used.

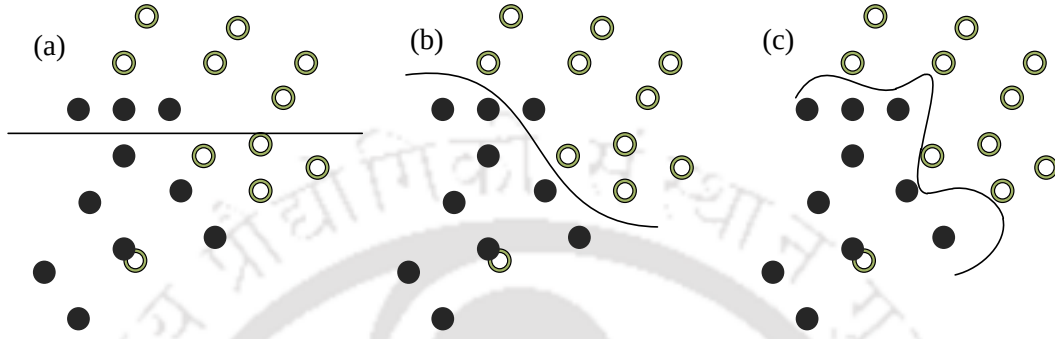


Figure 3.5 Different types of classifier fits, (a) under-fit classifier, (b) just-fit classifier and (c) over-fit classifier

Along with this CV technique the GST is used to refine the search grid range for the selection of (C, γ) parameters. The range of these grid points is generally selected on a logarithmic range so that a wide range is covered for both these parameters. In the current study the values of C are varied to take values from a geometric progression sequence of $3^{-5}, 3^{-4} \dots 3^{15}$, similarly γ is allowed to take values from geometric progression sequence $3^{-15}, 3^{-14} \dots 3^3$. Initially, a broad range of values is chosen for both C and γ parameters but once the CV accuracy is obtained, the grid is refined near the point giving maximum CV accuracy. Finally, the C and γ value combination giving the highest CV accuracy is retained. After this step, the SVM model is trained with whole training dataset at the given C and γ value combination to obtain the final classifier. The flowchart of the parameter selection is given in Figure 3.6.

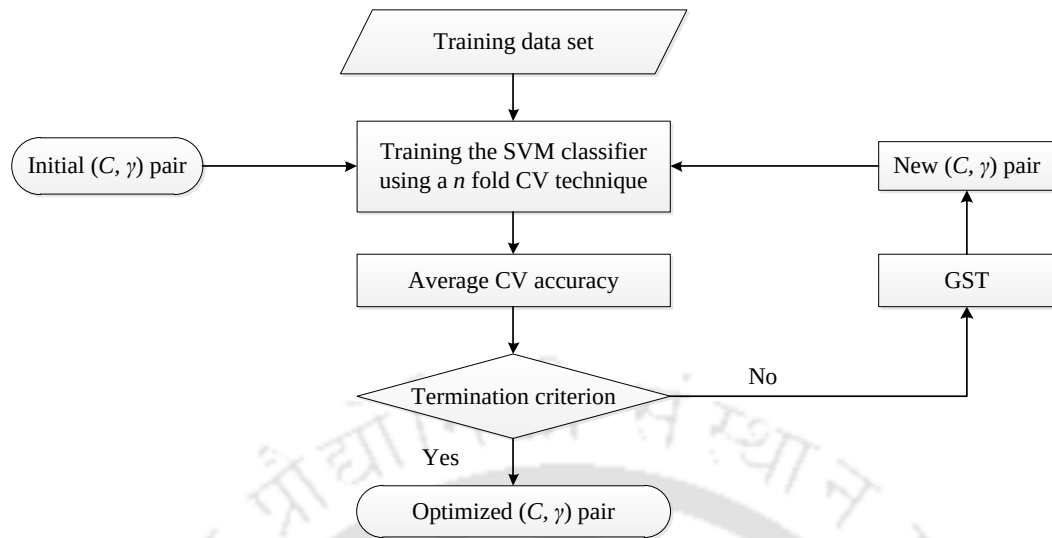


Figure 3.6: The CV method along with GST for optimum selection of SVM parameters

As it must have been understood by now, the feature values of different classes (faults) of data play a major role in the accurate representation of them in the feature space. Therefore, it is important that the best features representing different classes of data are systematically identified. In the next section, the selection of features using wrapper model is presented.

3.6 Wrapper Model for Feature Selection

The feature selection is very crucial in obtaining the best performance of a classifier. A good feature depicts the characteristics of the fault and remains unaltered with any change in external noise related factors or the operating parameters of the machinery. Selecting the right features decreases the training time and improves the interpretability, and thereby increasing the classification accuracy of the model using them. On the other hand, a bad feature quickly gets affected by the noise, carries redundant information, occupies a lot of memory and is computationally expensive to handle. Therefore, it is important to make right choice of features or feature combinations so that the desired classification performance is obtained. To

make such a choice, a wide selection of feature trials is done in each of the domains of the analysis considered and the selection of features is later done using a Wrapper method.

The wrapper model is simple to comprehend, less time consuming and is an efficient technique for feature selection. It is simple because it makes use of the prediction accuracy as the evaluation criterion for the selection of the features. Incidentally, the main objective of real time machine learning systems is to have improved correct predictions which means having an improved accuracy of the classifier. Therefore, classification accuracy can be taken as a criterion for the feature selection. There is no need to go for complex algorithms for feature selection because of their limitations to apply for practical problems.

For any chosen machine learning algorithm a general flowchart of applying a wrapper model is given in Figure 3.7. The procedure involves, (1) identifying a set of features from which the best features need to be selected, (2) training and testing the classifier with a class labeled feature data set, taking one feature at a time, (3) obtaining the classification accuracy of the classifier for the said feature, (4) repeat Steps 1 to 3 for all the features, (5) try different permutations and combinations of the features and repeat Steps 1 to 3, (6) identify the best performing feature(s), and (7) finalize the features. It is to be noted that if there are f number of features, there are various feature combinations possible (taking two features at a time, three at a time, and so on). Trying all the exhaustive list of feature combinations is not very effective and is also time consuming. Hence, after Step 5 the best performing features are chosen and a judicious combination of the best performing features is only carried out.

It is also important to note that there is no rule saying that if a particular feature works well for one type of data (e.g., features extracted from time domain data), the same features work for the other domains in which the data is used as well. Hence, feature selection needs to be performed for each of the domains independently. In the subsequent chapters, the feature selection performed for each of the time, frequency and time-frequency domains will be presented.

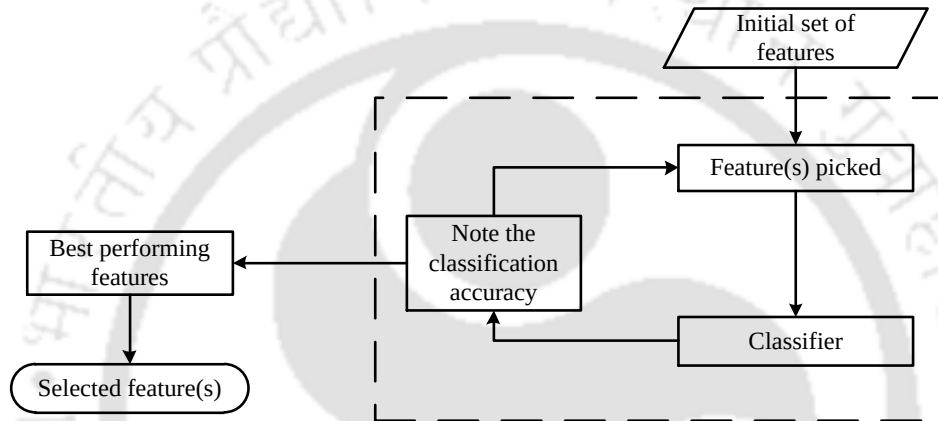


Figure 3.7: Feature selection using the wrapper model

3.7 Fault Diagnosis Procedure for the Current Study

The CP fault diagnosis is attempted in this study with the help of SVM algorithm described in the previous sections. The flowchart of the classification algorithm is presented in Figure 3.8. The fault classification is carried out in five steps: (a) the fault data is acquired from the experimentation, (b) the data is converted to the domain in which the analysis is required to be performed, (c) features are extracted systematically from the data, (d) best feature(s) (from the extracted lot) are selected, (e) the algorithm is trained with the selected features and the fault classification is performed. The first step involves the collection of the fault data from the CP, which is carried out using the experimental setup and procedure described in Chapter 2.

Once the raw signals are acquired, in the second step, these signals are converted to the required domain (frequency or time-frequency). Note that the signals acquired are already in time-domain so if fault classification needs to be performed in the time-domain, no further transformation is required. The third step revolves around extracting useful statistical features from the data in an organized format. Each of the faults is then assigned label (for example, features for the HP are assigned label 1, SB1 is given label 2, SB2 is given label 3, and so on). This step ensures that the data is ready to train a supervised machine learning algorithm.

The performance of the classifier majorly depends upon the choice of the features. Once the feature data is ready, in the fourth step using the wrapper model the best feature(s) are systematically selected. Note that the fourth and the fifth steps of the developed methodology go hand in hand. To train the SVM algorithm the feature data is randomly divided into training and testing data. The last step involves training of the SVM algorithm with the training sample (also in this step the SVM classifier is constructed using C and γ parameters selected from the CV and GST methods). The training of multiclass SVM algorithm is done using the training data and with the help of Gaussian RBF kernel. This model is optimized using the 5-fold cross validation and grid search techniques. Using the optimized algorithm and the finalized feature(s) from the wrapper model, the SVM is again trained with whole training data and is tested with the test data.

Lastly, the performance measure of the classifier is quantified as prediction rate/ percentage classification accuracy, as given in Eq. (3.28).

$$\% \text{ Classification accuracy} = C_a = \frac{\text{Number of correctly predicted data}}{\text{Total number of testing data}} \times 100\% \quad (3.28)$$

The classification of the faults is tested in two cases,

- I. The algorithm is trained using the fault data available at an operating speed and the algorithm is tested with data present at the same operating speed. This is generally the case with real-time systems.
- II. Further, as a separate study in the second case the faults are attempted to be classified at intermediate speeds, i.e., the algorithm is trained at two distinct speeds and tested at an intermediate speed. This is important in cases where the fault training data is not available at the present operating speed of the CP.

In the present work, LIBSVM (Chang and Lin, 2011) has been used as the SVM tool in MATLAB environment. For the analysis a CPU of HCL info systems limited make having Intel® Core™ i5-2400 CPU (3.10 GHz) with 8 GB RAM and 32 bit OS is used.

3.7.1 Noise Corruption Analysis

Though the fault data has been generated experimentally (and not in an isolated condition), in general plant operation, the signals may be further corrupted due to interfering signals from surrounding operating machines. Also, the change due to manufacturing tolerances in different CP systems can affect the assemblies and CP performance resulting in noise generation (or deviations in signals). In such cases, it would be very challenging to identify the CP faults, especially when the machine learning algorithm is trained using the lab generated data (which has much more controlled conditions than the real plant operation), and tested using the actual plant data.

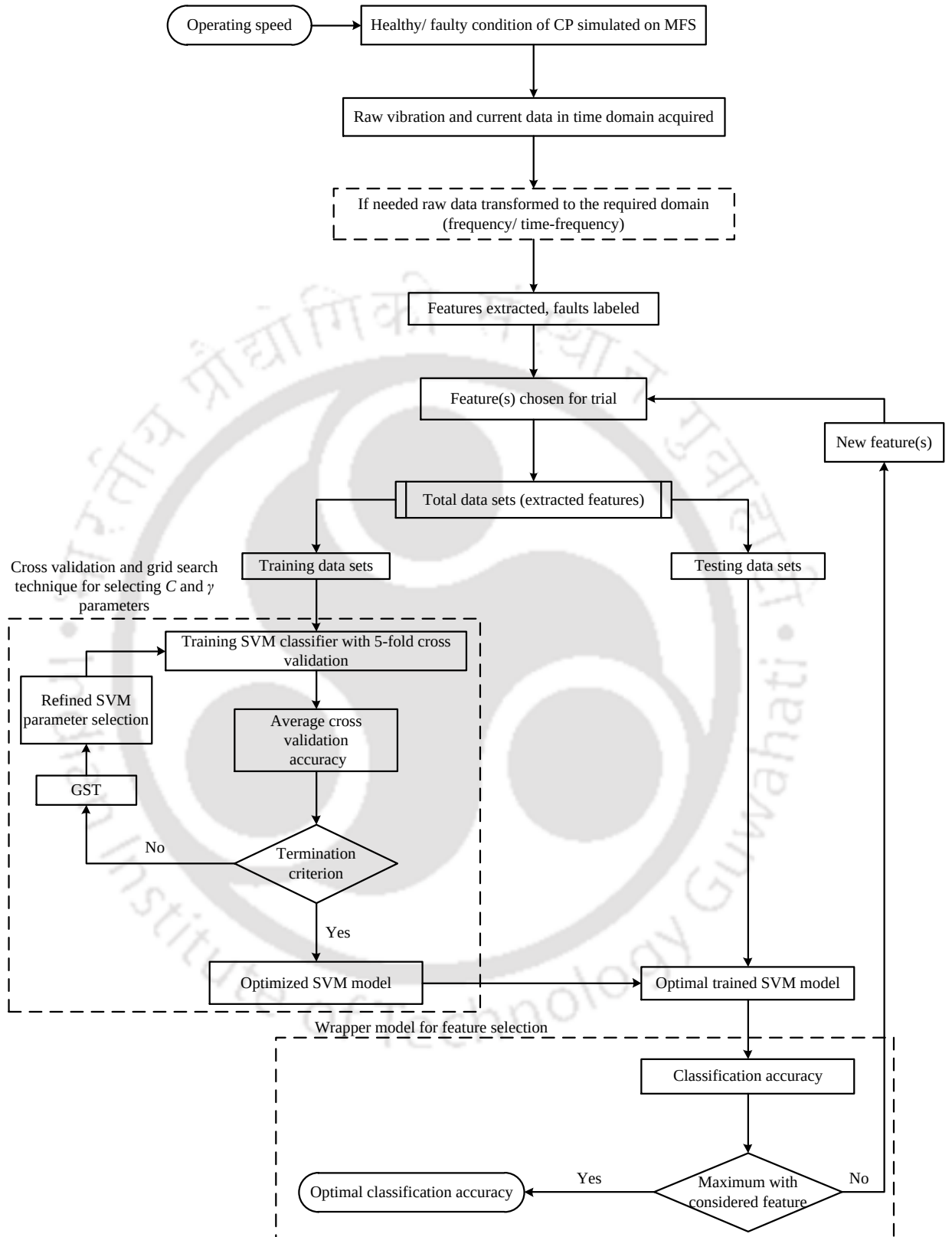


Figure 3.8: The flowchart for the proposed methodology

Therefore, to check the versatility of the algorithm the signals generated from the experiments are corrupted using additive white Gaussian noise (AWGN). White noise refers to a uniform power signal across the wide frequency band for the system. In an industry, we do not always know the frequency of noise. Also the probability theory suggests that the summation of many random processes generally has Gaussian distribution with zero-mean. Therefore, a white Gaussian noise model has been chosen. A noisy signal can be given as,

$$S = S_0 + N \quad (3.29)$$

where S_0 represents the original signal and N signifies the signal-independent noise. The amplitude of the noise is of Gaussian distribution and depends on the desired signal-to-noise ratio (SNR). The SNR in dB is defined as,

$$\text{SNR}_{dB} = 10 \log_{10} \left(\frac{A_s}{A_N} \right)^2 \quad (3.30)$$

where A_s and A_N are, respectively, the amplitude of the signal and the noise. In this study the algorithm is trained with the features extracted from the originally acquired data and it is tested using noisy data. Here, as the training and testing data is very different, the CV and GST techniques are not used for the parameter selection. Instead the parameters are selected using a systematic search. For a particular value of C different values of γ are tried and thus the search is continued until optimum combination is obtained. The flowchart for the methodology is given in

Figure 3.9.

Before adding AWGN to the signal, it is good to estimate how much noise is already present in the signal. To estimate the strength of the noisy signal the CP is detached from the drive shaft (i.e., the belt is removed) and the sensor readings are taken from their designated positions. Now using the equation given in 3.30 the SNR_{dB} is calculated considering HP condition as the original signal. This exercise is done for time-domain signal acquired at 5 kHz sampling rate. The SNR_{dB} values thus found are tabulated in Table 3.1.

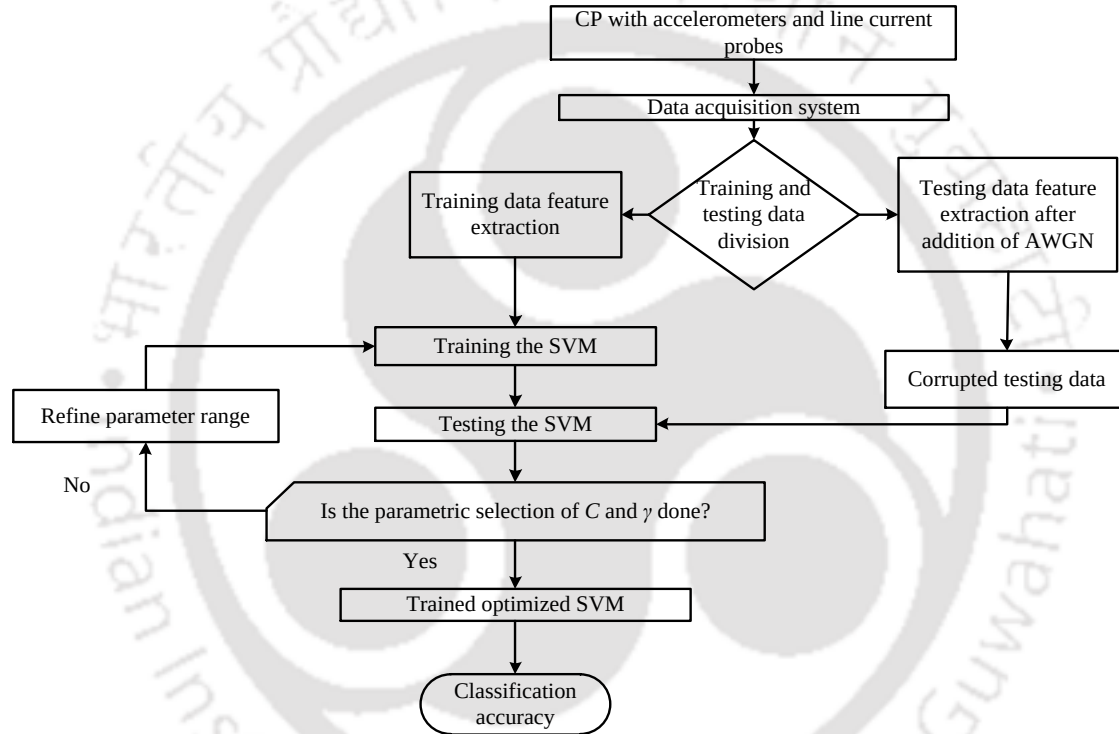


Figure 3.9: The flowchart for fault classification using noisy test data

Table 3.1: SNR_{dB} value for different speeds using accelerometers and line current signals

Speed		A-1x	A-1y	A-1z	LCF
30Hz	SNR_{dB}	56.20	49.65	50.86	25.20
	% noise	0.15	0.33	0.29	5.49
35 Hz	SNR_{dB}	57.21	52.34	53.19	22.86
	% noise	0.14	0.24	0.22	7.20
40 Hz	SNR_{dB}	58.22	55.63	54.02	25.90
	% noise	0.12	0.17	0.20	5.07
45 Hz	SNR_{dB}	60.92	61.51	57.55	33.12
	% noise	0.09	0.08	0.13	2.21

50 Hz	SNR _{dB}	62.73	58.56	57.65	42.43
	% noise	0.07	0.12	0.13	0.76
55 Hz	SNR _{dB}	64.72	60.23	59.74	52.06
	% noise	0.06	0.10	0.10	0.25
60 Hz	SNR _{dB}	67.51	58.74	60.58	54.95
	% noise	0.04	0.12	0.09	0.18
65 Hz	SNR _{dB}	66.94	57.59	58.09	54.57
	% noise	0.04	0.13	0.12	0.19
Average	SNR _{dB}	61.81	56.78	56.46	38.88
	% noise	0.09	0.16	0.16	2.67

From the table, it can be seen that the SNR value for the original signal is around 38 – 61 dB (0-3% noise).

3.8 Summary

The fault classification methodology (i.e., the SVM) has been discussed in the present chapter. The theory behind the binary and multiclass approaches to SVM is explained in detail. The theory behind effect of different parameters on the classification accuracy of SVM has been explained. The procedure for optimal selection of these parameters using cross-validation and grid-search techniques employed in this work is demonstrated. Thereafter, the wrapper model method for the selection of the best features is given out. The procedure for feeding the data into the SVM for the training and the classification has been presented. Also to check the robustness of the algorithm, a methodology is presented to test it with corrupted data. In subsequent chapters, the performance of the developed methodology is presented in time-domain, frequency-domain and time-frequency domain.

CHAPTER 4: Preliminary Time and Frequency Domain based Centrifugal Pump Fault Diagnosis

4.1 Introduction

This chapter presents the preliminary results of analyses based on the time and frequency domain studies. As explained in Chapter 2, Section 2.5, the experimentation has been performed in three phases. In this Chapter, the results based on experiments conducted in phase I and phase II are presented. Also, it is important to emphasize that the results of these preliminary studies paved the way for the phase III experimentation and subsequent studies presented in Chapters 5 to 9.

In phases I and II of the experimentations, vibration signals have been acquired from the CP laboratory set-up, and the fault diagnosis is performed based on the one-versus-one multi-class SVM based methodology. The objective of the analysis conducted based on phase I experiments is to identify different levels of suction blockages on the CP and the impending cavitation condition. On the other hand, in phase II experiments, mechanical faults (impeller defects) and a combination of mechanical and hydraulic anomalies (impeller defects and suction blockages) are considered, and their identification is attempted.

The data is acquired at a sampling rate of 20 kHz with 2000 samples per dataset, and 350 such datasets are obtained for each faulty or non-faulty condition. The signals from the faulty and non-faulty CP states are captured independently in both time and frequency domains. Statistical features are extracted from each of these domains and are used as inputs to the SVM based classifier.

In this chapter, the SVM parameters and kernel parameters (and their combination) are selected by a systematic trial and error method to build an optimal SVM classifier. In the present work, γ is varied from 5×10^{-4} to 5×10^3 , and C is varied from 1×10^{-1} to 1×10^4 , and the best parameter combination is chosen to get the maximum classification accuracy. This procedure is repeated for all the feature trials and each fault classification case. The fault diagnosis is performed when the classifier is trained and tested at the same operating conditions of the CP. During experimentation, it was observed that the fault manifestations change with the CP speed. However, a suitable classifier is expected to segregate faults irrespective of the operating speed. Therefore, to check the robustness of the algorithm, it is tested for a wide range of CP speeds. 80% of the data is used for training the algorithm, and 20% of it is used for testing. The training and testing data sets are non-overlapping. By ‘non-overlapping’ it is meant that the training and testing data sets are completely distinct.

4.2 Phase I Experimentation – Time Domain Analysis

To recapitulate, the experiments performed in phase-I involved acquiring vibration signatures from the CP bearing housing and the CP casing locations for the different suction blockage conditions (SB1 to SB4), the cavitation commencement condition (C0), and the healthy pump condition (HP). This section presents the initial results for the time-domain based CP fault classification for Phase-I experiments.

4.2.1 Fault Feature Extraction and Selection

From the literature survey summary presented in Table 1.8, features including the mean, standard deviation, skewness, kurtosis, crest factor, and entropy are selected to perform the

fault diagnosis in time-domain. The definition of the features and their physical significance are presented below

Mean (μ): The mean, is simply the average of all the data points in the time-domain data. With the change in fault severity or the damage in the system, the average load on the CP system is expected to increase/ decrease depending upon the fault characteristics. Therefore, the mean is expected to be an essential feature for the CP fault study. It is given as,

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (4.1)$$

Where, N is the number of points in the sample, x_i is the amplitude of the i^{th} data point in the sample.

Standard deviation (σ): Standard deviation is a dimensional quantity that measures the dispersion of the distribution from the mean value. It is a measure of the active energy/ power content of the signals. A low value of standard deviation implies that the data is placed close to the mean and a higher value implies that the data is spread out. It is given as,

$$\sigma = \sqrt{\sum_{i=1}^n (x_i - \mu)^2 / (n-1)} \quad (4.2)$$

Skewness (χ): The skewness is a non-dimensional feature that measures the degree of asymmetry of the probability distribution of real-valued variables around their mean. The value of skewness could be positive or negative or undefined. Physically, a negative skewness value represents a longer tail on the left side of the probability density function when compared to its right side, and the bulk values (including the median) lie to the left of the

mean. Similarly, a positive skewness value indicates an asymmetry towards the positive values. If skewness value is zero, the distribution is even about the mean. It is expressed as,

$$\chi = \frac{\frac{1}{N} \left(\sum_{i=1}^N x_i - \mu \right)^3}{\sigma^3} \quad (4.3)$$

Kurtosis (κ): The kurtosis is also a non-dimensional feature that reflects the extent of flatness or spikiness of the probability distribution of the signal. It gives an amount of the size of the tails of the distribution and is used as an indicator of significant peaks in a data set. Its value is generally expected to be high as the severity of the faults increases and vice versa. It is expressed as,

$$\kappa = \frac{\frac{1}{N} \left(\sum_{i=1}^N x_i - \mu \right)^4}{\sigma^4} \quad (4.4)$$

Crest factor (P): The crest factor is defined as the ratio of peak (maximum) value to the RMS of the signal. It is a non-dimensional feature that measures the spikiness of the signals. Also, it reflects the impulse corresponding to continuous signal at a lower level. It increases if peaks appear in the time-domain signal. It is given by,

$$P = \frac{\max |x_i|}{RMS} \quad (4.5)$$

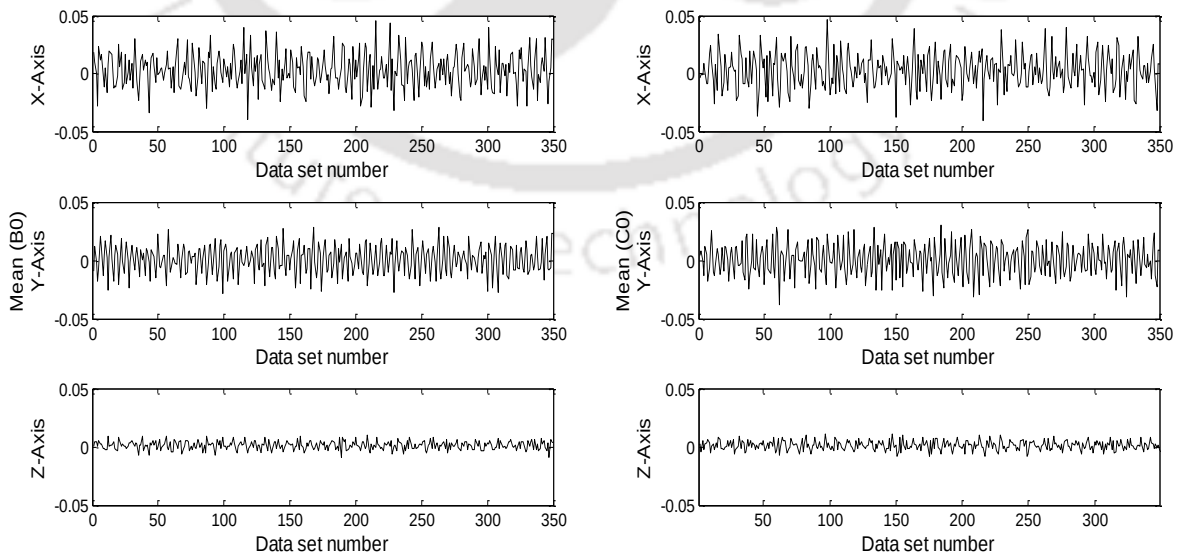
Where, the RMS value of the signal is the normalized second statistical moment of it. It is the square root of the arithmetic mean of the squares of the data value and is defined by,

$$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^N |x_i|^2} \quad (4.6)$$

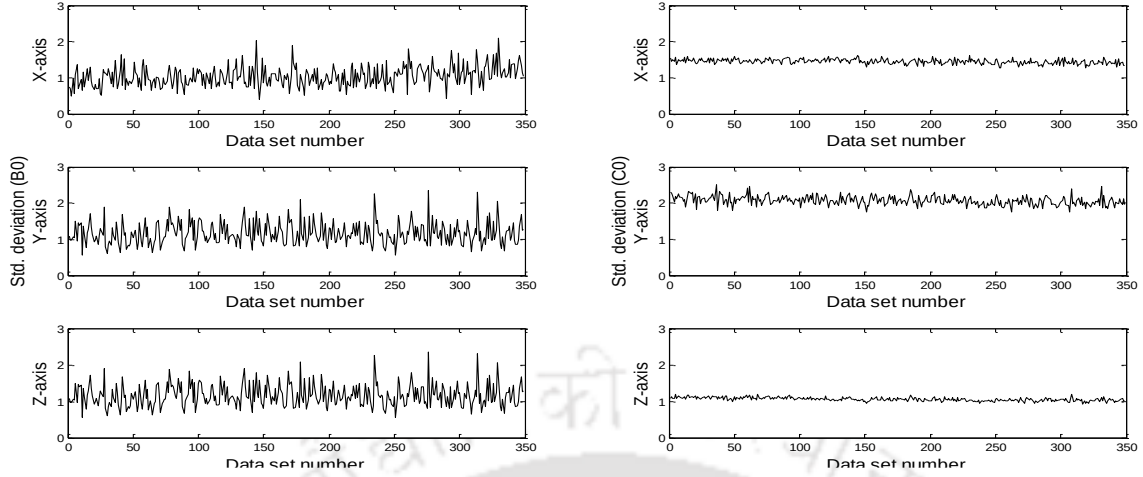
Entropy (S): The entropy is a generic measure of system disorganization and is mathematically presented as,

$$S = -\sum_{j=1}^N p(x_j) \log_{10} p(x_j) \quad (4.7)$$

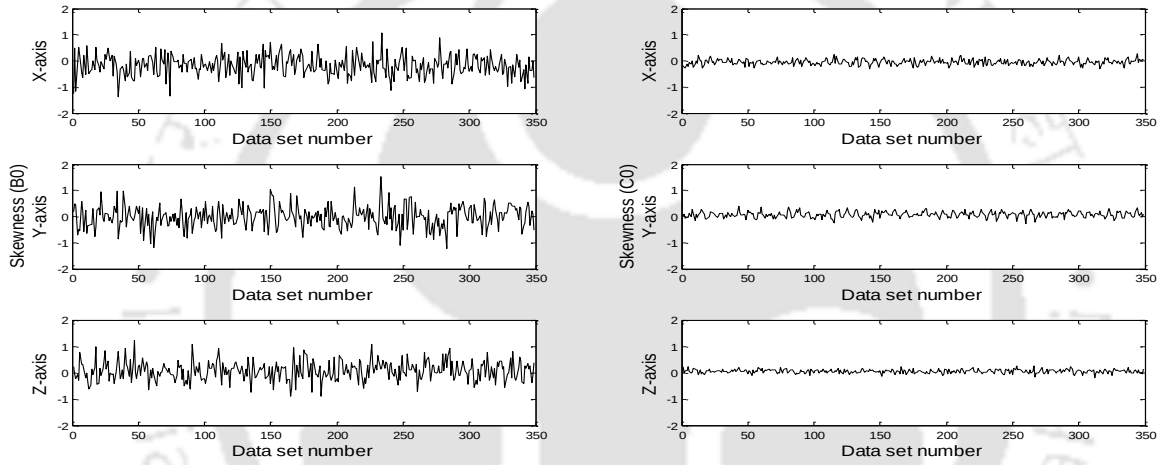
where, $x = \{x_1, x_2, \dots, x_N\}$ is a set of random phenomena, $p(x_j)$ is the probability of random phenomenon x_j . Aperiodic signals have highest entropy and periodic ones have significantly lower values of it. This characteristic of entropy in time domain can thus be used to distinguish normal operating condition of the pump to the faulty condition. On the whole, six statistical features are extracted from each of the three orthogonal directions of vibration data acquired from two accelerometers. Figure 4.1 shows six features plotted for the HP and C0 conditions at 65 Hz.



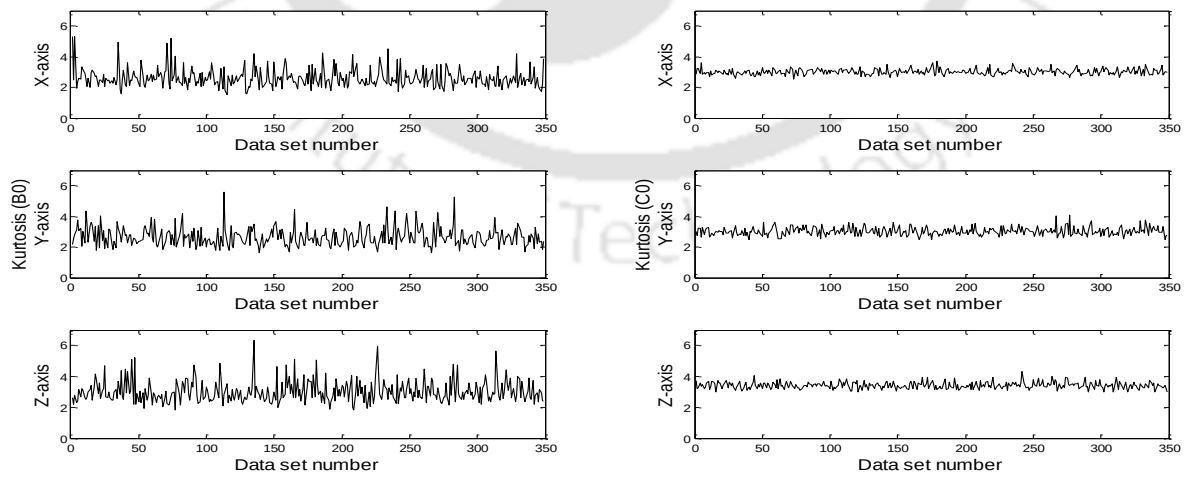
(a) Variation of mean with dataset for (left) HP and (right) C0



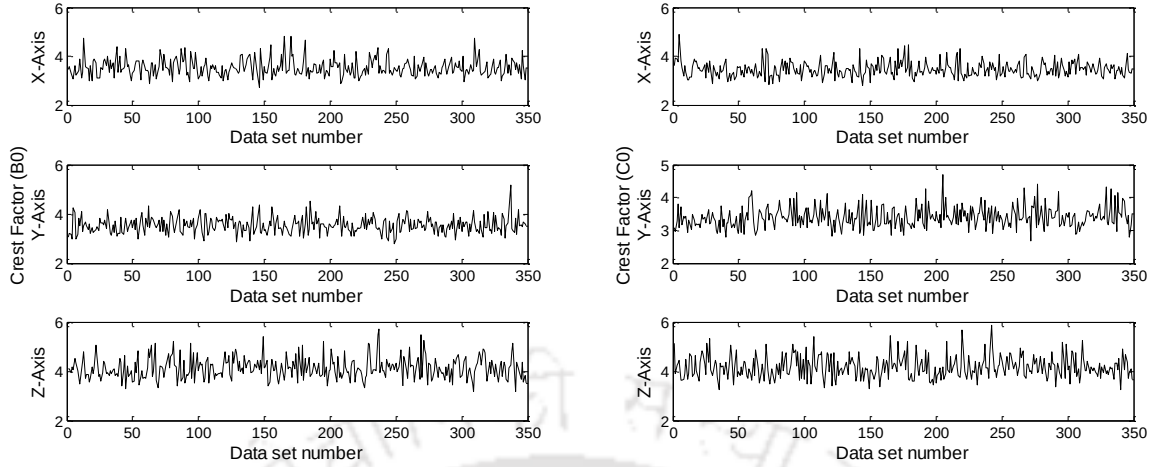
(b) Variation of standard deviation with dataset for (left) HP and (right) C0



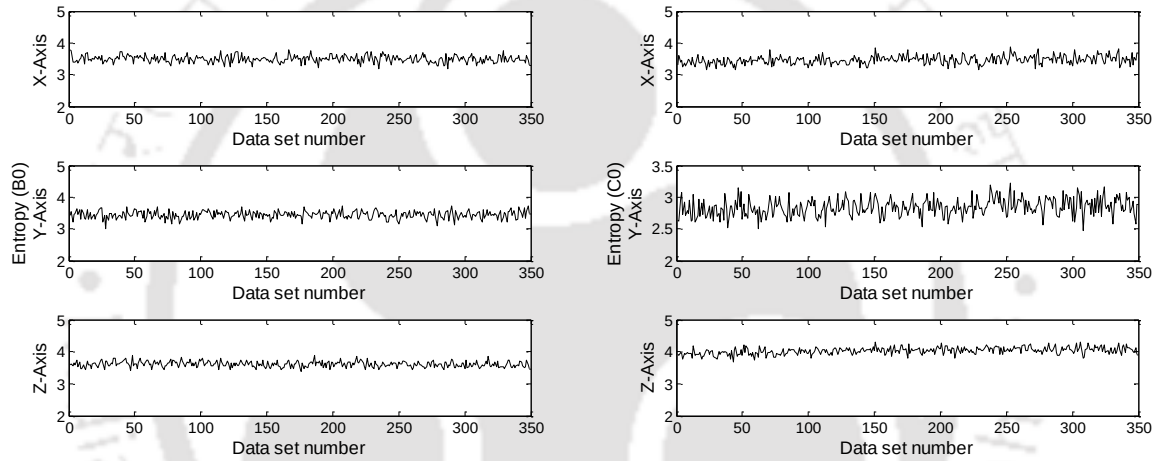
(c) Variation of skewness with dataset for (left) HP and (right) C0



(d) Variation of kurtosis with dataset for (left) HP and (right) C0



(e) Variation of crest factor with dataset for (left) HP and (right) C0



(f) Variation of entropy with dataset for (left) HP and (right) C0

Figure 4.1: Time domain features with dataset number for (a) mean (b) standard deviation (c) skewness (d) kurtosis (e) crest factor and (f) entropy

From Figure 4.1 it can be inferred that the range in which some of the statistical features vary with the fault condition of the CP is significant (for example standard deviation). Therefore these features may be useful in identifying CP faults when fed to the machine learning algorithm. After extracting the features, the best one or their best combination needs to be systematically selected. Therefore, for selecting the best features, a binary (SB1 versus HP at 40 Hz) and a multi-class classification are performed (classification of SB1, SB2, SB3, SB4,

and HP at 40 Hz). Classification accuracy is defined to quantify the performance of the classifier. It is the ratio of the number of feature points accurately predicted by the classifier for a specific condition to the total number of feature points tested for that condition. For example, in the binary classification of HP and SB1, the classification accuracy is given as,

$$\text{Classification accuracy of HP} = \frac{\text{No. of accurately predicted test feature points of HP}}{\text{Total no. of test feature points of HP}} \times 100 \quad (4.8)$$

$$\text{Classification accuracy of SB1} = \frac{\text{No. of accurately predicted test feature points of SB1}}{\text{Total no. of test feature points of SB1}} \times 100 \quad (4.9)$$

From the analyses tabulated in Table 4.1, it is observed that standard deviation feature outperforms all the other extracted features for multi-class classifications. Therefore, hereafter standard deviation would be utilized for all the subsequent classification analyses in this chapter. Initially, the binary fault classification is performed; where each blockage state would be compared with the HP condition. Later on, a multi-class classification involving all the blockage severities at a time is considered.

Table 4.1: Multiclass fault classification at 40 Hz for feature finalization

Feature	Classification accuracy					Average	γ	C
	HP	SB1	SB2	SB3	SB4			
Mean	50	50	50	50	100	60	0.5	100
Standard deviation	47.14	58.57	65.71	82.86	100	70.85	0.2	1000
Skewness	18.57	34.29	8.57	31.43	35.71	25.71	50	1
Kurtosis	31.43	22.86	14.29	14.29	55.71	27.71	0.05	1
Crest factor	34.29	10.00	30.00	5.71	51.43	26.28	1	1
Entropy	27.14	34.29	35.71	50	90	47.42	0.1	1000
Standard deviation and entropy	42.86	52.86	70.00	75.71	100.00	68.28	0.005	10000

4.2.2 Binary Classifications of Flow Blockages

In this section binary classification of suction blockage faults is attempted, where each blockage condition is compared to that of the healthy CP condition. These analyses would help to understand the complexity in identifying each blockage state and also establish if there is any operational speed dependency of the algorithm's classification performance. The results of the binary classification at different speeds with optimized C and γ parameters are shown in Table 4.2. From the table, it can be observed that SB1 (low severity suction blockage fault) shows the least classification performance, implying that the SB1 and HP fault data are not entirely separable this may be attributed to the decidedly less variation in vibration characteristics of the HP and SB1 conditions. Another observation that can be made from the results is that the classification accuracy of the algorithm improved with the increase in the suction restriction (or increase in the severity of suction blockage). It was experimentally observed that there was a perceivable flow anomaly between the SB2 and SB3 suction blockage levels (ref. Section 2.5). Therefore, it can be concluded that the CP is deviating from its healthy behavior rapidly with blockage beyond SB2 level. At higher speeds (> 50 Hz), SB2, SB3 and SB4 blockages show 100% classification accuracy. Also, SB4 blockage shows complete classification at all the CP speeds. Thus, the developed algorithm indicates the severity of the blockage and agrees with the experimental observations. Furthermore, it is observed that the classification accuracy for this case does not show any explicit operational speed dependency.

It is not always possible to estimate the amount of obstruction in a pipe and therefore training an algorithm for a binary classification may not be sufficient in all cases. Hence, a multiclass classifier is attempted to be built and tested in the next subsection.

4.2.3 Multiclass Classification of Different Levels of Blockages

In this section, a multiclass fault classification of different levels of suction blockages is attempted. The faults considered are HP, SB1, SB2, SB3, and SB4. Table 4.3 shows the fault prediction accuracy using the multiclass SVM classifier with the optimized values of C and γ . It can be observed from the table that the best average classification accuracy is obtained at 65 Hz speed. However, just like in the binary fault classification case, here also the accuracy of classification does not show any clear operational speed dependency. Similar to binary fault classification, even here SB1 fault showed least classification performance.

The SB4 level is predicted with 100% accuracy, SB3 with above 80% and after 55 Hz the SB2 is predicted with above 90% accuracy. The manifestation of CP faults depends very intricately on the operating speed, because as the speed changes, the flow inside the CP also modifies. It was observed during experimentation that above 50 Hz the bubble formation increases rapidly. Thus changing the vibration patterns and also the statistical features of the same. The more the deviation in the statistical patterns, the better the performance of the classifier. In the next subsection, an effort is made to predict the starting of cavitation (C0), which is very important for the pump safety.

4.2.4 Binary Classification of Impending Cavitation (C0)

Identification of the inception of cavitation in a CP system is crucial and has been a topic of extensive research. Many techniques involving usage of acoustic emission, noise spectrum, CP pressure data, and vibration signals have been studied to deal with the complex problem of cavitation identification.

Table 4.2: Percentage prediction accuracy in the binary classification of data

Speed						% classification accuracy														
(Hz)	SB1	HP	Avg.	C	γ	SB2	HP	Avg	C	Γ	SB3	HP	Avg	C	γ	SB4	HP	Avg	C	γ
40	68.6	54.3	61.4	15	5	81.4	87.1	84.2	10000	3	97.1	100	98.5	10	50	100	100	100	1	0.05
45	98.6	82.9	90.7	100	3	90	80	85	10	50	91.4	90	90.7	1	0.5	100	100	100	1	3
50	87.1	74.2	80.7	600	1	85.7	94.2	90	10000	0.1	100	100	100	1	3	100	100	100	1	0.005
55	98.5	74.2	86.4	10000	5	100	97.1	98.5	1	0.5	100	100	100	1	1	100	100	100	1	0.0005
60	90	68.6	79.2	10000	0.4	100	100	100	1	0.5	100	100	100	1	0.005	100	100	100	1	0.0005
65	90	97.1	93.5	1	0.2	100	100	100	1	0.05	100	100	100	1	0.005	100	100	100	1	0.0005

Table 4.3: Percentage accuracy in multiclass classification

Training and testing speeds (Hz)	% classification accuracy						γ	C
	HP	SB1	SB2	SB3	SB4	Average		
40	47.1	58.6	65.7	82.9	100	70.9	0.2	1000
45	70	85.7	75.7	85.7	100	83.4	2	100
50	74.3	71.4	68.6	100	100	82.9	1	100
55	68.6	100	98.6	97.1	100	92.9	50	10
60	67.1	90	100	100	100	91.4	4	1000
65	95.7	100	100	100	100	97.1	50	1

In the present study, during experimentation, it was observed that the bubble formation started between the SB2 and SB3 levels. If the bubbles are visible, then it means that the cavitation has already developed. This condition is termed as an impending cavitation state, C0. Also, with an increase in the suction flow restriction, there is a visible increase in the vibration level of the system. Hence, vibration signatures may prove to be useful to identify the C0 state.

In this section, it is attempted to identify the C0 state when compared to the HP condition via a binary fault classification. It needs to be noted that C0 is different for different speed levels, i.e., for low speeds it is close to SB3, and for high speeds, it is close to SB2 blockage level. The SVM algorithm is trained using features of C0 and HP and then tested to find the classification accuracy. The results are shown in Table 4.4.

From the table, it is evident that irrespective of the CP operating speed, the classifier identifies the C0 and the HP states with a near 100% accuracy. This classification is significant as it can accurately predict the start of cavitation. In industrial processes seeking the online condition

monitoring, this classification can help to curb the progression of cavitation right at its inception.

Table 4.4: Percentage prediction accuracy in the binary classification HP and C0

Training and testing speeds (Hz)	%classification accuracy			C	γ
	C0	HP	Average		
40	100	100	100	50	10
45	95.7	98.5	97.1	10	0.5
50	100	98.5	99.2	1	0.05
55	100	100	100	1	0.05
60	100	100	100	1	0.05
65	100	100	100	1	0.005

4.2.5 Conclusions: Phase I Time Domain Results

From the results presented in Tables 4.2 to 4.4, it can be observed that the developed classifiers have a limitation in identifying faults with low severity, especially at lower CP speeds. Nevertheless, the cavitation inception can be identified with almost 100% accuracy. However, it is essential to accurately detect the low suction restrictions with satisfactory performance to ensure that the fault progression is curbed.

In the subsequent section, the identification of suction blockage faults using frequency-domain based features is studied in order to establish the usefulness and limitations of these features.

4.3 Phase I Experimentation – Frequency Domain Analysis

In this section, the preliminary fault classification analysis using features derived from frequency spectra of the faulty and non-faulty CP conditions is used for the diagnosis.

4.3.1 Fault Feature Extraction and Selection

Features, such as the entropy, standard deviation, kurtosis, the product of amplitude and corresponding frequency, and root mean square have been considered in the frequency domain. The results for classification accuracy obtained for various features at 40 Hz operating speed is tabulated in Table 4.5. This classification accuracy of each of the features is used as a measure to select the best amongst them. Entropy is found to give the best performance for both binary and multiclass fault classifications, and therefore it would be used for all the further classifications.

Table 4.5: Percentage prediction accuracy of various features for binary and multiclass classifications at 40 Hz

Feature	Fault condition, % classification accuracy						
	C0	SB4	SB3	SB2	SB1	C0, SB4, SB3	C0, SB4, SB3, SB2
Entropy	98.6	100	99.3	82.9	60	95	88
Standard deviation	100	99.7	98.6	90	60.7	91.8	87.7
Product of amplitude and corresponding frequency	100	100	100	89.1	57.2	92.4	84.6
RMS	92.9	100	91.4	69.3	54.3	80.4	68

4.3.2 Binary Classifications of Flow Blockages

The binary classification is very crucial to determine the specific characteristics of a fault in comparison to the no-fault condition. Figure 4.2 (a to d) shows the classification accuracy of a specific fault, no-fault for all the experimented speeds. The following observations may be drawn from the results:

- For SB1 condition, the average classification accuracy over all the speeds has been found to be 73.8%.

- ii. The low classification accuracy may be attributed to the close behavior of the fault condition to that of no-fault.
- iii. For SB2 condition, the average classification accuracy is found to be 94.3% over the speed range. The classification accuracy just like in the previous case improved gradually with the speed. This may indicate that at higher speeds the effect of the blockage may be prominent.
- iv. For SB3 condition, the overall classification accuracy of 99.3% was achieved.
- v. For SB4 condition also, a high overall classification accuracy of 99.9% was obtained. This may indicate that there is a clear distinction of the fault from the no fault condition.
- vi. For cavitation inception condition C0, the overall classification accuracy was found to be 99.2%, which is a pretty good estimation of the beginning of cavitation.

Thus, SB3 and SB4 seem to be more serious conditions than SB1 and SB2. Also, it can be observed from the results that the classification accuracy is improving as the blockage increases.

4.3.3 Multiclass Classifications

Though the binary classification gives an overall picture of the fault severity, the multiclass classification has its share of advantages. It gives an understanding of the discreteness of each fault to one another. The following observations may be drawn from the results of the multiclass classification presented in Figure 4.3. The classification accuracies of HP and SB1 are almost comparable.

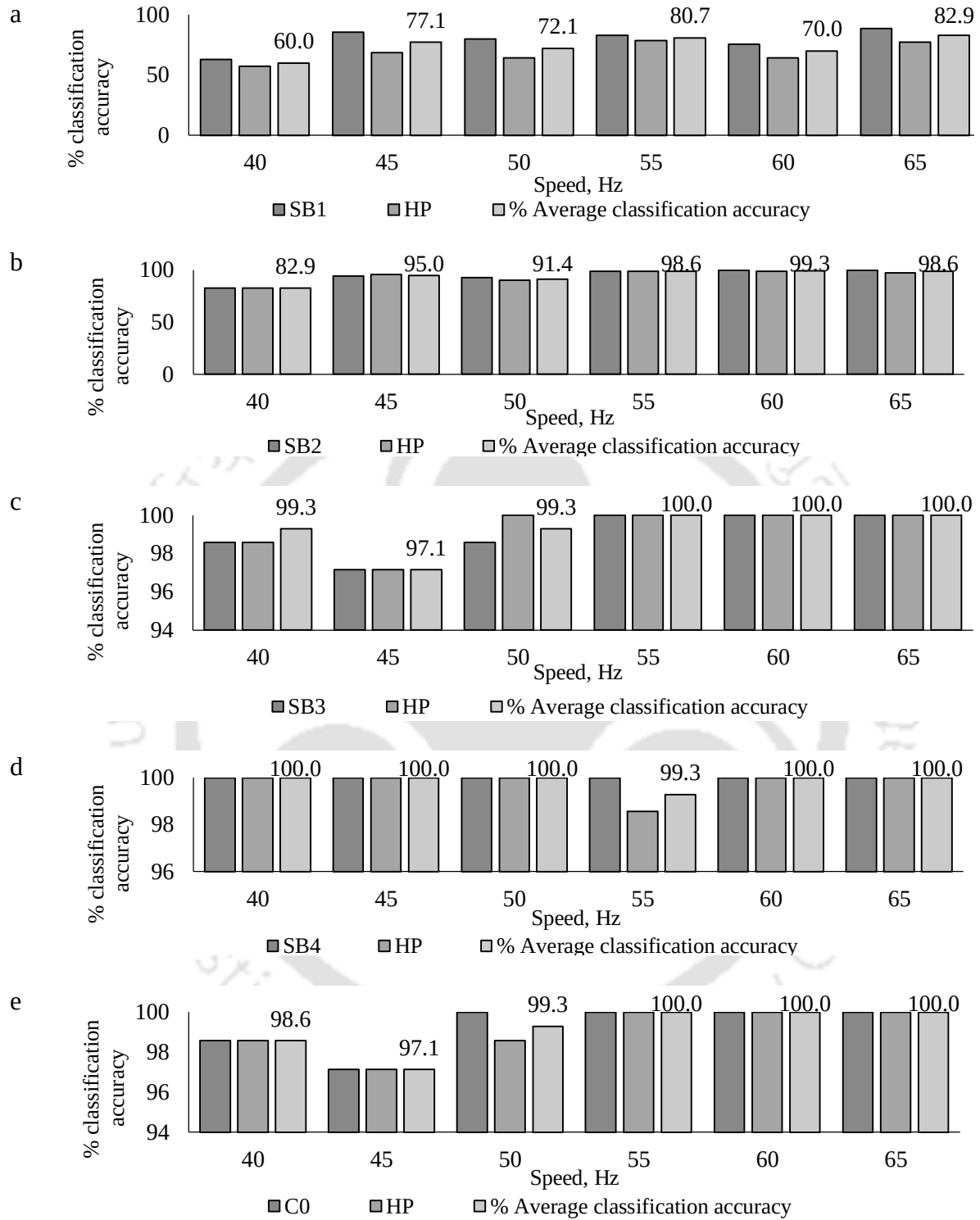


Figure 4.2: Binary classification of faults with respect to no fault condition, speed (Hz) versus percentage classification accuracy of, (a) SB1 and HP, (b) SB2 and HP, (c) SB3 and HP, (d) SB4 and HP and (e) C0 and HP

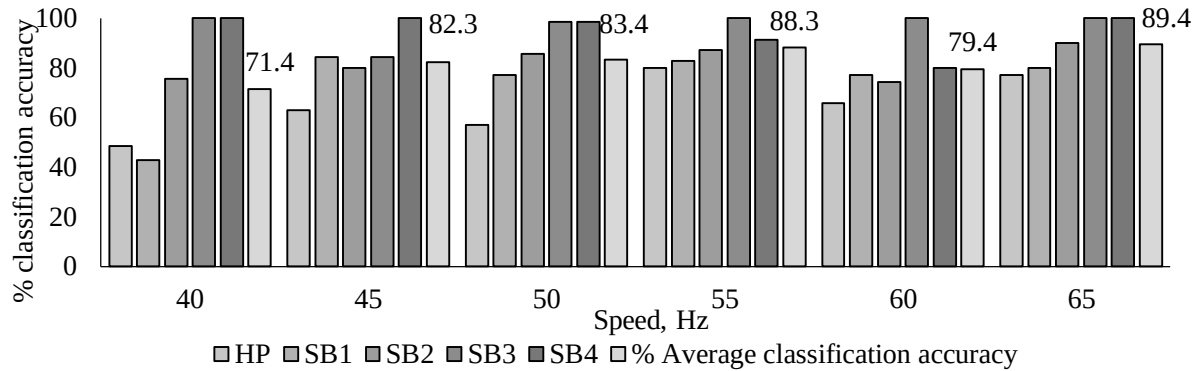


Figure 4.3: Multi-class classification of blockage: percentage of classification accuracy versus speed (Hz)

From the binary classification, we may infer that this may be due to the similarity of HP and SB1 vibration signatures.

- i. The classification accuracy of SB2 is low at 40 Hz and 60 Hz of pump speeds. Simultaneously, the classification accuracy of SB3 almost remained 100% throughout all the speeds.
- ii. The classification accuracy of SB4 remained 100% at all the speeds, except for 55 Hz and 60 Hz of pump speeds.
- iii. There is a sudden dip in the classification accuracy at 60 Hz pump speed. Hence, the classification of the faults seems to be independent of the operating speed of the pump.

4.3.4 Conclusions: Phase I Frequency Domain Results

Here also like in the time domain study, it can be observed that the classification accuracy is dependent on the amount of suction blockage and the operating CP speed. These developed methodologies need to be worked on further to give better predictability of the classifier. The

suction blockage faults and cavitation were addressed until now. However, no mechanical CP faults were considered. In the subsequent section, the Phase II experimentation analysis and results are described. Here, both the blockages and the impeller defects (mechanical faults) and their combinations are considered. Also, the phase II experiments, CP is run from 30 Hz to 65 Hz in steps of 5 Hz.

4.4 Phase-II Experimentation – Time Domain Analysis

In Phase-I experimentation, only the suction blockage faults of varying severity were considered. In the present experimentation, suction blockages, impeller defects and the combination of suction blockages and impeller defects are considered. From the previous set of experiments, it was observed that the classification performance of the algorithm depends on the type of the fault and the operating CP speed. Hence, in this study, it is attempted to delve deeper into understanding the characteristics of various features. It must also be noted that in this particular study to avoid the complexity of parameter selection, only the γ parameter is optimized, and the value of C parameter is taken as a constant. Also, extensive ranges of fault classification cases are studied. Therefore, the primary focus of this study is to identify the classification behavior using various feature(s) in normalized and non-normalized conditions. The cases of classification are considered in this study are presented in Table 4.6.

4.4.1 Phase II Feature Selection

A right choice of feature(s) helps to capture and interpret the physical behavior of the system with changing faults and aids in the better performance of the SVM classifier. In the present study, the feature performance has been tested on Case D with SB1 blockage, and the speed versus overall classification accuracy graph has been plotted as shown in Figure 4.4. The

reasons for choosing this case are, (1) IFSB1 is decidedly less severe and has a close resemblance to IF at low speeds, (2) in IFSB1, the bubble formation commences after 50 Hz speed, hence both steady and transient flow conditions can be perceived. In Figure 4.4, the left graph shows overall classification accuracies obtained using some of the prominent extracted features, like the mean (μ), peak-to-peak (p2p), standard deviation (σ), skewness (χ), kurtosis (κ) and entropy (S), at different speeds. In Figure 4.4, the right graph shows that, before cavitation commences, the μ consistently gives a classification accuracy of above 94%. However, with the commencement and progression of flow instability, the classification accuracy consistently drops. On the other hand, σ and S give good classification performance in the cavitation prone speed range. Thus, these three features are finalized for the fault classifications.

Table 4.6: Fault classification cases and their description

Case	Between	Description of type
A	HP, IF	*, Classified with respect to healthy CP
B	HP, IFSB	**, Classified with respect to healthy CP
C	SB, IFSB	*, Classified with respect to CP with blockage fault
D	IF, IFSB	**, Classified with respect to *
E	HP, SB, IF, IFSB	Healthy CP with different blockage levels and CP with co-existing blockage and impeller faults, *, **
F	IF, IFSB1, IFSB2, IFSB3, IFSB4	*, **, Classification taking all the faults at a time

* IF – Impeller defects coexist with flow instabilities

** IFSB – It has co-existing blockage and impeller faults along with flow instabilities due to both of them

Also, to reduce the variations in these features, a normalization technique has been employed. The generated feature data is divided by its respective absolute maximum. Hence, the range of data reduces to -1 and 1. $N(a)$ stands for the normalized form of the feature a . To present the results in a concise form a new term called average classification accuracy over entire speed range is defined. For this, first the overall classification accuracy of various faults is calculated for each blockage level at a given speed. Then the *overall classification accuracy for entire speed range* is calculated as the mean of the overall classification accuracies obtained at each speed. Mathematically, it is represented as,

$$O_j = \frac{\sum_{k=1}^p C_{j_k}}{p} \quad (4.10)$$

Where, j is the feature(s) considered, k is the speed, and p is the total number of speeds. O_j is the overall classification accuracy for entire speed range for a particular blockage level and j^{th} feature. $(C_j)_k$ is the average classification accuracy of all faults considered at a specific blockage level, using the j^{th} feature at k^{th} speed. Table 4.7, shows a sample calculation of $O_{(\mu-\sigma-S)}$ for Case F at SB2 blockage level.

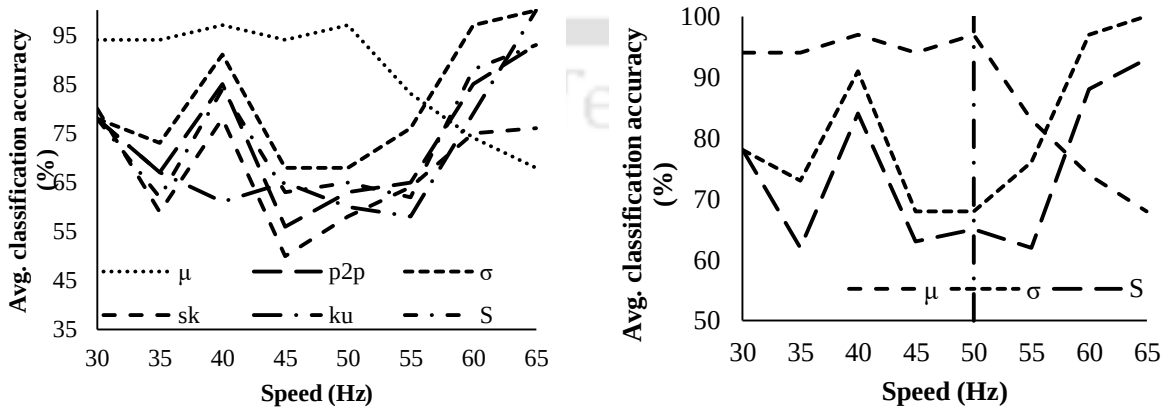


Figure 4.4: Speed versus classification accuracy of different features: Case D with B1

4.4.2 Binary Classification - Case A, Case B and Case C

A CP may develop impeller cracks due to manufacturing defects, handling issues, corrosion pitting, erosion, foreign particles inducing damage and flow instabilities (low-pressure pulsations). All the three cases considered in this section are multiple coexisting fault classifications. Case A classification is done between a healthy CP and a CP with impeller faults accompanied by flow instabilities. It is highly probable that the blockage and impeller defects may co-exist. Case B attempts to distinguish these co-existing blockage and impeller defects with a healthy CP condition. On the other hand, Case C classification attempts to distinguish these co-existing defects from a CP with blockage condition alone. In all the above three fault cases, all the finalized features and their combinations could classify the faults with 100% classification accuracy at all speeds.

Table 4.7: Calculation of $O_{(\mu-\sigma-S)}$ with Case F at SB2

Speed (Hz)	NB	OB	NBIF	BIF	$C_{(\mu-\sigma-S)}$
30	90	83	93	83	87
35	95	82	92	78	87
40	100	82	98	72	88
45	77	73	92	85	82
50	78	85	65	78	77
55	100	100	100	100	100
60	100	100	100	100	100
65	100	100	100	100	100
$O_{(\mu-\sigma-S)} =$					91

4.4.3 Prediction Performance - IF and IFSB - Case D

In this case, a CP with co-existing blockage and impeller cracks is classified with respect to a CP, which has impeller faults alone. This case will help to identify any additional fault development in a CP with impeller defects. In this classification, each blockage level is separately considered and classified. Of all the normalized and non-normalized feature(s) tried out, $(\mu-\sigma-S)$, $(\sigma-\mu)$ and σ performed better. Figure 4.5 shows the variation of $O_{(\mu-\sigma-S)}$, $O_{(\sigma-\mu)}$ and O_{σ} with the blockage level. From the graph, it can be observed that the prediction performance is improving with the blockage level. Here again, it can be attributed to the increase in the flow instability as the blockage increases. As the blockage approaches the SB3 or SB4 level, the accuracy is increased to more than 95%, indicating their deviation from IF signature and thus establishing the severity of these co-existing faults.

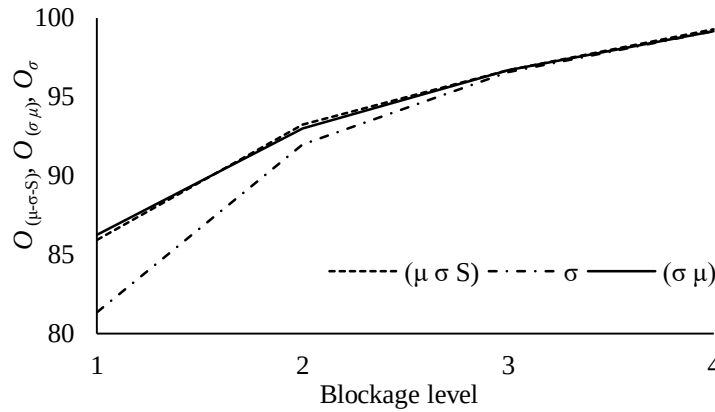


Figure 4.5: Blockage level versus average classification accuracy over entire speed range
Case D

4.4.4 Prediction Performance - HP, SB, IF and IFSB – Case E

Table 4. shows O_j for different features at different blockage levels. From the table, it can be vividly seen that feature(s) $(\mu-\sigma-S)$ and σ almost give similar classification accuracies.

Classification accuracies of non-normalized features are better than the normalized ones. In Table 4., SB1 represents blockage induced in the SB and IFSB cases, i.e. SB1 and IFSB1, similarly for all the other blockage levels as well.

Figure 4.6 shows the classification accuracy for each fault, using $(\mu-\sigma-S)$ features. The classification performance of the algorithm improves with the blockage severity, and for a particular blockage level, it improves with speed. When SB3 is induced, all the faults gave a classification accuracy of more than 90%, and with SB4 they gave almost 100% classification accuracy. With the increase in the severity of blockage, the fault signatures deviate from the healthy CP signatures. In other words, the classification accuracy is lower with less severe blockages (SB1 and SB2) and at low speeds.

Table 4.8: O_j of different features: Case F

	$O_{(\mu-\sigma-S)}$	$O_{N(\mu-\sigma-S)}$	O_{μ}	$O_{N(\mu)}$	O_{σ}	$O_{N(\sigma)}$	O_S	$O_{N(S)}$
SB1	80	81	69	69	78	73	67	64
SB2	90	85	68	69	90	70	79	70
SB3	96	85	61	67	96	75	88	70
SB4	99	87	72	69	99	75	96	71

4.4.5 Multiclass Classification IF, IFSB1, IFSB2, IFSB3 and IFSB4 – Case F

In this case, unlike Case D, all the faults are considered at a time. It, therefore, helps to identify the severity of the co-existing blockage and impeller faults. This fault classification is attempted in the following ways, (i) the training and testing is done with all the faults simultaneously, and (ii) the training is done with all the faults and testing are done only with

one fault at a time. Figure 4.7 shows the speed versus classification accuracy of each fault using $(\mu-\sigma-S)$ feature. It can be observed that at all speeds, IF has $O_{(\mu-\sigma-S)}$ of 100%. As the blockage level increases from IFSB1 to IFSB4, the $O_{(\mu-\sigma-S)}$ steadily improved.

In a scenario where the operator knows that there is a blockage fault, and is not sure of the severity of it, the second method of classification can be used. In this, training of the algorithm is done using all the fault conditions simultaneously, and testing of it is done using only one fault at a time. In the SVM classification, each fault is assigned a label. In the present case, each level of blockage has been attached with a distinct label, i.e. IF is assigned Label I, IFSB1 - Label II, IFSB2 - Label III, IFSB3 - Label IV, and IFSB4 - Label V.

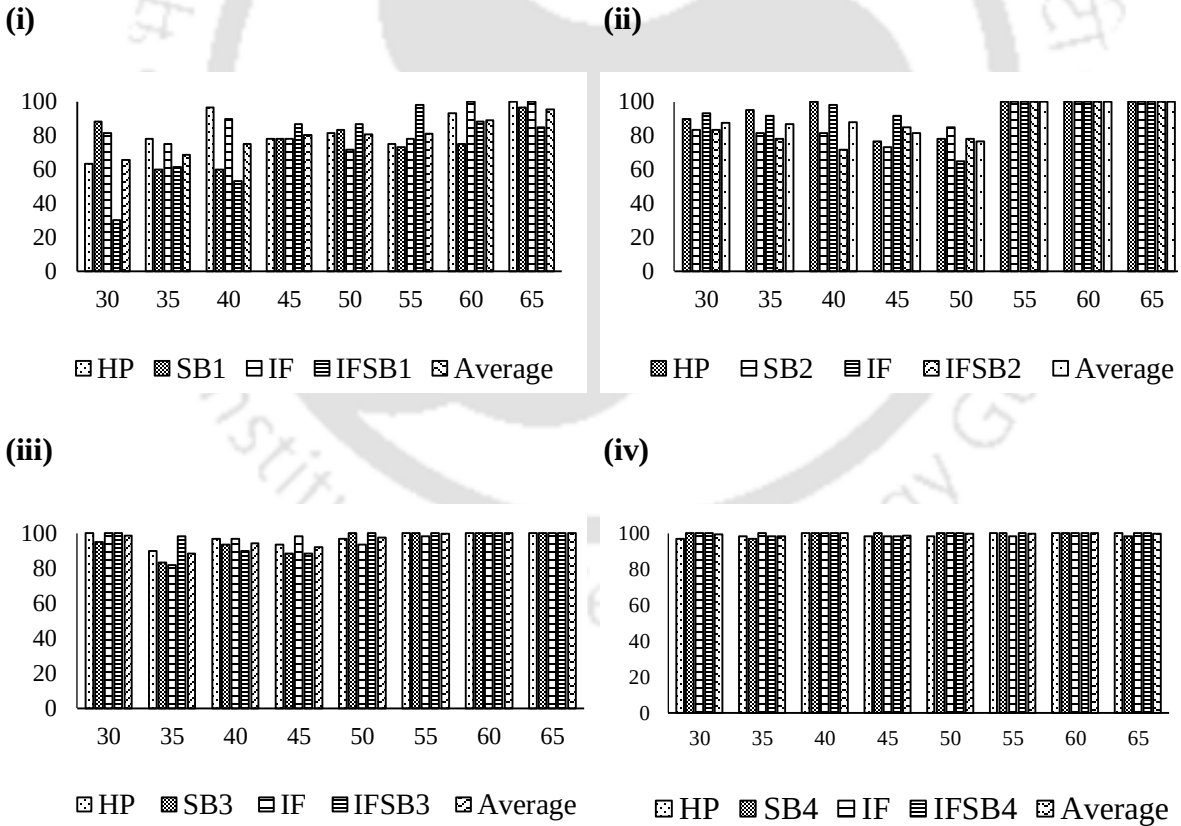


Figure 4.6: Classification accuracy of each fault at different speeds for (i) SB1, (ii) SB2, (iii) SB3, (iv) SB4; with $(\mu-\sigma-S)$ feature

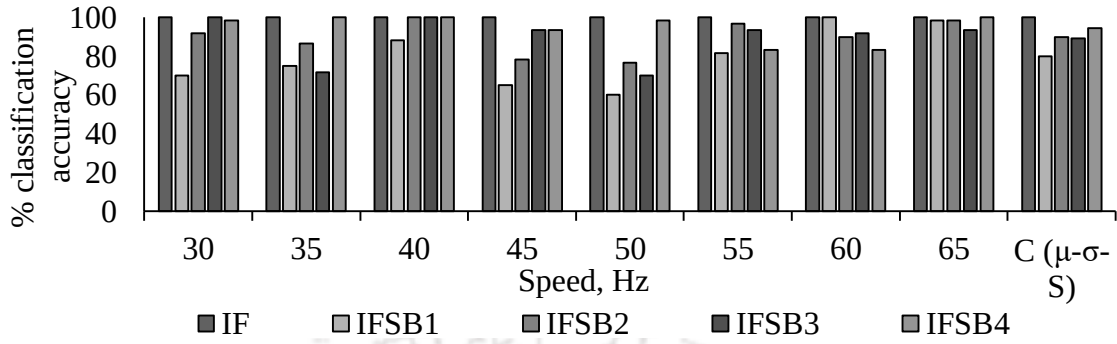


Figure 4.7: Classification accuracy by testing with one fault at a time for different speeds with $(\mu-\sigma-S)$ feature Case F.

The signal whose blockage level needs to be determined is assigned one fault label at a time and is tested on the trained algorithm. From the preliminary results, it is observed that every fault, irrespective of the label assigned, classifies as IF fault after a specific value of γ ($= 50$) due to over-fitting of data. Hence, the final range of γ has been chosen as 5×10^{-3} to 5. Further, if the data is over-fit, then most of the misclassified data points form support vectors themselves (a separate cluster). Hence, any amount of increase in C would not change the classification accuracy. However, with the finalized range, it has been observed that there is a change in classification accuracy with C , justifying the selected γ range. With the results of the second investigation, a confusion matrix is developed. Each column of the matrix represents the instances of different predicted labels, and each row represents the actual fault (label).

Note that in the confusion matrix presented here, each entry is a percentage of classification accuracy at the optimized γ value as shown in Tables 4.9 to 4.12. The sum of classification accuracies in any row would therefore always be 100. From the table, it can be evidently seen that the fault has been accurately classified into its correct fault category.

Table 4.9: Confusion matrix with a prescribed γ value, for 30 Hz CP speed

Speed (Hz)	Test with	Label I	Label II	Label III	Label IV	Label V	γ
30	IF	78.33	11.67	6.67	0	3.33	0.2
	IFSB1	18.33	70	1.67	1.67	8.33	0.005
	IFSB2	3.33	3.33	91.67	0	1.67	1
	IFSB3	0	0	0	100	0	1
	IFSB4	0	0	1.67	0	98.33	0.005

Table 4.10: Confusion matrix with a prescribed γ value, for 35 Hz and 40 Hz

Speed (Hz)	Test with	Label I	Label II	Label III	Label IV	Label V	γ
35	IF	68.33	20	3.34	8.33	0	0.4
	IFSB1	16.67	75	3.33	5	0	0.4
	IFSB2	5	3.33	86.67	5	0	5
	IFSB3	18.33	8.33	3.34	70	0	0.1
	IFSB4	0	0	0	0	100	all
40	IF	93.33	3.33	0	3.34	0	1
	IFSB1	8.33	88.33	0	3.34	0	0.3
	IFSB2	0	0	100	0	0	5
	IFSB3	1.67	3.33	0	95	0	0.1
	IFSB4	0	0	0	0	100	0.5

4.4.6 Conclusions: Phase-II Time Domain Results

Based on the experiments performed in phase II and the results of their analyses, following essential deductions can be made,

1. By considering the combination of hydraulic and mechanical CP faults, the complexity of fault classification increased.
2. Here also, just like from phase I results, the classification accuracy improved with the increase in SB severity level.

3. Though classification accuracy does not show any explicit operational speed dependency, there is a definite increase in the classification performance at higher speeds which may be attributed to the increase in fault severity.
4. The non-normalized features always outperformed the normalized ones. Implying that there is no necessity to normalize the feature values.
5. The developed confusion matrix helps to visualize the performance of the classification algorithm, and the faults could be accurately classified into their respective fault label. However, if more faults are considered, then it is difficult and tedious to make such a table.

Table 4.11: Confusion matrix with a prescribed γ value, from 45 Hz to 55 Hz

Speed (Hz)	Test with	Label I	Label II	Label III	Label IV	Label V	γ
45	IF	61.66	16.67	18.33	1.67	1.67	0.5
	IFSB1	20	65	11.67	3.33	0	0.4
	IFSB2	5	10	80	5	0	0.4
	IFSB3	0	0	3.33	93.33	3.34	0.1
	IFSB4	1.67	1.67	0	3.33	93.33	1
50	IF	66.66	15	15	1.67	1.67	0.4
	IFSB1	11.67	60	23.33	5	0	0.1
	IFSB2	13.33	23.33	60	3.34	0	0.3
	IFSB3	5	1.67	0	70	23.33	0.1
	IFSB4	1.67	0	0	8.33	90	5
55	IF	73.33	26.67	0	0	0	0.2
	IFSB1	18.33	81.67	0	0	0	0.3
	IFSB2	0	0	96.67	3.33	0	1
	IFSB3	0	0	1.67	93.33	5	0.1
	IFSB4	0	0	0	16.67	83.33	5

Table 4.1: Confusion matrix with a prescribed γ value, from 60 Hz to 65 Hz

Speed (Hz)	Test with	Label I	Label II	Label III	Label IV	Label V	γ
60	IF	95	5	0	0	0	1
	IFSB1	0	100	0	0	0	1
	IFSB2	0	0	90	1.67	8.33	0.3
	IFSB3	0	0	0	91.67	8.33	0.5
	IFSB4	0	0	11.67	5	83.33	0.4
65	IF	100	0	0	0	0	0.05
	IFSB1	0	98.33	1.67	0	0	0.2
	IFSB2	0	1.67	98.33	0	0	0.1
	IFSB3	0	0	0	93.33	6.67	0.1
	IFSB4	0	0	0	0	100	5

4.5 Summary

From all the classifications presented in this study (in both time and frequency domains), the following significant conclusions can be drawn:

1. The feature(s) selection is a very crucial step in the fault classification, and there seems to be some speed (or fault manifestation) dependency of the feature classification performance.
2. With the increase in the fault severity, the fault classification accuracy is improving, owing to the change in feature values.
3. The choice of optimal kernel parameter combination is vital in building a reliable classifier.
4. Though the performance of the developed classifier is satisfactory, there is significant scope for improvement of the same.
5. With the consideration of a combination of mechanical and hydraulic faults, the complexity of fault classification is increasing.

6. Binary classification and multiclass classification go hand-in-hand, and therefore attempting a multiclass classification is sufficient to analyze the adaptability of the developed classifier.
7. The flow irregularities developed in the CP create transient signals. Accurately capturing these signals is essential to identify fault characteristics better. One step ahead is by considering the effect of data sampling rate on the fault prediction performance.

Based on the derived conclusions, the objectives of the subsequent studies would be:

1. To choose the features more systematically, so that,
 - (i) The classifier shows good classification ability at all CP operating speeds. This is important because the developed methodologies should be independent of the operating CP speed.
 - (ii) There is an improvement in the fault classification performance
 - (iii) The classifier gives good prediction performance while dealing with faults of lower severity as well. This is crucial because the identification of CP faults at their formative stages can help curb its progression.
2. From the literature, it is observed that there is no thumb rule for the choice of the sampling rate of data acquisition. However, it is important to understand the effect of the same on the fault classification performance. Hence, in the subsequent studies two sampling rates of data acquisition are studied, and their performances and characteristics are compared.
3. Through the Phase II experiments, it was observed that the mechanical and hydraulic faults are inter-related, and it is important to consider such fault combinations.

Therefore, in the next phase of experiments, the suction and discharge blockages of varying severities, impeller faults, pitted cover faults, dry run faults and their combinations will be considered.

4. The present method (i.e., hit and trial), used for the selection of best kernel and SVM parameter combinations, is very tedious and cannot be used for automation of the algorithm. Hence, in the subsequent studies a more effective yet simple technique, called the *cross validation* technique, would be employed for the optimal parameter selection.
5. In this study 80% data is used for training the algorithm and 20% is used to test the algorithm. But using more data for training and less for testing may not be able to accurately describe the generalized classification performance of the classifier. Therefore, in the next studies, it is attempted to decrease the training data and still build a reliable classifier.

CHAPTER 5: Multi-Fault Diagnosis of Centrifugal Pumps in Time Domain using Data Acquired at Low Sampling Rate

5.1 Introduction

Time domain analysis methods deal with studying the amplitude and phase information of a CP system and thus determining its working condition. It gives a practical understanding of vibration patterns of the system and thus offers a physical feel of the extent of the damage. Mainly, the short-lived transients generated by the flow instabilities can be very well analyzed using time-domain studies.

In Chapter 2, it was mentioned that the experimentation was carried out in three phases. Chapter 4 describes the fault classification analyses results for the data acquired from various faults using Phase I (described in Section 2.5.1) and Phase II (described in Section 2.5.2) of the experiments. It is clear from the previous chapter that there is a need for selection of better features as the classifier's performance majorly depends upon the choice of the features. Therefore, a number of features, which are extracted from the time domain vibration and current data are considered in this work. These features are fed into the SVM classifier as inputs to perform the fault classification. Unlike the previous chapter, here the cross-validation technique along with the grid-search technique is used for the optimum selection of the SVM parameters. The wrapper model (as detailed in Chapter 3) is used for the selection of the best features. Thus, the multi-class SVM based CP fault diagnosis methodology is established on an assimilated approach using the wrapper model and the cross-validation in conjunction with the grid-search technique.

Also, to study the effect of the data sampling rate, a comparison of the classification performance of the methodology using two different data resolutions is made in this study. In addition to the work explained in the previous chapter, in this work, current signals have also been extracted for various faults. Therefore, the effect of using both the features extracted from the time domain of vibration signals and current signals is compared with the one obtained using features extracted only from vibration signals. To check the robustness of the method, it is tested at a wide range of distinct CP operating conditions. The CP fault diagnosis is presented when the algorithm is trained and tested at the same speed, trained at two distinct speeds and tested at an intermediate speed. The intermediate speed case is studied to check the capability of the algorithm in handling data unavailability at certain speeds. Adding to this, the versatility of the algorithm in handling corrupted fault signatures is also studied.

5.2 Fault Feature Extraction

The feature selection is very crucial in obtaining the best performance of a classifier. Therefore, a wide selection of feature trials is done for the best feature(s) selection, which is done based on the wrapper method. Some of the common features used by various researchers in the CP fault diagnosis are presented in (Table 1.8). In the current work, features like the mean, variance, standard deviation, skewness, kurtosis, 5th statistical moment, 6th statistical moment, 7th statistical moment, root mean square, root sum of squares, sum of squares, shape factor, impulsion index, crest factor, tolerance index, power to average ratio and kurtosis index are extracted from the raw data. Some of the features have already been described in Chapter 4. The definitions and explanations of the other features are given as follows:

k^{th} statistical moment (μ_k): The statistical moment provides an indication of the general feature distribution of the data variables, x . A number of statistical features which describes the behavior of the data are constructed by the first few moments. The k^{th} moment about the mean of discrete data can be calculated using the following equation,

$$\mu_k = \frac{1}{N} \sum_{i=1}^N x_i - \mu^k \quad (4.11)$$

Variance (μ_2): Variance, as the name suggests works by evaluating the mean of squares of the deviations of the data points from the mean value. Therefore, the more the deviation, the more the variance and positive/negative deviations are treated equivalent. This is a very lucrative statistical measure to calculate the spread of the data. This is also the second statistical moment of the data.

Sum of squares (SS) and root sum of squares (RSS): Sum of squares is defined as the addition of the square values of each data point in the data set. Another statistical feature RSS is given as the square root of the sum of squares of the data points of the signal. The definition is given as,

$$SS = \sum_{i=1}^N |x_i|^2 = RSS^2 \quad (4.12)$$

Shape factor (w): Shape factor is given as a ratio of the root mean square and the mean. It is given by the formula,

$$\omega = \frac{RMS}{\mu} \quad (4.13)$$

Impulsion index (I): The impulsion index is a non-dimensional quantity that is given by the ratio of the maximum value of a data set to the mean value of the data set. It can be given by,

$$I = \frac{\max |x_i|}{\mu} \quad (4.14)$$

Power to average ratio (PA): The power to average ratio is the square of crest factor (Eq. (4.5)). They can be related as,

$$P = \frac{\max |x_i|}{RMS} = \sqrt{PA} \quad (4.15)$$

Tolerance index (T): The tolerance index is given as the ratio of maximum value to the square of the mean square root values of the data points. This is also a non-dimensional value. It is given by,

$$T = \frac{\max |x_i|}{\left(\frac{1}{N} \sum_{i=1}^N |x_i|^2 \right)^{1/2}} \quad (4.16)$$

Kurtosis index (K): It is the ratio of kurtosis to fourth power standard deviation. The importance of this feature is explained in the subsequent sections. It is written as,

$$K = \frac{\kappa}{\sigma^4} \quad (4.17)$$

All the aforementioned seventeen statistical features are extracted from the time-domain data (both vibration and current data). The same features are extracted from the data collected with both the high sampling rate data and the low sampling rate data. The details of the data collection are given in Table 5.1.

5.3 Fault Feature Selection

The multi-fault diagnosis of fault feature data generated for nine faults are considered for the feature finalization. The faults considered are the HP, SB3, DB3, IF, IFSB3, IFDB3, PC, PCSB3, and PCDB3¹ (refer Table 2.5). Figure 5.1 shows the average classification performance (it is the average of individual fault classification accuracies of each fault) for each feature at 30 Hz speed.

The features μ , μ_2 , σ , RMS , RSS and K show promising classification abilities and are as shown in Figure 5.1. Also, μ and K show 100% classification. While the mathematical descriptions of all the other features are well defined, the kurtosis index is newly introduced in this paper. As seen in Figure 5.1, κ (kurtosis) is not a suitable feature for the fault classification. But on dividing κ by σ^4 (i.e., K), the classification performance drastically improved. This implies that the quantity in the denominator of K has a significant role to play.

Hence, different powers (-4, -3, -2, -1, 1, 2, 3, 4, 5) of σ are taken and overall classification performance obtained using each of them is compared. The comparison is shown in Figure 5.2. It can be vividly seen from the figure that σ^{-1} (shown in pattern fill in the graph), which is the reciprocal of standard deviation gives the best classification performance. When the values of σ and σ^{-1} are plotted for all the 150 datasets of a healthy pump at 30 Hz as shown in top and bottom diagrams of Figure 5.3, respectively; it can be observed that σ^{-1} is almost a mirror image of σ .

¹ HP – healthy pump, SBk – k^{th} suction blockage, DBk – k^{th} discharge blockage, IF – impeller fault, IFSBk – impeller fault along with k^{th} suction blockage, IFDBk – impeller fault along with k^{th} discharge blockage, PC – pitted cover fault, PCSBk – pitted cover fault along with k^{th} suction blockage, PCDBk – pitted cover fault along with k^{th} discharge blockage

Another interesting observation that can be made is that the variation in σ is reduced drastically implying that the newly introduced σ^{-1} feature preserves the intricacies of σ feature and also reduces its variation, thereby proving to be a very effective. Hence, from Figure 5.1 and Figure 5.3, μ and σ^{-1} are chosen as the suitable features for the fault classification.

5.3.1 Data Resolution Effect

As explained in Chapter 2 (Section 2.3.2), two sets of data are collected for the same faults at different sampling rates. The scatter plots using μ and σ^{-1} features for HP, SB3, DB3, IF, IFSB3, IFDB3, PC, PCSB3, and PCDB3 faults in the input space for the low and high data sampling rate are shown in Figure 5.4 and Figure 5.5, respectively. It can be seen that all the nine faults form distinct clusters in the input space when low sampling rate data is used (Figure 5.5), while it is observed that they overlap with each other when using the high sampling rate data (Figure 5.4).

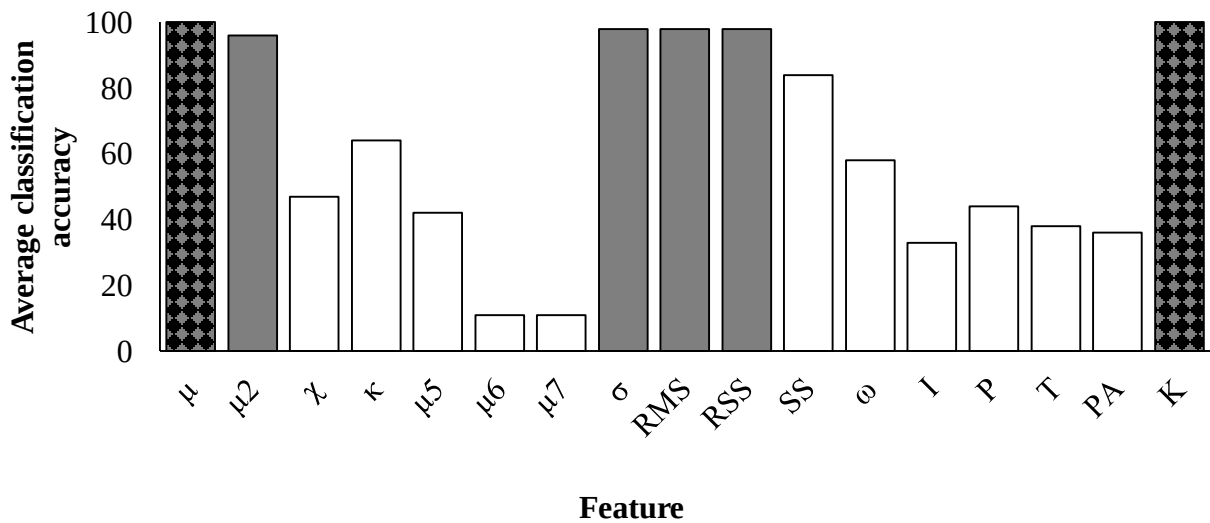


Figure 5.1: Feature classification performance

Table 5.1: Total data collected and features extracted for different data acquisition rates used

Sampling rate (kHz)	No. of samples/dataset in time domain (a)	No. of data sets (b)	No. of measurements using two tri-axial accelerometers (c)	No. of measurements using three line current probes (d)	No. of faults (e)	Total data collected per fault $f=\{a \times b \times (c+d)\}$	Total data collected for all faults $g=(f \times e)$	No. of features collected per dataset (h)	No. of features collected for each case $v=h \times b$
20	2000	350	6	3	33	63×10^5	2079×10^5	17	5950
5	5000	150	6	3	33	67.5×10^5	2227.5×10^5	17	2550

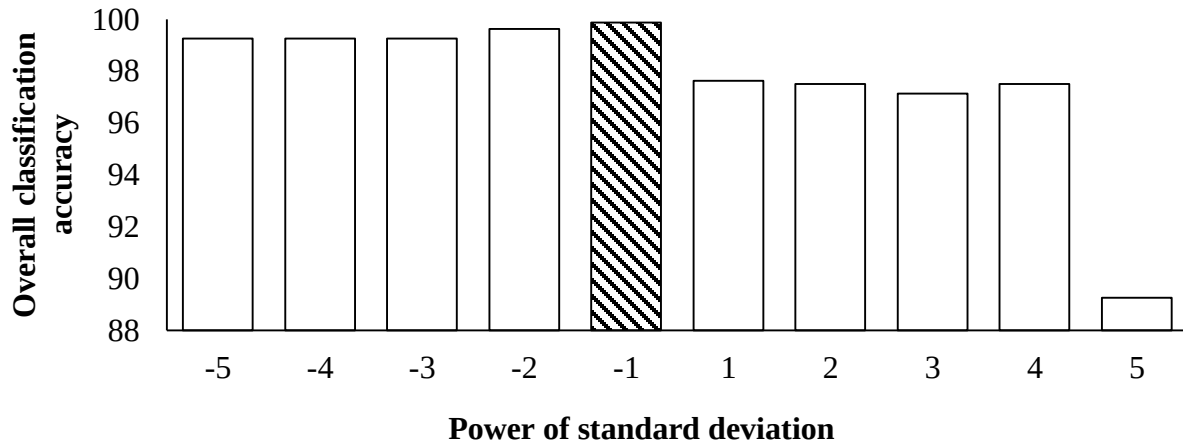


Figure 5.2: Feature classification performance of different powers of σ

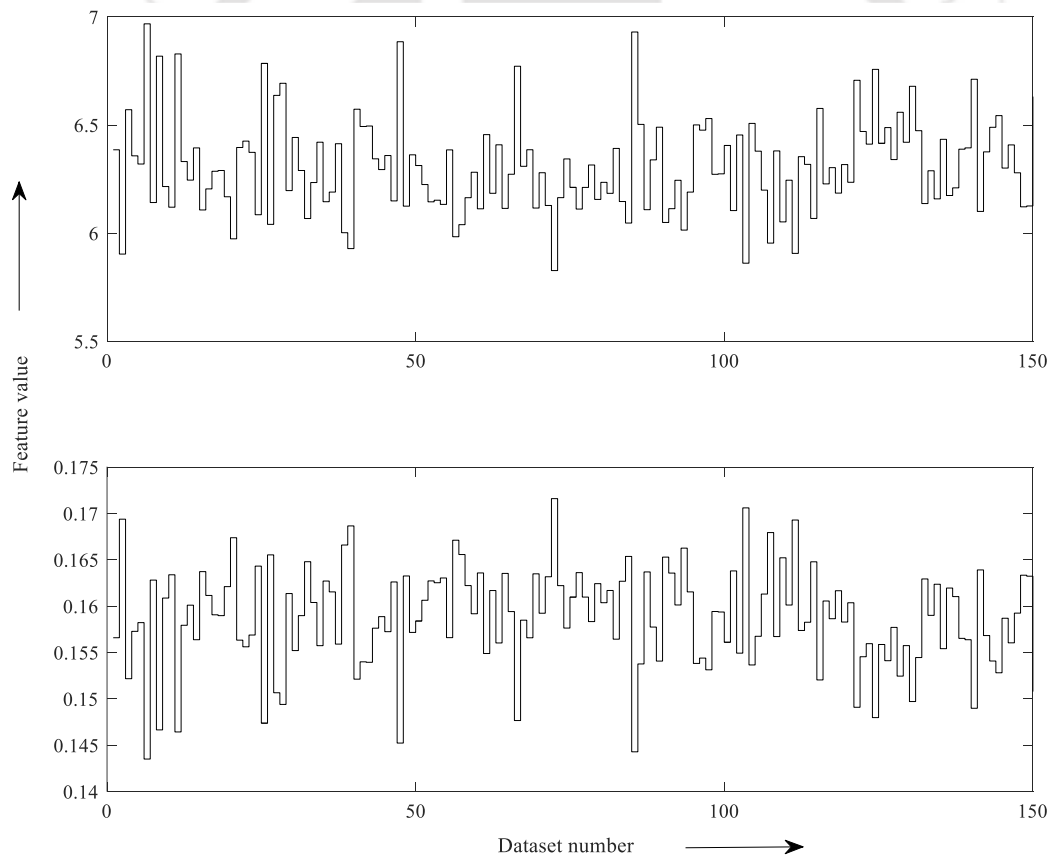


Figure 5.3: Feature values versus dataset number for σ (top) and σ^{-1} (bottom)

By using a kernel function, these data clusters can be made linearly classifiable. For the same aforementioned faults, an average classification accuracy of 89% is obtained at 30 Hz speed with high sampling rate data, while by using the low sampling rate data the average classification accuracy reached 100%. That is, there is an 11% improvement in the classification accuracy on increasing the data resolution (i.e., low sampling rate data). Owing to the data resolution effect, hereafter, the results of the classifications using low sampling rate data are only presented.

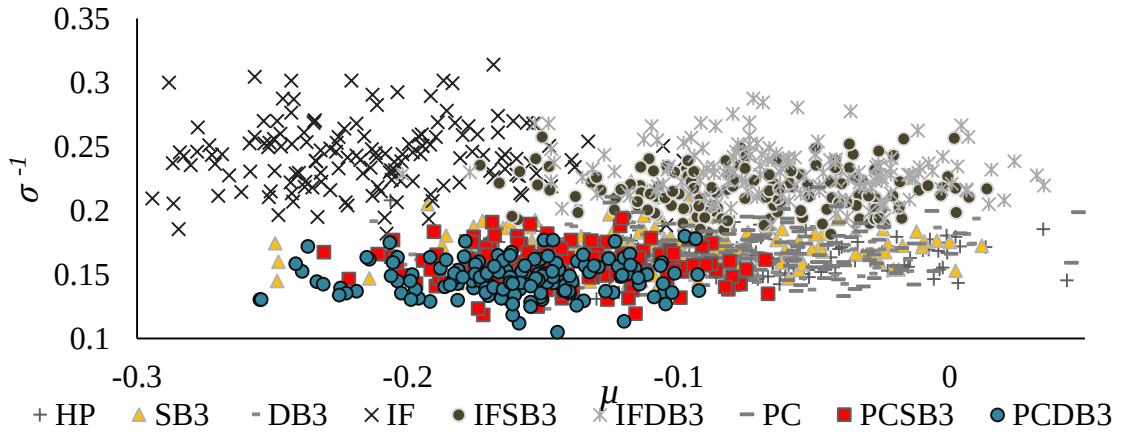


Figure 5.4: Faults scatter plot for the low resolution data

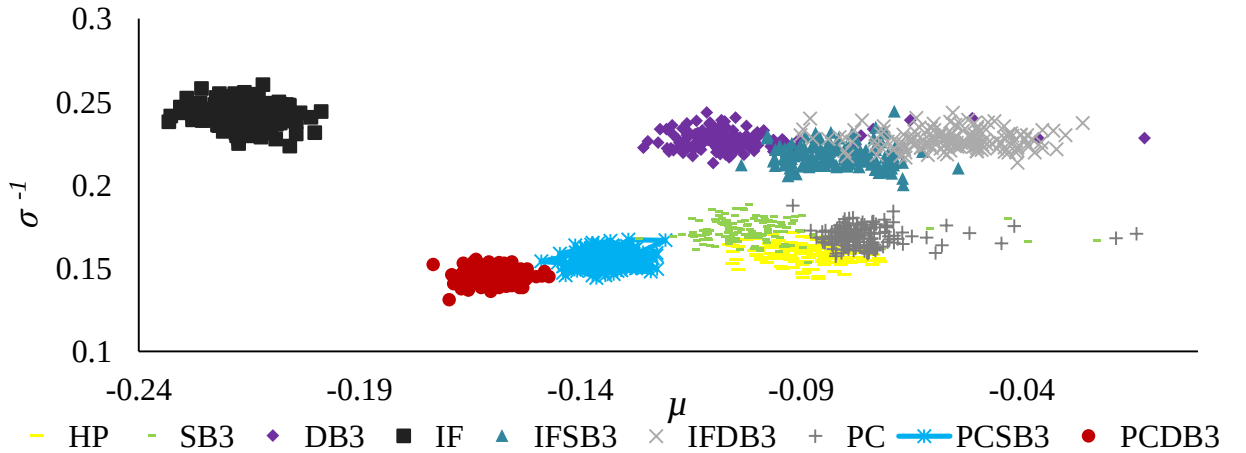


Figure 5.5: Fault scatter plot for the high resolution data

5.4 Fault Diagnosis Using the SVM

To facilitate fault classifications, the faults are broadly segregated into the following groups,

- (i) the CP with pure blockage faults (BF) constituting the SB1, SB2, SB3, SB4, DB1, DB2, DB3, DB4, and DB5 faults,
- (ii) the CP with both impeller defects and blockage (IFBF) faults constituting the IF, IFSB1, IFSB2, IFSB3, IFSB4, IFDB1, IFDB2, IFDB3, IFDB4, and IFDB5 faults,
- (iii) the CP with both the pitted cover faults and the blockage faults constituting (PCBF) the PC, PCSB1, PCSB2, PCSB3, PCSB4, PCDB1, PCDB2, PCDB3, PCDB4, and PCDB5 faults,
- (iv) the CP with dry runs constituting the SB5, IFSB5 and PCSB5 faults, and
- (v) all faults (AF) constituting the HP, SB2, SB5, DB1, DB5, IF, IFSB2, IFSB5, IFDB1, IFDB5, PC, PCSB2, PCSB5, PCDB1 and PCSB5 faults.

Here, it may be noted that the AF set of faults is chosen such that it considers individual faults, co-existing faults with intermediate severity and the dry-runs. The fault classification cases that are studied in the present work are given in Table 5.2.

It may be noted that the type of signal used for the fault classification also plays a major role in the classification performance obtained. When the motor is directly coupled to the CP, the current signature varies with any load variation on the CP. But in the present experiment, the CP is driven by a belt-drive and so the load variations on the CP may not correspondingly affect the motor current as lot of damping is involved. Nevertheless, the current signature compliments the performance of the vibration signals. Figure 5.6 shows the average and overall classification accuracies (it is the average of average classification accuracies at all

speeds) obtained using the vibration signature alone, and both vibration and current signatures for PCBF faults.

Table 5.2: Fault classification cases

Case	Faults for training	Faults for testing	Training speed (Hz)	Testing speed
A	BF, HP	BF, HP	30 – 65	Same as training
	IFBF, HP	IFBF, HP		
	PCBF, HP	PCBF, HP		
	Dry runs, HP	Dry runs, HP		
B	AF	AF	30 – 65	Intermediate speed
	AF	AF		
C	AF	AF	30 – 65	Noisy signal at same training speed

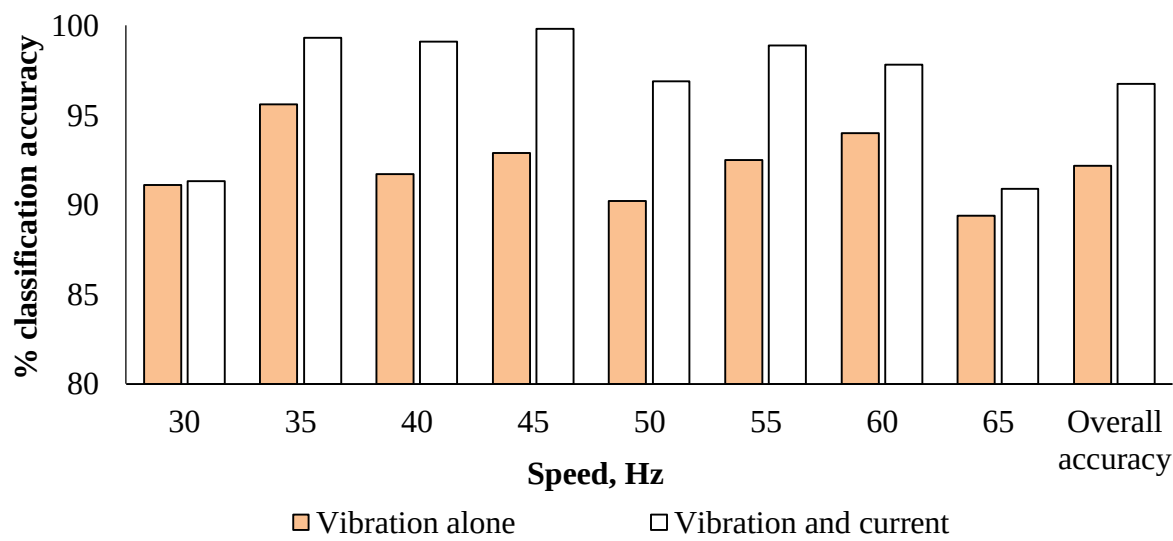


Figure 5.6: Performance of the SVM classifier using vibration data alone and with both vibration and current data for the PCBF

It is evident that there is an improvement in the classifiers performance with the usage of both the signals together. Henceforth, the combination of current and vibration data features will be used in the classifier.

5.4.1 Case A: Same Speed Classification

In Case A, five categories of faults are classified, namely, the blockage faults, impeller faults, pitted cover faults, dry run faults and a prescribed combination of all the above faults. It must be noted that the algorithm is trained and tested at the same operating speed of the CP in this case.

A classification algorithm is proved to be effective if it can learn even using very less samples of data. Hence, in the present study, 50% of the data is used for training and 50% for testing. Here, the classification accuracy of the algorithm is calculated for two cases, (i) training and testing of the algorithm with 70 non-overlapping data sets each, (ii) training and testing of the algorithm with 40 non-overlapping data sets each. Average classification accuracies obtained in both the cases for AF faults classification are compared in the bar graph as shown in Figure 5.7.

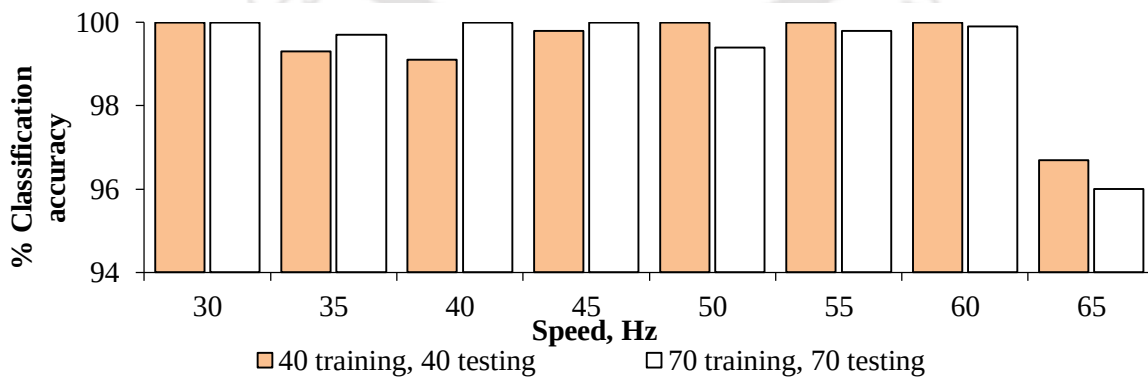


Figure 5.7: Effect of number of training/ testing datasets on the classifiers performance for AF

It can be lucidly observed that in both the cases, the results obtained are comparable. Also, the time taken to train the algorithm with 40 datasets is much lesser. Hence, for all the forthcoming fault classifications, the results using 40 datasets each for the training and the testing are presented.

5.4.1.1 Blockage faults classification

In this section, the faults of the CP systems with the inlet and the outlet blockage faults are classified with respect to the healthy pump condition. The CP is given each blockage (both inlet and outlet) one at a time and the severity of both the inlet and outlet blockages is varied. It is very essential to identify the blockage faults as and when they form, so that they do not set off the formation of other CP faults. The results of fault classification are presented in Table 5.3. It can be observed that the overall classification accuracy of the 10 faults is over 97.8%. The SB1, SB2 and SB3 faults are perfectly classified at all the speeds. On the other hand, the HP and SB4 faults are not 100% classified at 30 Hz, 40 Hz and 35 Hz, respectively.

The outlet blockage conditions are showing diminishing classifiability with increased speeds. This may be attributed to the reduction in the vibration levels due to the drop in the CP discharge flow rate even though the CP is full. The increase in the noise-to-signal ratio may be the reason for the reduced fault segregation. Furthermore, from the table, it can be noted that the maximum average classification accuracies can be seen at 35 Hz and 40 Hz. In addition, at all the speeds, the average classification accuracy is more than 94%. All the blockage levels are identifiable by using the above developed methodology. This is very useful, as it can identify the initial stages of fault formation and the developed algorithm does not show any operational speed dependency.

Table 5.3: Case A performance of the classified with BF and HP faults²

Speed (Hz)	% fault classification accuracy										C_a	$\log_3 C$	$\log_3 \gamma$
	HP	SB1	SB2	SB3	SB4	DB1	DB2	DB3	DB4	DB5			
30	97.6	100	100	100	100	100	100	97.5	95.1	100	99	13	-9.25
35	100	100	100	100	97.5	100	97.6	100	100	100	99.5	16.5	6.5
40	97.6	100	100	100	100	97.6	100	100	100	100	99.5	16.5	6.25
45	100	100	100	100	100	100	100	100	100	90.2	99	15	-9
50	100	100	100	100	100	95.1	100	95.1	97.5	97.5	98.5	14.5	-8.5
55	100	100	100	100	100	97.6	95.1	97.6	100	92.7	98.3	13.5	-6.5
60	100	100	100	100	100	82.9	90.2	100	78	95.1	94.6	12	-6
65	100	100	100	100	100	90.2	100	100	82.9	70.7	94.4	13.5	-4.25

² Case A – Same speed fault classification, BF – Blockage faults family, HP – healthy pump, SBk – $(k/6)^{th}$ suction flow restriction, DBk – $(k/6)^{th}$ discharge flow restriction

5.4.1.2 Classification of impeller and blockage faults

In this section, the impeller faults, the impeller faults inclusive of inlet blockage faults and the impeller faults inclusive of outlet blockage faults have been considered. The classification is done with respect to the healthy pump condition. The CP under consideration in this section has a faulty impeller. Initially, this CP is run with the full inlet and outlet flows (the IF case) and then, like in the previous case, the individual inlet and outlet blockages are given to the CP with impeller defects.

The results of the classification are presented in Table 5.4. It can be observed that the overall classification accuracy of the 11 faults is over 96%. The HP and IFSB4 faults gave 100% classification accuracy at the entire speed range. However, the IF, IFSB2 and IFSB3 faults show less than 100% classification accuracy at one or two speed levels.

Like in the previous case, the outlet blockage faults on the faulty impeller also showed reduced performance. However, at all the speeds, the average classification accuracy is above 91%. The best performance can be seen at 30 Hz, with an average classification accuracy of 99.1%. The classifier is able to identify and isolate the healthy pump condition, impeller fault condition and different severities of blockages in combination with the impeller defects. Here also, like in the previous case of the blockage fault classification, the algorithm's performance is independent of the operating conditions.

Table 5.4: Case A performance of the classified with IFBF and HP faults³

Speed (Hz)	% fault classification accuracy											C_a	$\log_3 C$	$\log_3 \gamma$
	HP	IF	IFSB1	IFSB2	IFSB3	IFSB4	IFDB1	IFDB2	IFDB3	IFDB4	IFDB5			
30	100	100	100	100	100	100	100	100	97.6	92.7	100	99.1	12.75	-7.25
35	100	97.6	97.6	100	100	100	100	87.8	100	100	100	98.4	16.5	6.5
40	100	100	100	100	100	100	100	92.7	58.5	100	92.7	94.9	16.5	-8.5
45	100	100	100	100	100	100	100	65.8	92.7	100	100	96.2	14.5	-11
50	100	100	100	100	92.7	100	100	95.1	73.1	80.5	97.6	94.5	11.25	-5
55	100	100	100	100	100	100	95.1	100	100	75.6	87.8	96.2	13.75	-6
60	100	100	100	100	95.1	100	90.2	95.1	100	100	95.1	97.7	13	-4.75
65	100	95.1	95.1	97.6	100	100	43.9	97.6	82.9	95.1	100	91.6	16.5	6.5

³ Case A – Same speed fault classification, IFBF – Impeller fault and blockage faults family, HP – healthy pump, IF – impeller fault, IFSBk – impeller fault with $(k/6)^{th}$ suction flow restriction, IFDBk – impeller fault with $(k/6)^{th}$ discharge flow restriction

5.4.1.3 Classification of pitted cover plate and blockage faults

In this section, the pitted cover plate faults, and the pitted cover plate faults inclusive of the blockage (both inlet and outlet) are classified with respect to the healthy pump condition. The CP is fitted with a pitted cover plate. Initially, this CP is run with the full inlet and outlet flows (PC case) and then, like in the previous case the individual inlet and outlet blockages are given on the CP with pitted cover plate faults. The results of classification are presented in Table 5.5. It can be observed that the overall classification accuracy of the 11 faults is over 96.7%. The HP, PCSB3 and PCSB4 faults tend to be perfectly segregated at all the operating speeds. From the table, it can also be observed that some of the faults are showing low classification performances specifically at certain speeds. At all the speeds, the classification is above 90%. However, the best performance is seen at 45 Hz speed with an average classification accuracy of 99.8%. The classifier is able to segregate all the individual as well as the combined faults. The classifier is also indifferent to the operation speed of the CP.

5.4.1.4 Dry run faults

In this section, the dry run faults with each of the pump condition (e.g., the healthy impeller, the faulty impeller and the pitted cover plate) are classified with respect to the healthy pump condition. The results of the classification are presented in Table 5.6. The overall classification accuracy is found to be 99.4%. The healthy pump condition gives a 100% classification accuracy at all the speeds. However, the other three dry run conditions miss 100% classification accuracy at certain speeds. Nevertheless, the average classification accuracy at all the speeds is greater than 96%. The best performance, however, is found at the 35, 40, 50, 55 and 65 Hz. The algorithm is again speed-independent in this case also.

Table 5.5: Case A performance of the classified with PCBF and HP faults⁴

Speed (Hz)	% fault classification accuracy											C_a	$\log_3 C$	$\log_3 \gamma$
	HP	PC	PCSB1	PCSB2	PCSB3	PCSB4	PCDB1	PCDB2	PCSB3	PCSB4	PCSB5			
30	100	100	100	100	100	100	100	100	87.8	24.4	92.7	91.3	11	-5
35	100	97.6	100	97.6	100	100	100	100	97.6	100	100	99.3	16.5	6.25
40	100	100	97.6	100	100	100	100	100	97.6	95.1	100	99.1	16.5	5.5
45	100	100	100	100	100	100	97.6	100	100	100	100	99.8	16.5	6.5
50	100	100	100	100	100	100	100	95.1	70.7	100	100	96.9	1	5
55	100	97.6	100	100	100	100	100	97.6	92.7	100	100	98.9	3.25	4
60	100	87.8	97.6	100	100	100	100	97.6	92.7	100	100	97.8	16.25	5
65	100	46.3	100	56.1	100	100	97.6	100	100	100	100	90.9	1.5	5

⁴ Case A – Same speed fault classification. PCBF – Pitted cover and blockage faults family, HP – healthy pump, PC – pitted cover fault, PCSBk – pitted cover fault with $(k/6)^{th}$ suction flow restriction, PCDBk – pitted cover fault with $(k/6)^{th}$ discharge flow restriction

Table 5.6: Case A performance of the classified with dry runs⁵ and HP faults

Train/ Test	% fault classification accuracy						
Speed (Hz)	HP	SB5	IFSB5	PCSB5	C_a	$\log_3 C$	$\log_3 \gamma$
30	100	90.2	100	97.6	96.9	16.5	6.5
35	100	100	100	100	100	16.5	6.25
40	100	100	100	100	100	16.5	6.75
45	100	100	97.6	100	99.4	16.5	7.25
50	100	100	100	100	100	16.5	6.25
55	100	100	100	100	100	16.5	7
60	100	100	95.1	100	98.8	16.5	7.25
65	100	100	100	100	100	16.5	6.75

5.4.1.5 Combinations of all faults

AF⁶ faults are classified with respect to the healthy pump condition. This fault classification is done to check the classifiers capability to identify all the aforementioned fault categories. The results of this classification are presented in Table 5.7. The overall classification accuracy of this classification over the entire speed range is found to be 99.85%. The HP, SB2, SB5, DB5, IFDB1, IFDB5, PCSB5 and PCDB5 faults demonstrate perfect classification at all the speeds. Here also, the algorithm shows detachment to the operating speed of the CP. The average classification accuracy at all the speeds is above 99.5%. The best performance of

⁵ SB5 – dry run, IFSB5 – impeller faults with dry run, PCSB5 – pitted cover fault with dry run

⁶ AF – All faults family, HP – healthy pump, SBk – $(k/6)^{th}$ suction flow restriction, DBk – $(k/6)^{th}$ discharge flow restriction, IF – impeller fault, IFSBk – impeller fault with $(k/6)^{th}$ suction flow restriction, IFDBk – impeller fault with $(k/6)^{th}$ discharge flow restriction, PC – pitted cover fault, PCSBk – pitted cover fault with $(k/6)^{th}$ suction flow restriction, PCDBk – pitted cover fault with $(k/6)^{th}$ discharge flow restriction

classification (with an average classification accuracy of 100%) is found at speeds of 30 Hz, 50 Hz, 55 Hz and 60 Hz. However, the classifier seems to be very adaptable and versatile in the identification of all the faults and is unaffected by the operating conditions of the CP.

5.4.2 Case B: Fault Identification at Intermediate Test Speed

Though the classification algorithm that can perform well with the same training and testing speeds is very important, there might be cases where the CP fault data history may not exist at the current operating speed. In such cases, there needs to be a way to deal with the identification of faults in the machinery. If the raw signal data of the faults exists at a speed lower and higher than the current operating speed, Case B (refer Table 5.2) can be used to identify the fault. This case is therefore said to classify faults at intermediate speeds. The procedure used to develop the classification algorithm is very different from that of Case A.

In Case A, 50% of the data is used for the training and the rest for the testing. But in this case, as the training and testing data are completely independent of each other and are acquired at different pump fault conditions, a total of 140 training datasets ($\times 2$) and 140 testing datasets are used. As in this case, the grid search method/ cross validation technique does not guarantee a good result (because of the independence of training and testing data), therefore a parametric search is used to find optimum hyper parameter as well as the kernel parameters.

In the present case, training is done at 30-40 Hz, 40-50 Hz and 50-60 Hz speeds, and testing is done at speeds of 35 Hz, 45 Hz and 55 Hz, respectively. The faults of AF and healthy pump are attempted to be classified. The results of the classification can be seen in Figure 5.8. In the figure, 30/40-35 means that the training is done at 30 and 40 Hz, and testing is done at 35 Hz and likewise the other numbers.

Table 5.7: Case A performance of the classified with AF and HP faults⁷

Speed (Hz)	% fault classification accuracy															C_a	$\log_3 C$	$\log_3 \gamma$
	HP	SB2	SB5	DB1	DB5	IF	IFSB2	IFSB5	IFDB1	IFDB5	PC	PCSB2	PCSB5	PCDB1	PCDB5			
30	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	16.5	2.75
35	100	100	100	100	100	97.6	100	100	100	100	100	97.6	100	100	100	99.7	16.5	6.75
40	100	100	100	99.5	100	100	97.6	100	100	100	100	100	100	100	100	99.5	16.5	6
45	100	100	100	100	100	100	100	100	100	100	100	100	100	97.6	100	99.8	16.5	6
50	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	16.5	5.75
55	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	16.5	6
60	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	16.5	5.25
65	100	100	100	100	100	100	100	100	100	100	100	100	100	97.6	100	99.8	16.5	6.5

⁷ Case A – Same speed fault classification, AF – All faults family, HP – healthy pump, SBk – $(k/6)^{th}$ suction flow restriction, DBk – $(k/6)^{th}$ discharge flow restriction, IF – impeller fault, IFSBk – impeller fault with $(k/6)^{th}$ suction flow restriction, IFDBk – impeller fault with $(k/6)^{th}$ discharge flow restriction, PC – pitted cover fault, PCSBk – pitted cover fault with $(k/6)^{th}$ suction flow restriction, PCDBk – pitted cover fault with $(k/6)^{th}$ discharge flow restriction

The overall classification accuracy is above 97.5%. This means that, the classifier is very versatile and can even identify faults lying at an intermediate speed. The best performance (with the average classification accuracy of 100%) can be seen in predicting the faults at 45 Hz operational speed, while training with 40 and 50 Hz.

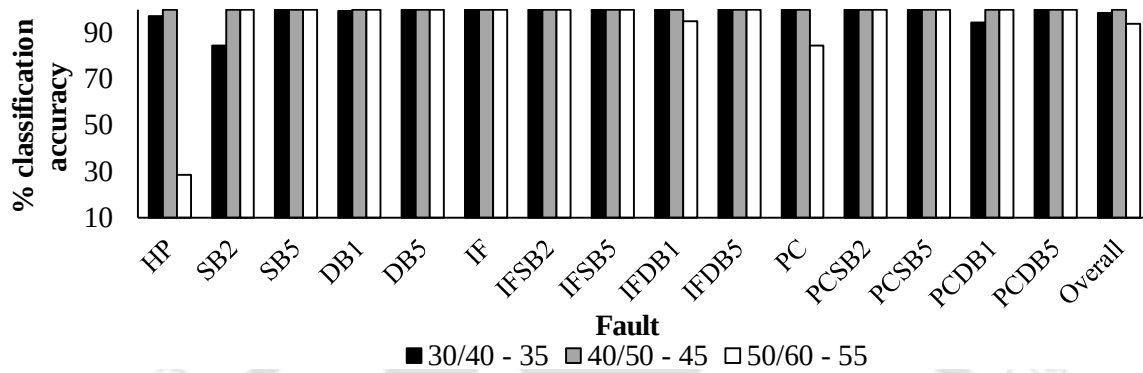


Figure 5.8: Performance of the SVM classifier at intermediate speeds⁸

5.4.3 Case C: Fault Identification with Noisy Test Data at Same Test Speed

In the next multi-class classification, the SVM algorithm is trained with the features extracted from uncorrupted signal and tested with the features extracted from signal corrupted with 5% additive white Gaussian noise (AWGN). The results of this classification are tabulated in Table 5.8. Here, it must be noted that, the CV technique has not been used for the selection of best kernel and SVM parameters. Alternatively, the best features are selected based on a parametric approach. That is, the parameters are given a range of values, and the combinations of the parameters that give the best classification performance are presented. From the table, it can be seen that the classification accuracy is found to be 100% at all the operating speeds of the CP. This also means that the overall classification accuracy of the algorithm is 100%.

⁸ 30/40 – 35, train with fault data 30 and 40 Hz speed and test with fault data at 35 Hz speed, 40/50 – 45, train with fault data 40 and 50 Hz speeds and test with fault data at 45 Hz speed, 50/60 – 55, train with fault data 50 and 60 Hz speeds and test with fault data at 55 Hz speed.

Here, it seems that the classification accuracy has improved with noise addition, but in engineering perspective, there is no significant difference between 99.85% and 100% classification. Therefore, it can be said that adding 5% AWGN does not deteriorate the classifier's performance.

The next multi-fault classification is to train the algorithm with the features extracted from the uncorrupted signals and test them with the features extracted from the signal corrupted with 10% AWGN. The results of the classification are presented in Table 5.9. All the classification accuracy values are above 99.5%. The least classification is found at 55 Hz speed of CP operation, which is due to the DB5, PC and PCDB1 faults. The overall classification accuracy is found to be 99.9%. Here also, there seems to be no operation speed dependency on the fault classification. The performance of the classifier remains unaltered with an addition of 10% AWGN.

The last multi-fault classification in Case C, is to train the algorithm with the features extracted from the uncorrupted signals and test them with the features extracted from the signal corrupted with 25% AWGN. The classification results are tabulated in Table 5.10. All the classification accuracy values are above 89.7% (which is at 60 Hz). The overall classification accuracy is found to be 96.2%. For higher speeds (beyond 55 Hz), the PC and PCSB2 faults are showing a low classifiability. Clearly, it can be seen that these faults are sensitive to the high noise at the high operating speeds. It means that the feature values are changing significantly. With the addition of 25% AWGN the performance of the classifier decreased by 4% approximately.

Table 5.8: Prediction performance of classifier when testing with uncorrupted data and training with 5% AWGN data

Speed (Hz)	% fault classification accuracy															C_a	C	γ
	HP	SB2	SB5	DB1	DB5	IF	IFSB2	IFSB5	IFDB1	IFDB5	PC	PCSB2	PCSB5	PCDB1	PCDB5			
30	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	20	0.5
35	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	20	0.5
40	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	20	2
45	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	20	0.5
50	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	20	0.5
55	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	20	2
60	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	20	50
65	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	20	1

Table 5.9: Prediction performance of classifier when testing with uncorrupted data and training with 10% AWGN data

Speed (Hz)	% fault classification accuracy															C_a	C	Γ
	HP	SB2	SB5	DB1	DB5	IF	IFSB2	IFSB5	IFDB1	IFDB5	PC	PCSB2	PCSB5	PCDB1	PCDB5			
30	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	20	5
35	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	20	5
40	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	20	2
45	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	20	0.5
50	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	20	1
55	100	100	100	100	97.6	100	100	100	100	100	97.6	100	100	97.6	100	99.5	20	3
60	100	100	100	100	100	100	100	100	100	100	99.7	99.7	100	100	100	99.7	20	50
65	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	20	50

Table 5.10: Prediction performance of classifier when the testing with the uncorrupted data and the training with 25% AWGN

Speed (Hz)	% fault classification accuracy															C_a	C	Γ
	HP	SB2	SB5	DB1	DB5	IF	IFSB2	IFSB5	IFDB1	IFDB5	PC	PCSB2	PCSB5	PCDB1	PCDB5			
30	100	100	100	100	95.2	100	100	100	100	100	100	100	100	100	100	99.7	20	4
35	100	100	100	100	97.6	100	100	100	100	100	100	88.1	100	100	100	99	20	70
40	100	100	100	100	100	95.2	100	100	100	92.9	100	88.1	100	100	100	98.4	20	1
45	100	100	100	100	95.2	100	100	100	100	90.5	100	90.5	100	100	100	98.4	20	0.5
50	100	100	100	100	97.6	100	100	100	100	100	100	97.6	100	100	100	99.7	20	50
55	97.6	97.6	100	100	83.3	100	97.6	100	90.5	85.7	66.7	88.1	100	88.1	100	93	20	5
60	83.3	97.6	100	85.7	92.9	100	88.1	100	76.2	100	61.9	64.3	100	95.2	100	89.7	20	2
65	81	95.2	100	90.5	100	97.6	95.2	92.9	100	85.7	76.2	73.8	88.1	100	97.6	91.3	20	9

However, the algorithm seems helpful in identifying the faults (giving a prediction of more than 50% for each fault). The summary of results of case C is shown in Figure 5.9. It can be seen from the figure that the algorithm is performing well even with the addition of noise. However, with 25% AWGN, the overall classification decreased by 4%.

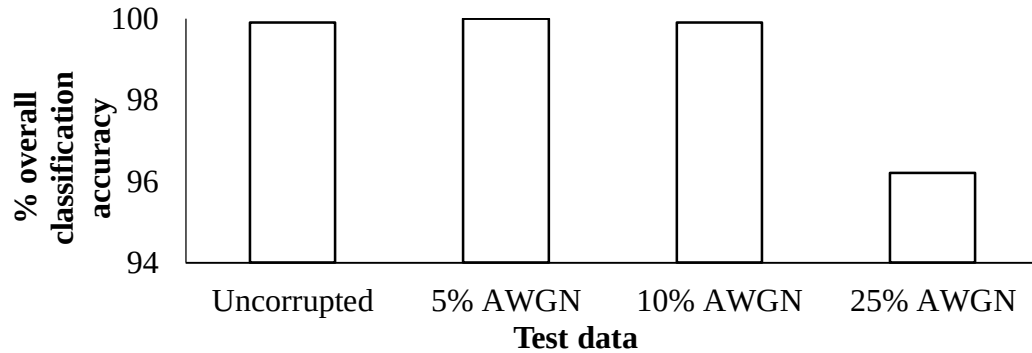


Figure 5.9: Percentage overall classification accuracy versus the quality of data used for testing the algorithm

5.5 Comparison of the results of the methodology with other works in literature

To understand the quality of the results obtained from the developed methodology in time-domain, a comparison of them with similar results from literature is presented in Table 5.11.

Following are the important observations from the table:

1. Predominantly, vibration signatures are used in time-domain for the CP fault diagnosis
2. Unique operating CP speed is considered for the fault diagnosis
3. Most of the times, independently existing CP faults are considered
4. Sakthivel et al. classified combination of mechanical faults
5. Azizi et al., 2017 and Shervani-Tabar et al., 2017 classified varying severity of cavitation

Table 5.11: Comparison of prediction performance of the current methodology with other works in literature in time-domain

References	Faults	Signal	Machine learning algorithm	Data sampling rate/ sample length	Features considered	Accuracy	Remarks
Zouari et al., 2004	Loosening of front or rare CP attachments, misalignment, cavitation, partial flow, and air injection	Vibration	NN, neuro-fuzzy network	50kHz	Fischer, PCA	n/a	1. Unique CP operating speed 2. Combination of mechanical, hydraulic faults not considered
Sakthivel et al., 2010b	Bearing faults, seal defects, impeller defects, impeller and bearing faults, cavitation	Vibration	decision tree fuzzy, rough set fuzzy		Mean, standard error, median, standard deviation, sample variance, kurtosis, skewness, range, minimum, maximum, and sum	99.33%	
Sakthivel et al., 2010a	Bearing faults, seal defects, impeller defects, impeller and bearing faults, cavitation	Vibration	C4.5 decision tree	24kHz/ 1024	Standard error, standard deviation, variance, kurtosis, skewness, range, minimum, maximum, sum	100%	1. Single speed characterization 2. Combination of mechanical faults considered
Sakthivel et al., 2012b	Bearing faults, seal defects, impeller defects, impeller and bearing faults, cavitation	Vibration	Roughset – fuzzy		mean, standard error, median, standard deviation, sample variance, kurtosis, skewness, range, minimum, maximum, and sum	97.5%	3. Combination of hydraulic and mechanical faults not considered
Sakthivel et al., 2014	Bearing faults, seal defects, impeller defects, impeller and bearing faults, cavitation	Vibration	decision tree, Bayes Net, Naïve Bayes and kNN classifiers		PCA, LLE, LTSA, HLL, Isomap, kPCA, Laplacian Eigenmap, Manifold chart, MVU	100%	
Zhao et al., 2016	Inner race wear, outer race wear, roller wear, and impeller wear	Vibration	Deep learning, softmax regression	10.24 kHz	n/a	>99%	1. Single speed characterization 2. Only mechanical faults considered
Azizi et al., 2017	No cavitation, limited cavitation and developed cavitation	Vibration	GRNN	16 kHz/ n/a	RMS, mean, variance, standard deviation, skewness, kurtosis, crest factor, upper bound, lower bound, range, shape factor, margin factor, impulse factor, entropy, zero crossing rate	100%	1. Single speed characterization 3. Varying severity of hydraulic faults considered
Shervani-Tabar et al., 2017	Five levels of cavitation	Vibration	SVM and ANN	n/a	Mean, peak, root mean square, standard deviation, kurtosis, skewness, kurtosis, crest factor, clearance factor, shape factor, impulse factor	97%	
Current research	Suction/ discharge blockages (with varying severity), impeller defects, impeller defects together with blockages, pitted cover plate faults, pitted cover plate together blockages, dry run	Vibration and Motor line current	MSVM	20 kHz/ 2000 and 5 kHz/ 5000	Mean, Standard deviation, reciprocal of standard deviation, entropy, 5 th , 6 th , 7 th moments, variance, RMS, RSS, SS	99.9%	1. Two signal types used
						(Same train and test speed)	2. Multiple speed characterization
						97.5%	3. Combination of mechanical and hydraulic faults considered
						(Intermediate speed)	4. Varying severity of faults considered
						99.9%	5. Faults classified at intermediate speeds
						(Noise)	6. Effect of data resolution presented 7. Effect of additive noise on the classification performance of the algorithm presented

From the literature survey it is understood that identifying faults with closely varying severity is a challenging task. Also, identifying combination of mechanical and hydraulic faults is a difficult task. In the current research both the aforementioned problems are addressed. The effect of data resolution on the classification results are elaborated. Also, the classification of faults is presented at intermediate speeds and with the addition of noise.

5.6 Summary

To perform the fault diagnosis of the CP faults, seventeen features have been extracted from the time-domain vibration and the current data for all the thirty-three faults. Also, a new feature which is the 'reciprocal of standard deviation' has been extracted that retains the trend of the standard deviation feature but reduces the variation in it. The wrapper model was used to select the suitable features for the fault classification. It was found that the combination of mean and reciprocal of standard deviation features performed the best at the attempted fault classification. Also, using a combination of features extracted from the vibration and current signatures, the classification accuracy improved. On comparing the time-domain based results obtained in Chapter 4 to that of the analyses presented in the current chapter, it was found that the classification accuracy indeed depended upon the sampling rate of data acquisition, i.e. lower sampling rate gave better prediction performance than that of the higher sampling rate. Also, the choice of the features plays a crucial role in the accuracy of the classifier.

From the fault classifications presented, it can be inferred that, the classification performance of BF fault family (involving suction and discharge blockages of varying severities) improved significantly by lowering the data sampling rate. This may be because the transients generated due to bubble collapses or flow instabilities were better captured using the lower sampling

rate of data acquisition. When the classification results of the combination of the mechanical and the hydraulic faults are considered (IFBF and PCBF fault families), it can be said that there is still a scope for improvement of the classifier's performance. Especially, the impeller faults (IF) generate the fault harmonics and utilizing the spectral information of the impeller faults may help improve the performance of the classifier. On the other hand, the dry run fault and AF fault family were clearly distinguished because the behavior of each fault was very different from that of the other. Therefore, it is safe to say that distinguishing/ identifying the fault severity is much more complicated than identifying the faults belonging to various families. In addition, the identification of the combination of the faults is also challenging.

SVM's capability to distinguish CP faults has been established, but it is seldom attempted to classify the faults at intermediate speeds, at which data is not available. Taking this point into account, in this chapter an attempt is made to classify multiple CP faults by training the algorithm at two distinct speeds and testing it at an intermediate speed. The algorithm worked very efficiently when trained and tested at the same speed. Even when the algorithm was tested at intermediate speeds, the performance of the classifier was very appreciable. Furthermore, to check the robustness of the algorithm, the test data was corrupted with 5, 10 or 25% additive white Gaussian noise. The algorithm seems to be immune to this noise addition. But there is a slight decrease in the classifier's performance with the 25% additive noise. Experimental fault predictions shown in this study are based on the RBF kernel with the time-domain vibration and the motor line current data. In the next chapter, the classifiers ability to distinguish the faults using the frequency-domain data will be presented.

CHAPTER 6: Multi-Fault Diagnosis of Centrifugal Pumps in Frequency Domain using Data Acquired at Low Sampling Rate

6.1 Introduction

In the previous chapter, the CP fault diagnosis was attempted using the features extracted from the time-domain data. In this chapter, the fault diagnosis of the CP faults is based on the features extracted from the frequency domain data.

The time and the frequency domain features represent the fault patterns differently. Frequency-domain data depicts the energy distribution of a signal over a range of frequencies, whereas the time-domain demonstrates the variation of the signals over time. The mechanical rotor faults are generally periodic in nature and therefore, have good frequency representation. That is the reason why frequency representation is used for the visual inspection of the signal to identify the faults. Similarly, in this chapter, it is attempted to see how the features extracted from the frequency domain can be useful in identifying the different CP faults when the SVM algorithm is trained with them.

In continuation of the methodology adopted in Chapter 5, here also many statistical features are extracted from the frequency-domain representation (power-spectrum transformation) of the faults. Therefore, when compared to the previous chapter, there is one additional step involved here, which is the transformation of the signal from the time-domain to the frequency domain. Once the features are extracted, the best representative features are selected and also the SVM parameters combination giving the highest prediction performance is chosen. The

fault diagnosis of the algorithm is tested for same speed and intermediate speed fault classification cases with both baseline and noisy data. Here also, to check the versatility of the algorithm, it is tested for a wide range of distinct speeds of the CP operation.

6.2 Fault Feature Extraction

The power spectrum for all the faults at 30 Hz speed using high and low sampling rate data is presented in Appendices D and E. From the raw-data it can be observed that it is of very high dimensionality and contains redundant information. Therefore, if this data is directly used to train the SVM classifier, the classification results may not be satisfactory. Hence, a proper selection of the statistical features needs to be carried out so that they are sensitive to the faults and to their varying severities. The statistical features extracted from the frequency domain capture the energy distribution in the signal and also the position of various characteristic frequency. They also capture the change in the position of the characteristic frequency with the operating speed of the CP.

Here, the wrapper model is used for the feature selection. Features like, the mean (μ), variance (μ_2), skewness (χ), kurtosis (κ), the 5th statistical moment (μ_5), the 6th statistical moment (μ_6), the 7th statistical moment (μ_7), standard deviation (σ), reciprocal of the standard deviation ($1/\sigma$), root mean square (RMS), root sum of squares (RSS), sum of squares (SS), shape factor (w), impulsion index (I), crest factor (P), tolerance index (T), and power to average (PA), are extracted from the raw frequency domain data. The definition of all these features is given in Table 6.1. $1/\sigma$ that has been used in this study is less frequent and often referred to as precision. This feature retains the trends of the standard deviation but reduces its variation. Higher the spread in the data, smaller would be the value of $1/\sigma$.

Table 6.1: Statistical features extracted in the frequency domain

S.N	Features	Mathematical formula	Description
1	Mean	$\mu = \frac{1}{N} \sum_{\omega=0}^N x_{\omega}$	Here, x_{ω} is the amplitude of frequency spectrum at ω frequency, and N is the Nyquist frequency
2	k^{th} statistical moment	$\mu_k = \frac{1}{N} \sum_{\omega=0}^N x_{\omega}^k - \mu^k$	Here, $k = 5, 6, 7$ for 5 th to 8 th statistical moments
3	Variance	$\mu_2 = \sigma^2; \sigma = \sqrt{\mu_2}; \frac{1}{\sigma} = \frac{1}{\sqrt{\mu_2}}$	Here, the standard deviation is given by σ and the reciprocal of the standard deviation is $1/\sigma$
4	Skewness	$\chi = \frac{\mu_3}{\sigma^3}$	Here, μ_3 is the third statistical moment
5	Kurtosis	$\kappa = \frac{\mu_4}{\sigma^4}$	Here, μ_4 is the fourth statistical moment
6	Root mean square	$RMS = \sqrt{\frac{1}{N} \sum_{\omega=0}^N x_{\omega} ^2}$	
7	Sum of squares	$SS = \sum_{\omega=0}^N x_{\omega} ^2 = RSS^2$	Here, RSS is the root sum of squares feature
8	Shape factor	$w = \frac{RMS}{\mu}$	It is the ratio of root mean square value to mean
9	Impulsion index	$I = \frac{\max x_{\omega} }{\mu}$	Here, $\max x_{\omega} $ is the maximum amplitude
10	Crest factor	$P = \frac{\max x_{\omega} }{RMS} = \sqrt{PA}$	Here, PA is power to average feature
11	Tolerance index	$T = \frac{\max x_{\omega} }{\left(\frac{1}{N} \sum_{\omega=0}^N x_{\omega} ^2 \right)^{1/2}}$	Here, $\max x_{\omega} $ is the maximum amplitude.

To demonstrate the same, Figure 6.1 is plotted between the variation of σ and $1/\sigma$ with each of the datasets for the healthy pump (HP) at 30 Hz speed. From the figure, it can be seen that there is a significant decrease in the variation of the standard deviation feature by taking a reciprocal of it. If previously its range was from 13 to 16, by taking its reciprocal it reduced to 0.064 to 0.076. Giving a data normalization-like effect.

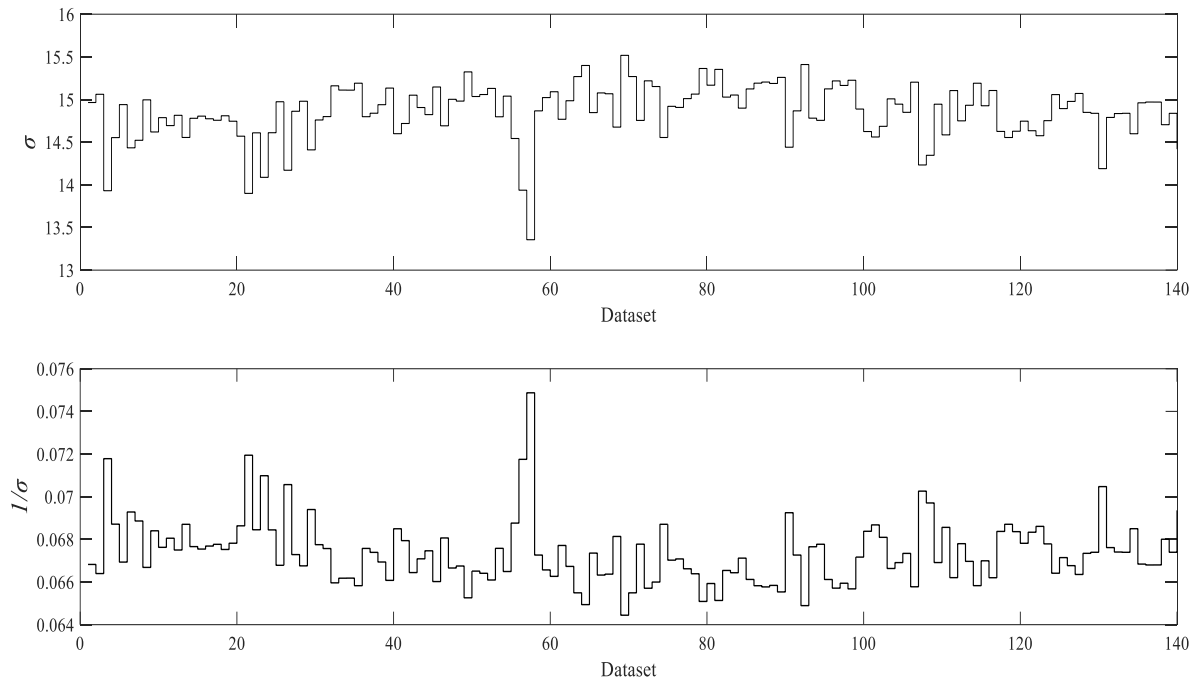


Figure 6.1: Feature values of σ (above) and $1/\sigma$ (below) for each data set of HP at 30 Hz.

Here also, like in the previous chapter, the effect of the data sampling rate (and thereby the data resolution) on the classification performance of the classifier are compared. Therefore, the features are extracted from the data (vibration and current) sampled at two different rates. The details of the amount of the data acquired and the number of features extracted is given in Table 6.2. After feature extraction, the very next important step is to select the best

performing features. In the next section, the details of the feature selection in the frequency-domain for the CP fault classification is presented.

6.3 Fault Feature Selection

Every fault in the CP contributes to the changes in its flow patterns and thus, has a unique effect on the CP signatures produced. In a multi-fault classification, it is critical to find the feature(s) that is sensitive to every fault.

The wrapper model is a performance-based feature selection technique. Therefore, to select the best features, a multi-class classification case of 15 faults is created, using, HP, SB2, SB5, DB1, DB5, IF, IFSB2, IFSB5, IFDB1, IFDB5, PC, PCSB2, PCSB5, PCDB1, and PCDB5 faults (these set of faults are described as ‘AF’ fault case in the previous chapter). These faults represent each of the fault families at intermediate severity (if applicable). The performance of different features in the aforementioned fault classification at 30 Hz operating speed is shown in Figure 6.2. From the figure, it can be seen that μ , σ , $1/\sigma$, χ , RMS and w give a prediction performance of more than 85%. The combinations of these good performing features are then taken in triplets, and it is found that the combination of μ , σ and $1/\sigma$ gives the best classification accuracy of 99.2%. Upon including any additional feature to this set of features (μ , σ , and $1/\sigma$), there is no improvement in the classification performance. Therefore, the combination of μ , σ and $1/\sigma$ features is sufficient to represent each of the fault types accurately. Henceforth, a combination of these three features (μ , σ , and $1/\sigma$) would be used for all the multi-fault classification cases.

Table 6.2: Total data collected using different sampling rates of the data acquisition

Sampling rate (kHz)	No. of samples/dataset in the time domain	No. of samples/dataset in the frequency domain (a)	No. of datasets (b)	No. of measurements using two tri-axial accelerometers (c)	No. of measurements using three line current probes (d)	No. of faults (e)	Total data collected per fault $f=\{a \times b \times (c+d)\}$	Total data collected for all faults $g=(f \times e)$	No. of features collected per dataset (h)	No. of features collected for each case $v=h \times b$
20	2000	1000	350	6	3	33	3150000	103950000	17	5950
5	5000	2500	150	6	3	33	3375000	111375000	17	2550

This feature selection has been made based on the data acquired at the low sampling rate (high-frequency resolution). To analyze the effect of data resolution, the finalized features extracted from data sampled at both the rates (high and low) are used to perform the classification of faults mentioned above and their results are compared as shown in Figure 6.3. It can be seen that the classification accuracy has improved significantly (by 8%) by using the data from the high-frequency resolution which may be attributed to the more delicate details that are captured by the statistical features. Furthermore, owing to the better performance of the higher resolution sampling, in the rest of the thesis, this would be used for various multi-fault classifications. It would be interesting to observe the scatter plot of various faults using these three features to understand their effectiveness in clustering various faults. For this purpose, a scatter plot of HP, SB2, SB5, DB1, IF, IFSB2, PC, and PCSB2 faults at 45 Hz is shown in Figure 6.4. It can be observed from the figure that all the faults are placed in separate clusters. There is a slight overlap of features from HP, SB2, PC, and PCSB2 which may be because of the closeness of the severity levels of these faults.

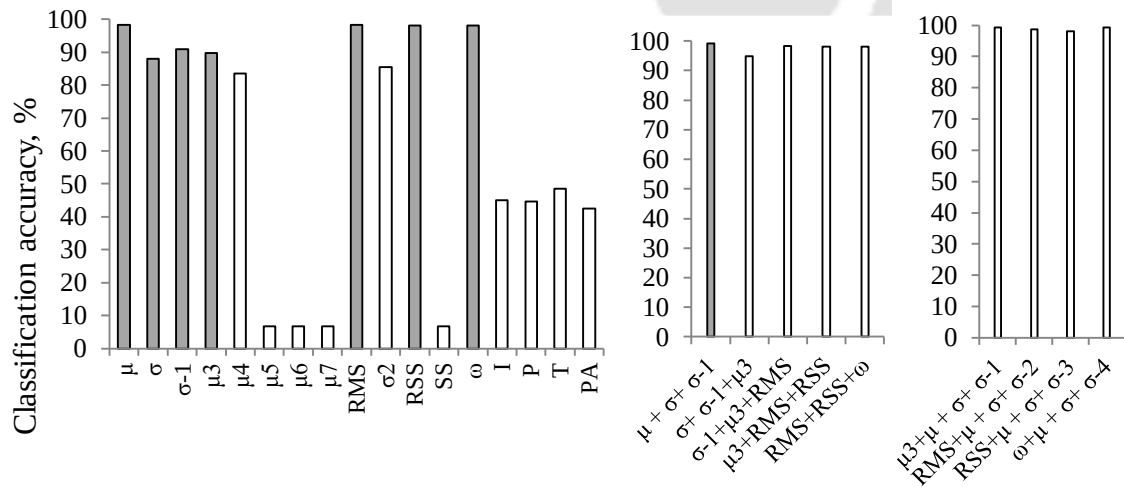


Figure 6.2: Average classification performance of various features to segregate HP, SB2, SB5, DB1, DB5, IF, IFSB2, IFSB5, IFDB1, IFDB5, PC, PCSB2, PCSB5, PCDB1 and PCDB5 faults.

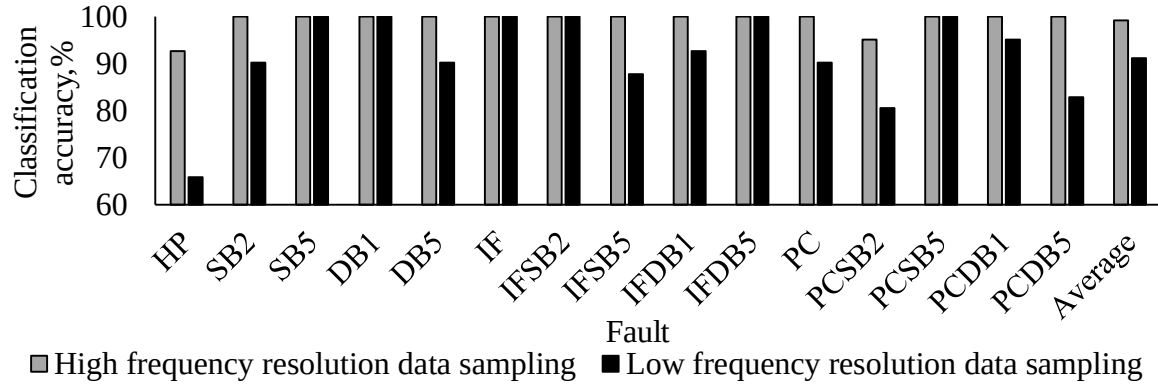


Figure 6.3: Performance comparison of the classifier using the data sampled at high frequency resolution and low frequency resolution to segregate HP, SB2, SB5, DB1, DB5, IF, IFSB2, IFSB5, IFDB1, IFDB5, PC, PCSB2, PCSB5, PCDB1 and PCDB5 faults.

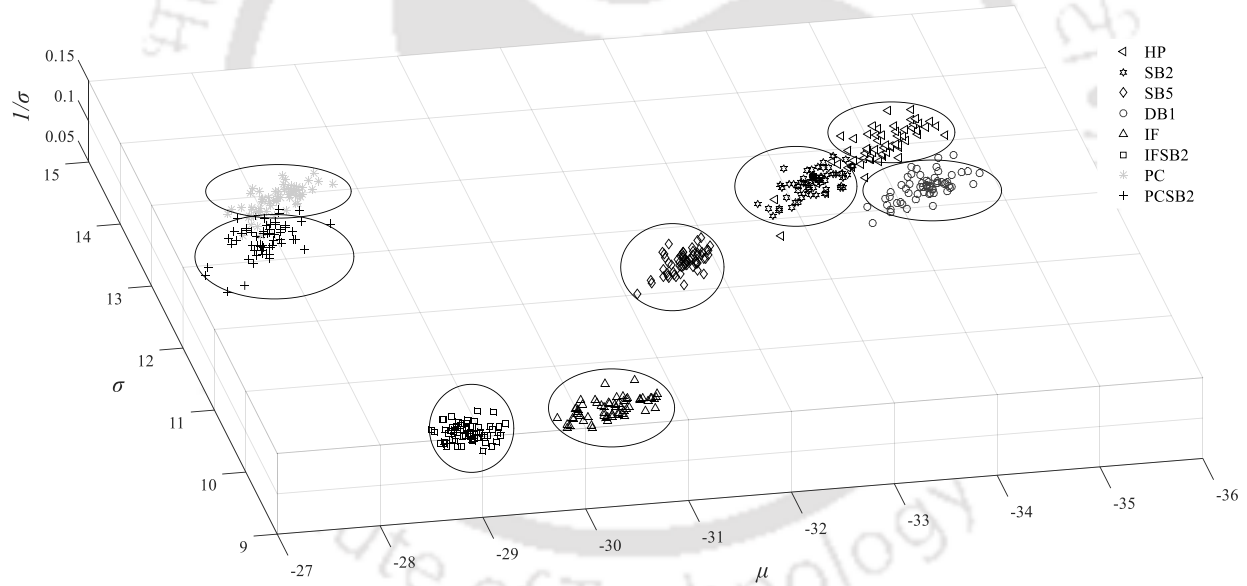


Figure 6.4: Scatter plot of HP, SB2, SB5, DB1, IF, IFSB2, PC, and PCSB2 faults at 45 Hz using features μ , σ , and $1/\sigma$

6.4 Fault Diagnosis Using the SVM

In this section, results of the multi-class classification of various faults are presented. First, different categories of the fault classifications are given, later, the significance of the

considered fault classification category is illustrated, and their classification results are presented. The fault classification cases considered are summarised in Table 6.3. Please refer to Section 5.4 for the detailed description of various fault families.

6.4.1 Case A: Same Speed Classification

In Case A, five categories of faults are classified, namely, the blockage faults, the impeller faults, the pitted cover faults, the dry run faults and a prescribed combination of all the above faults. It must be noted that the algorithm is trained and tested from the same operating speed data of the CP in this case.

Table 6.3: Fault classification cases considered

Case	Faults for training	Faults for testing	Training speed (Hz)	Testing speed
A	BF, HP	BF, HP	30 – 65	Same as training
	IFBF, HP	IFBF, HP		
	PCBF, HP	PCBF, HP		
	Dry runs, HP	Dry runs, HP		
B	AF	AF	30 – 65	Intermediate speed
	AF	AF		
C	AF	AF	30 – 65	Noisy signal at same training speed

6.4.1.1 Blockage faults classification

The blockage faults primarily constitute the suction and the discharge blockages with varying severity levels. The abbreviation – BF represent this family. In this case, an attempt is made

to segregate BFs and identify their severities. The accurate recognition of these faults at their commencement will help us to take the corrective measures to stop their progression. In this classification, ten faults are considered, four severities of suction blockages, including, SB1, SB2, SB3 and SB4, five severities of discharge blockages, including, DB1, DB2, DB3, DB4, DB5 and a healthy CP condition, HP. All these blockages are administered on a healthy CP one at a time. The training and the testing of the algorithm is done at the same operating speed of the CP. The results of this classification are presented in Table 6.4.

From the table, it can be seen that the average classification accuracy at all the speeds is above 91% and the overall classification accuracy is more than 96%. Also, the best classification performance of 98.3% is found at 45 Hz speed. It can also be observed that the average classification accuracy is improving with speed up to a speed of 45 Hz, after which there is no clear trend which may be attributed to the stabilization of the faults at higher speeds even though they are severe. The DB1 fault shows a complete classification at all the speeds. The classification results are encouraging, and this algorithm seems useful in identifying each of the blockage fault types along with their severities.

6.4.1.2 IFBF fault classification

Impeller cracks are mechanical faults; they not only create the CP unbalance but also disturb the fluid-flow pattern causing a pseudo-flow re-circulation. They even act as stress-hot spots. If impeller defect is present in combination with the fluid-flow blockages, it may result in an accelerated CP failure. In this classification case, eleven faults are considered, including, (i) the impeller defect alone (IF), (ii) the four levels of the suction blockages on the CP with the impeller defects (IFSB1 to IFSB4), (iii) the five levels of the discharge blockages on the CP

with the impeller defects (IFDB1 to IFDB5), and (iv) the healthy pump condition (HP). The reason for considering the HP condition is to check if the algorithm is precisely able to distinguish between the faulty and the non-faulty classes. It may be noted that the algorithm here is attempting to classify, healthy pump (HP), mechanical faults (IF at low speeds), mechanical faults and pseudo-flow re-circulation (IF, at higher speeds of CP operation), mechanical faults, pseudo-flow re-circulation combined with the discharge re-circulation (IFSB1, IFSB2, IFSB3, and IFSB4), mechanical faults, pseudo-flow re-circulation combined with the suction re-circulation (IFDB1, IFDB2, IFDB3, IFDB4 and IFDB5).

The training and the testing of the classifier is done using the features obtained from the data acquired at the same CP operating speed. The results of this classification are presented in Table 6.5. It can be seen from the results that the average classification accuracy at all the speeds is above 92%. Also, the overall classification accuracy is found to be 97.1%, and the best classification performance of 99.6% is found at 60 Hz the CP speed. HP shows a perfect classification at all the speeds.

The IF, IFSB2 and IFSB3 faults also show nearly perfect classifications. It can also be observed that there is no clear trend in the fault classification performance, which means that the algorithm is independent of the operating speeds of the CP. The complex fluid-structure interactions may be the reason for some faults not being able to achieve a 100% fault classification. However, the methodology seems to be very useful in identifying the different flow instabilities and the mechanical impeller faults.

Table 6.4: Case A BF fault classification performance

Train/ Test Speed (Hz)	% fault classification accuracy										C_a	$\log_3 C$	$\log_3 \gamma$
	HP	SB1	SB2	SB3	SB4	DB1	DB2	DB3	DB4	DB5			
30	92.7	90.2	100	92.7	100	100	100	95.1	60.9	80.5	91.2	11	-11
35	100	95.1	100	100	92.7	100	97.6	95.1	82.9	78	94.1	15.25	-12.5
40	97.6	100	100	100	97.6	100	100	95.1	87.8	100	97.8	8	-7.75
45	95.1	100	100	97.6	100	100	100	100	97.6	92.7	98.3	14	-14
50	97.6	95.1	97.6	100	100	100	100	95.1	92.7	95.1	97.3	16.5	-10.75
55	85.4	100	97.6	87.8	95.1	100	100	100	97.6	92.7	95.6	11	-10
60	97.6	100	100	100	100	100	90.2	100	100	92.7	98	16.5	-7
65	100	97.6	100	100	97.6	100	100	100	82.9	80.5	95.9	16.5	-12

6.4.1.3 PCBF fault classification

The defects on the cover plate may result in the turbulence inside the CP leading to a decrease in the pump performance. Additionally, a combination of the pitted cover plate fault with blockage faults may lead to the enhancement of the flow turbulence in the CP. In this case, the suction and the discharge blockages are given individually on the CP with pitted cover plate faults. These faults are classified with respect to the HP condition and the PC condition. Like in the previous case of fault classification (Section 6.4.1.2), here also, the mechanical faults and the mechanical faults coupled with different types of blockage faults are classified. The training and the testing of the algorithm is done at the same operating speed of the CP.

The results of the classification are presented in Table 6.6. From the table, it can be observed that at all the speeds the classification accuracy is above 89.6%. In addition to this, the overall classification accuracy at all the speeds is found to be 93.9%. The highest classification accuracy of 98.7% is found at the CP operating speed of 55 Hz. The SB and the PCSB4 faults give the complete classification at all the operating speeds. There is no clear trend in the fault classification, and hence it can be said that the developed algorithm is versatile at all the operating speeds of the CP. The methodology is working desirably in classifying the mechanical faults as well as the flow instabilities.

6.4.1.4 Dry run fault classification

In most of the CPs, the priming is essential. Non-priming of the CP leads to a heat generation, which may result in bearing and seal failures. Hence, the dry run condition needs to be identified instantly.

Table 6.5: Case A IFBF fault classification performance

Train / Test Speed (Hz)		% fault classification accuracy										C_a	$\log_3 C$	$\log_3 \gamma$	
		HP	IF	IFSB1	IFSB2	IFSB3	IFSB4	IFDB1	IFDB2	IFDB3	IFDB4				IFDB5
30		100	100	100	100	100	100	100	100	92.7	87.8	90.2	97.3	15	-14
35		100	100	97.6	100	100	97.6	100	100	100	97.6	100	99.3	16.5	-11.75
40		100	100	100	97.6	100	90.2	92.7	97.6	100	97.6	73.2	95.4	15.5	-12.75
45		100	97.6	100	100	97.6	100	97.6	60.9	65.8	97.6	100	92.5	15.5	-14.25
50		100	100	100	100	100	100	100	95.1	48.8	97.6	87.8	93.6	13	-12
55		100	100	100	100	100	100	100	97.6	95.1	95.1	100	98.9	12.5	-11
60		100	100	100	100	100	100	100	100	100	100	100	100	13	-13
65		100	100	100	100	100	97.6	100	100	100	100	100	99.8	14	-13.5

In this case, four faults are classified which include, the dry run of a healthy CP (SB5), the dry run of a CP with impeller defects (IFSB5), the dry run of CP with pitted cover plate faults (PCSB5) and a healthy pump condition (HP). The results of the classification are presented in Table 6.7. It can be observed from the table that, the classification accuracy at all the speeds is above 99.4%. The overall classification accuracy for all the speeds is 99.9%. There is perfect segregation of faults at the speeds of 30, 40, 45, 55, 60 and 65 Hz. The IFSB5 and PCSB5 faults are correctly classified at all the operating speeds. There is no relation between the classification performance and the CP operating speed. Hence, the algorithm can be considered robust and can be used to identify the dry run conditions readily.

6.4.1.5 AF fault classification

All the classification cases presented so far considered the segregation of the faults within their specific family/group. Those cases would be applicable in places where the CPs repeatedly fail due to a particular identified cause, and it is essential to recognize the fault severity. However, there may be cases where the causes of the failure may be multiple, and the faults may exist simultaneously. In such cases, the present classification would be useful. In this classification case, fifteen faults, viz., no blockage, SB2, dry runs, DB1, and DB5 for all the three CP configurations along with the HP condition are considered. This fault family is abbreviated as – AF. The algorithm is trained and tested at the same operating speed of the CP. Results of this classification are presented in Table 6.8. It can be seen from the table that the classification accuracy is above 96.9% at all the speeds. The overall classification accuracy at all the speeds is found to be 99.1%. DB1, DB5, IFSB2, IFSB5, PC and PCSB5 faults give a complete classification at all the speeds. At the speeds of 40 and 60 Hz, there is a 100% fault classification.

Table 6.6: Case A PCBF fault classification performance

Train/ Test Speed (Hz)	Fault classification accuracy											C_a	$\log_3 C$	$\log_3 \gamma$
	HP	PC	PCSB1	PCSB2	PCSB3	PCSB4	PCDB1	PCDB2	PCDB3	PCDB4	PCDB5			
30	100	100	100	87.8	100	100	100	70.7	95.1	60.9	87.8	91.1	13	-11.25
35	100	95.1	80.5	97.6	97.6	100	97.6	90.2	82.9	85.4	95.1	92.3	8	-8
40	100	100	95.1	100	100	100	95.1	97.6	85.4	95.1	87.4	95.8	13.25	-12.5
45	100	100	100	100	100	100	92.7	95.1	90.2	100	73.2	95.6	12.5	-11.5
50	100	92.7	100	80.5	100	100	68.3	100	85.4	85.4	100	92	12	-10.75
55	100	97.6	97.6	97.6	100	100	95.1	100	100	100	97.6	98.7	14.5	-14
60	100	92.7	97.6	100	100	100	85.4	29.7	87.8	92.7	100	89.6	15.5	-13
65	100	97.6	100	100	97.6	100	97.6	100	90.2	95.1	75.6	95.8	11.5	-10.25

Table 6.7: Case A dry run fault classification performance

Train/ Test Speed (Hz)	% fault classification accuracy				C_a	$\log_3 C$	$\log_3 \gamma$
	HP	SB5	IFSB5	PCSB5			
30	100	100	100	100	100	16.5	-6.75
35	97.6	100	100	100	99.4	16.5	-4.25
40	100	100	100	100	100	16.5	-4.5
45	100	100	100	100	100	16.5	-5.75
50	97.6	100	100	100	99.4	16.5	-5
55	100	100	100	100	100	16.5	-5
60	100	100	100	100	100	16.5	-3.5
65	100	100	100	100	100	16.5	-5.25

There seems to be no correlation between the classification accuracy and the operating speed. It is interesting to note that this classification aims at identifying the pure mechanical faults, the flow faults, the pseudo-re-circulation faults and the mechanical faults combined with the different re-circulations mentioned above at varied operating conditions of the CP. The performance of the methodology is very impressive.

6.4.2 Case B: Fault Identification at an Intermediate Test Speed

Further, as a separate study, the faults of case B are attempted to be classified at intermediate speeds, i.e., the algorithm is trained with the data at two distinct speeds and tested with the data at an intermediate speed. This is important in cases where the fault training data is not available at the present operating speed of the CP. For this, the algorithm is trained at 30-40, 40-50, and 50-60 Hz speeds, tested at 35, 45 and 55 Hz, respectively. The results of the classification are presented in Table 6.9.

Table 6.8: Case A AF fault classification performance

Speed (Hz)	% fault classification accuracy															C_a	$\log_3 C$	$\log_3 \gamma$
	HP	SB2	SB5	DB1	DB5	IF	IFSB2	IFSB5	IFDB1	IFDB5	PC	PCSB2	PCSB5	PCDB1	PCDB5			
30	92.7	100	100	100	100	100	100	100	100	100	100	95.1	100	100	100	99.2	16.5	-7
35	100	100	100	100	100	100	100	100	97.6	100	100	97.6	100	97.6	100	99.5	15	-16
40	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	16.5	-9.5
45	100	100	100	100	100	97.6	100	100	100	100	100	100	100	92.7	85.4	98.4	13	-11.75
50	100	95.1	75.7	100	100	100	100	100	100	100	100	82.9	100	100	100	96.9	16.5	-14.75
55	82.9	100	100	100	100	100	100	100	100	100	100	100	100	100	100	98.9	16.5	-10
60	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	16.5	-7.5
65	100	100	100	100	100	100	100	100	100	100	100	100	100	95.1	100	99.7	16.5	-8.25

Here, the multi-class fault classification of AF faults is attempted. It can be seen from the table that the overall test classification accuracy is 83.2%. Clearly, the classification performance has decreased in comparison to the same speed training and testing condition. DB1, IFSB2, and IFSB5 faults are classified perfectly at the intermediate speeds. At higher speeds, the fault signatures get very distinct, and hence the classification performance is decreasing with increasing speed. However, the results are promising, and the developed algorithm may be used for the fault identification at intermediate speeds.

6.4.3 Case C: Fault Identification with Noisy Test Data at the Same Test Speed

In this fault classification case, the robustness of the classifier is tested in handling noisy data. In the previous chapter, the additive white Gaussian noise of up to 25% was added to the raw time-domain data. Also, it was found that the classifier was able to handle so much noise without much deterioration in the performance. However, in the frequency domain, with the addition of noise, there is a significant shift in the feature values which directly affects the classifier's performance.

Since the classifier is trained with the features extracted from the original power spectrum and tested with the features extracted from the power spectrum of the corrupted time-domain data, it is essential that the feature values be unaffected by the noise addition for adequate performance of the classifier.

Figure 6.5 shows the mean feature value for 50 datasets. From the graph, it can be observed that there is no significant variation in μ with the addition of noise. However, if the standard deviation is plotted against the sample set number, it is observed that there is much difference

between the σ feature values of the original data and the corrupted data, as shown in Figure 6.6. That means to say that the fault data is shifting in the feature space while using the σ feature.) Hence, μ would be employed for further analysis of the corrupt data. Also, since the variation in the feature values is evident, as an initial study, only 3% and 5% additive Gaussian noise have been added to the signals. In the previous chapter, it was demonstrated that the use of features extracted from the current signal along with that of the features extracted from the vibration signals improved the classifier's performance. A similar study is carried out here, on how the combination of the signal features affects the classifier's performance especially when the features extracted from the corrupted data are used for testing.

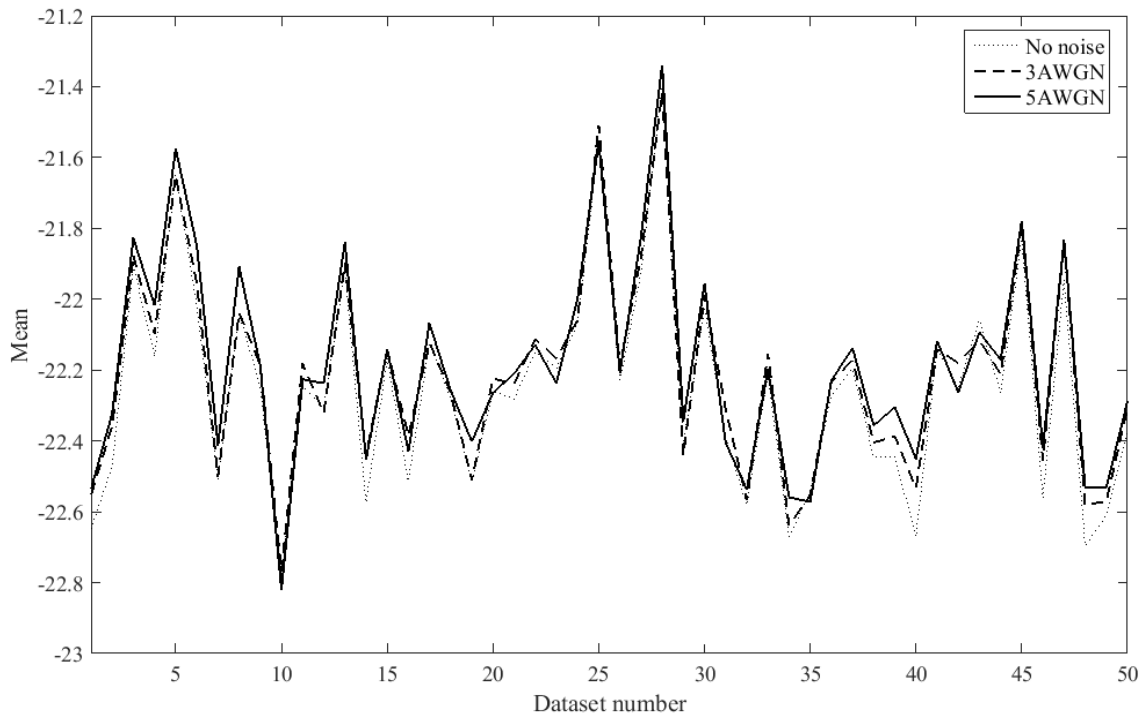


Figure 6.5: Mean versus sample set

Table 6.9: Case B AF fault classification performance at intermediate speeds

Train /Test Speed (Hz)	% fault classification accuracy															C_a	C	I
	HP	SB2	SB5	DB1	DB5	IF	IFSB2	IFSB5	IFDB1	IFDB5	PC	PCSB2	PCSB5	PCDB1	PCDB5			
30,40/35	88.7	23.4	100	100	100	100	100	100	99.3	69.5	90	100	100	95	99.3	91	1e7	5e-7
40,50/45	92.2	87.2	82.9	100	87.9	90	100	100	100	100	100	91.5	61.7	100	0	86.2	1e7	5e-7
50,60/55	73.2	92.7	100	100	31.7	65.8	100	100	95.1	100	100	100	29.3	0	0	72.5	1e7	5e-7

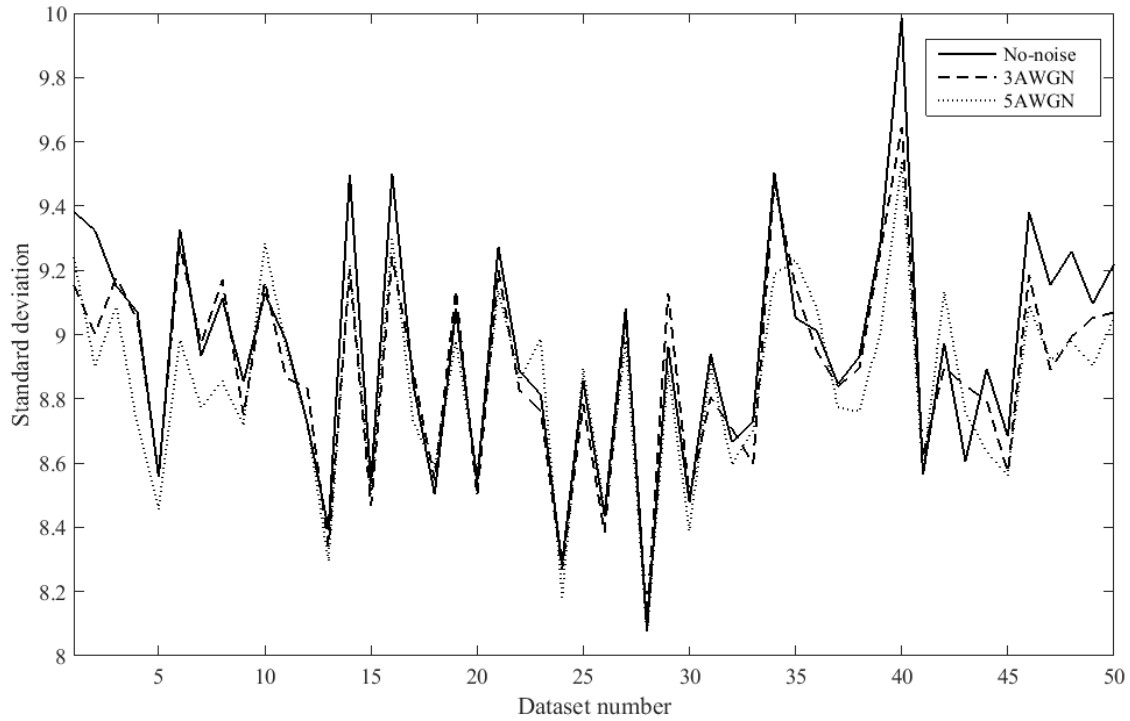


Figure 6.6: Standard deviation versus sample set

Figure 6.7 shows the classification performance of the corrupt data using the vibration data (V) and the combination of vibration and current (V&C) data. It can be observed that for 3AWGN, the overall accuracy improved from 84.9% to 96%, with the addition of current signals which is an increment of approximately 11% in the classification accuracy.

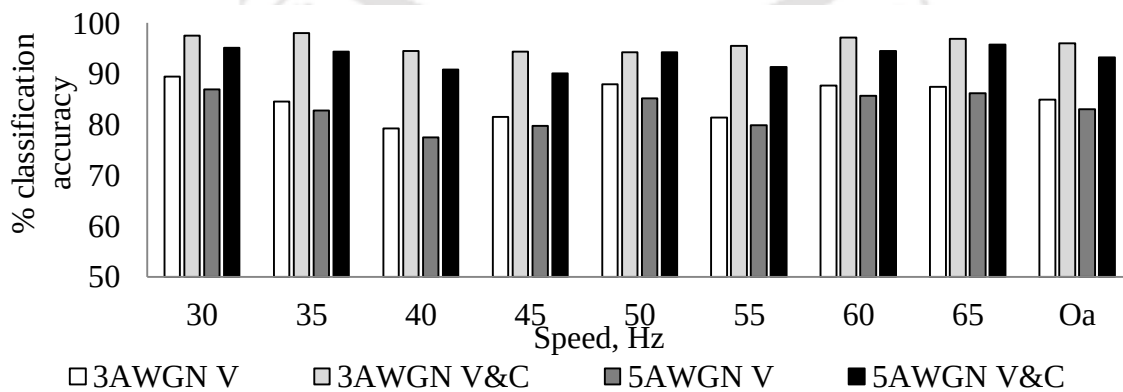


Figure 6.7: Performance of vibration and current signals

Similarly, for 5AWGN, the overall accuracy improved from 83% to 93.3% which is again a 10.3% increment. The results suggest that the combination of the vibration and the current signal is beneficial, especially when external noise corrupts the data.

The results of the classification of the 3AWGN data are presented in Table 6.10. It must be noted that the classifier is tested for a multi-class classification of AF faults. The classifier is performing very well and is giving an overall classification accuracy of 96%. However, it must be noted that, with the addition of noise, overall classification accuracy decreased by 3% (refer Table 6.8 and Section 6.4.1.5) when compared to the classification accuracy obtained while using the original data. Nonetheless, even with the addition of noise, the algorithm can distinguish all the faults. SB2, SB5, IF, IFDB5, and PCSB5 faults are correctly classified at almost all the operating speeds of the CP.

Similarly, the results of the classification of the 5AWGN data are presented in Table 6.11. The performance of the classifier decreased further, giving an overall classification accuracy of 93.3%. However, the SB5 and the PCSB5 faults give nearly a perfect classification at all the operating speeds of the CP. The reason why the other faults are showing slightly lesser prediction may be due to the closeness of the characteristics of the fault signatures of the various CP flow instabilities. That is, when noise is added to the signal, the fault signatures are getting mixed up.

The overall classification performance of no-noise, 3AWGN, and 5AWGN data are presented in Figure 6.8. From the figure, it can be observed that there is approximately a 3% decrease in the overall classification accuracy with close to 3% increase in noise. However, for noise

levels of less than 10%, the algorithm seems to be effective giving a satisfactory classification of the faults.

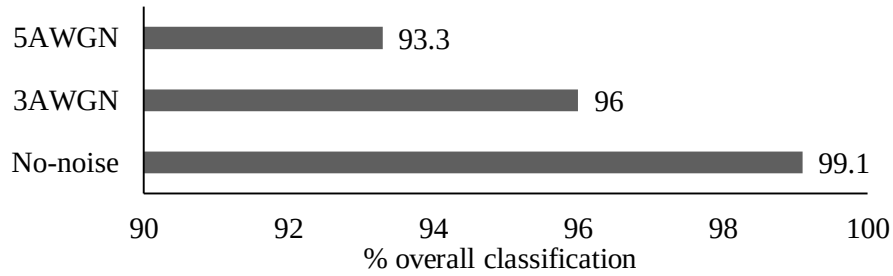


Figure 6.8: Overall classification of uncorrupted, 3AWGN and 5AWGN data

6.5 Comparison of the results in frequency-domain with other similar work in literature

To understand the performance of the developed methodology in frequency-domain to that of the other reported studies, a table of comparisons, Table 6.12 is presented. Following are the salient points derived from the table,

1. Perovic et al., 2001, used motor line current signals to diagnose CP faults. Also, the authors tried to classify combination of mechanical and hydraulic CP faults. However, the classification accuracy obtained was less.
2. Xue et al., 2014, tried to identify varying severity of cavitation.
3. Most of the researchers employed vibration signatures for the diagnosis of faults in frequency domain.
4. Compared to the time-domain results presented in Table 5.11, the accuracy of frequency domain based analysis is slightly less, in general.
5. All the classifications were attempted at a unique CP operating speed.

Table 6.10: Multi fault classification with 3AWGN corrupted data

Speed (Hz)	% fault classification accuracy															C_a	C	γ
	HP	SB2	SB5	DB1	DB5	IF	IFSB2	IFSB5	IFDB1	IFDB5	PC	PCSB2	PCSB5	PCDB1	PCDB5			
30	98.4	95.9	100	100	95	100	100	95.9	96.7	100	100	96.7	93.4	95.9	94.2	97.5	5e5	15
35	98.4	100	100	100	96.7	100	98.9	98.4	97.5	100	90.1	98.4	96.7	98.4	96.7	98.0	1e3	20
40	95.9	98.4	98.4	98.4	94.2	100	96.7	96.7	93.4	97.5	86.8	96.7	100	72.7	91.7	94.5	5e5	15
45	90.1	100	100	100	92.6	100	92.6	99.2	88.4	100	89.3	93.4	100	89.3	81	94.4	5e5	15
50	90.1	100	100	95.9	98.4	100	100	100	98.4	95.9	89.3	95	100	72.7	77.7	94.2	5e5	0.09
55	100	100	100	98.4	100	95.9	91.7	96.7	92.6	93.4	90.1	88.4	100	92.6	92.6	95.5	5e5	15
60	96.7	96.7	100	99.2	98.4	95	100	97.5	86.8	100	94.2	95.9	100	98.4	97.5	97.1	5e5	0.01
65	98.4	99.2	100	91.7	97.5	97.5	99.2	100	94.2	100	91.7	94.2	100	91.7	97.5	96.9	5e5	0.01

Table 6.11: Multi fault classification with 5AWGN corrupted data

Speed (Hz)	FAULT CONDITION															C_a	C	γ
	HP	SB2	SB5	DB1	DB5	IF	IFSB2	IFSB5	IFDB1	IFDB5	PC	PCSB2	PCSB5	PCDB1	PCDB5			
30	93.4	95.9	100	100	93.4	100	97.5	91.7	99.2	99.2	85.6	96.7	86.8	93.4	93.4	95.1	5e5	1e-3
35	95	96.7	100	96.7	94.2	99.2	90	100	95.9	100	77.7	95	87.6	89.3	98.4	94.4	5e5	1e-1
40	83.5	99.2	100	91.7	79.3	97.5	93.4	98.4	92.6	85.1	81	96.7	100	66.1	98.4	90.9	5e5	1e-1
45	76	99.2	100	100	85.6	99.2	74.4	100	76	98.4	75.2	99.2	100	77.7	90.1	90.1	5e5	1e-1
50	94.2	98.4	100	93.4	90.9	100	100	100	96.7	93.4	81.8	95	100	71.1	98.5	94.2	5e5	1e-1
55	98.4	100	100	96.7	95.9	74.4	93.4	85.1	95	90.9	82.6	82.6	100	88.4	86	91.3	5e5	9e-1
60	95.9	95	100	93.4	98.4	94.2	99.2	98.4	71.9	97.5	94.2	92.6	100	98.4	87.6	94.4	5e5	1e-1
65	96.7	95.9	100	89.3	98.4	95.9	100	100	93.4	98.4	92.6	92.6	100	91.7	90.9	95.7	5e5	1e-1

From the literature survey it is understood that identifying faults with closely varying severity and combination of mechanical and hydraulic faults is a challenging task. In the current research both the aforementioned problems are addressed. The effect of data resolution on the classification results are elaborated. Also, the classification of faults is presented at intermediate speeds and with the addition of noise. The classification accuracies obtained are not just at one CP operating speed, but are in a range of CP operating speeds. This is important because with the changing CP speed, the frequency spectrum of the fault also changes, thus directly effecting the extracted features. This important issue is addressed in developing the current methodology in frequency-domain.

6.6 Summary

As it was observed, the spectral data is handy for fault identification by itself. With the use of developed methodology, the frequency domain classification was found to give satisfactory classification performance both with and without additional noise, at same and intermediate speeds. However, compared to the results from Chapter 5, it could be said that the performance with the frequency-domain was not at par with that of the time-domain at intermediate speeds and also with noise addition. This may be because the fault frequencies change with CP speed and thus shift the characteristic a change in feature values. Thus, making it is difficult to identify the faults at intermediate speeds. Nevertheless, the performance of the classifier to identify the IFBF fault family improved using the frequency-domain features in contrast to the time-domain features (results presented in Chapter 5).

Table 6.12: Comparison of fault prediction performance in frequency domain to that presented in literature

References	Faults	Signal	Machine learning algorithm	Data sampling rate/ sample length	Features considered	Accuracy	Remarks
Perovic et al., 2001	Discharge blockage faults, impeller defects and both discharge blockage and impeller defects	Motor current	Fuzzy logic	n/a	Slip, current amplitude, amplitude of the shaft speed component, accumulative noise	83.7%	1. Combination of mechanical and hydraulic faults classified 2. Faults classified at a single operating CP speed
Yuan and Chu, 2006	Gear damage, structure resonance, rotor radial touch friction, shaft crack, bearing damage, body join looseness, bearing looseness, rotor part looseness, pressure pulse, cavitation, vane rupture, rotor eccentricity, shaft bend	Vibration	SVM	n/a	Relative spectral amplitude, PCA	Precession ratio >95%	1. Faults classified at a single operating CP speed
Zhao et al., 2010	Impeller leading edge and trailing edge damage	Vibration	Probabilistic neural network, k nearest neighbours, SVM	9kHz	Noise based features	99.56%	1. Faults classified at a single operating CP speed
Farokhzad et al., 2012	Impeller defects, seal defects and cavitation	Vibration	ANN	n/a	Mean, standard deviation, variance, skewness, kurtosis, crest factor, slippage, root mean square	100%	1. Faults classified at a single operating CP speed
Xue et al., 2014	Varying severities of cavitation, unbalance and misalignment	Vibration	SVM	50kHz	NSP	0.9649/1	1. Faults classified at a single operating CP speed 2. Varying severity of cavitation considered 1. Two signal types used 2. Multiple speed characterization
Current research	Suction/ discharge blockages (with varying severity), impeller defects, impeller defects together with blockages, pitted cover plate faults, pitted cover plate together blockages, dry run	Vibration and Motor line current	MSVM	20 kHz/ 2000 and 5 kHz/ 5000	Mean, Standard deviation, reciprocal of standard deviation, entropy, 5 th , 6 th , 7 th moments, variance, RMS, RSS, SS	99.1% (Same train and test speed) 83.2% (Intermediate speed) 96% (Noise)	3. Combination of mechanical and hydraulic faults considered 4. Varying severity of faults considered 5. Faults classified at intermediate speeds 6. Effect of data resolution presented

Hence, it is evident that the frequency content (especially of rotating faults) needs to be taken into account to get better classification results. However, while doing the frequency analysis (like the one presented in this chapter), the vibration and the current signals are considered to be stationary which is not always a correct assumption especially while dealing with faults that produce transients. Hence, in the next chapter, a time-frequency analysis is presented to overcome the shortcomings of the frequency analysis.





CHAPTER 7: Multi-Fault Diagnosis of Centrifugal Pumps using Continuous Wavelet Transform Analysis

7.1 Introduction

In the previous chapter, the fault diagnosis of the centrifugal pumps was made using the features extracted from the frequency-domain (power spectrum) data in which the fault signals were considered to be stationary in nature. However, in reality, the fluid-flow faults and the complex combinations of the mechanical and the hydraulic faults are not always stationary.

For example, a bubble implosion creates a short transient signal which may not be accurately captured using simple frequency-domain studies, because in frequency-domain, the time aspect of the frequency components is ignored. Therefore, there is a need for evaluating the changing frequency spectra with respect to time, that is, time-frequency domain analysis is required. This time-frequency transformation may be obtained by using a wavelet transform (WT) analysis. Projection of a time signal onto the wavelet basis functions is called 'wavelet transform.'

In this chapter, the multiple fault diagnosis of the mechanical and the hydraulic CP faults, and their severity estimation is performed using the multi-class SVM based on one variant of WT, i.e., the continuous wavelet transform (CWT). Since wavelets are capable of handling the non-stationary signals, they are being used very widely for the condition monitoring of the machinery. In Fourier transform, the sine and the cosine functions are used as the basis functions, these functions are global in time-domain but are localized in frequency-domain. However, in the WT, the 'wavelet' functions which are localized both in the time and the

frequency domains are used. That is, wavelets are short-lived functions. The CWT splits the signal into scaled and transformed versions of the mother wavelet.

The low sampling rate time-domain data acquired is transformed to the time-frequency domain using the CWT. There is no thumb-rule on the choice of the mother wavelet. However, this choice is very much dependent on the nature of the problem, and it affects the intricacy of the features extracted.

Hence, nine different base wavelets are chosen to investigate the effect of various wavelet families on the CP fault classification. To select the best scales and the corresponding coefficients, the wavelet energy method and the principal component analysis (PCA) method are adopted. The relative performance of both these methods is also compared. The features are extracted from the wavelet coefficients. The appropriate features for the fault classification are selected using the wrapper model. These features are then fed to the SVM classifier whose parameters are optimized using a 5-fold cross-validation technique. The algorithm has been tested for both benchmarks as well as noisy time-domain data.

7.2 Fault Feature Extraction

In this section, the details of the continuous wavelet transform (CWT), the method applied for wavelet scale selection, and feature selection procedure is explained.

7.2.1 Continuous Wavelet Transform

The choice of the domain of signal used for the classification greatly relies on the behavior of the faults. In general, mechanical faults such as the impeller cracks, bearing damages,

unbalances and misalignments produce characteristic fault frequencies, which would depend on the operating speed of the rotor. These signals are periodic and stationary. Thus, these frequencies can be captured using a fast Fourier transform (FFT) approach. However, the hydraulic faults produce transient signals. FFT considers the signal to be stationary and so, it may not be able to capture the transients in a signal. Therefore, while dealing with a combination of mechanical and hydraulic faults, a time-frequency domain analysis may be useful.

The continuous wavelet transform (CWT) converts the time-domain signal into the time-frequency domain. The basic CWT theory and its application in fault diagnosis have been discussed in refs. (Tse et al., 2004; Wu and Chen, 2006). For an original time domain signal $f(t)$, the CWT is obtained by convolution between the original signal and the complex conjugate of the considered mother wavelet. It is given by,

$$\psi(s, T) = \int_{-\infty}^{\infty} f(t) \bar{h}(t) dt \quad (4.18)$$

With,

$$\bar{h}(t) = \frac{1}{\sqrt{s}} h\left(\frac{t-T}{s}\right) \quad (4.19)$$

Where $\bar{h}(t)$ denotes the windowing function and is known as the mother wavelet, and s is the scale value which can dilate or compress the signal. The scale value is inversely proportional to the frequency of the signal. High scale implies low frequency and vice versa. T represents the translation parameter, which decides the time-location of the window. It is clear from the above definition that the signal transformation efficiency greatly depends upon the choice of

the mother wavelet. In general, the wavelets are chosen based on their ability to represent the signal under consideration.

Among the available mother wavelets, Haar is most straightforward in construction. It is discontinuous and represents a step function. The other family of wavelets is the Daubechies wavelets. They are used in many wavelet transform applications. The Symlets wavelets are nearly symmetrical and are modified forms of the Daubechies family. The wavelets Haar, Daubechies, Symlets, and Coiflets are compactly supported orthogonal wavelets. These, along with the Meyer wavelet are capable of perfect construction of the signal. The Meyer, Mexican hat and Morlet wavelets are perfectly symmetrical. The wavelet families and the order of wavelets considered in the present work are given in Table 7.1. The CWT spectrum for various faults using db9 wavelets is shown in Figure 7.1.

Table 7.1: List of wavelet families and order of wavelets considered

S. No	Name of Wavelet	Order of Wavelets
1	Haar	db1
2	Daubechies	db2, db3, db4, db5, db6, db7, db8, db9, db10
3	Symlets	sym2, sym3, sym4, sym5, sym6, sym7, sym8
4	Coiflets	coif1, coif2, coif3, coif4, coif5
5	Meyer	Meyr
6	Biorthogonal	bior1.1, bior1.5, bior2.2, bior2.8, bior3.3, bior3.7, bior6.8
7	Reverse Biorthogonal	rbior1.1, rbior1.5, rbior2.2, rbior2.8, rbior3.3, rbior3.7, rbior6.8
8	Gaussian	gaus1, gaus2, gaus3, gaus4, gaus5, gaus6, gaus7, gaus8
9	Morlet	Morl
10	Mexican Hat	Mexh

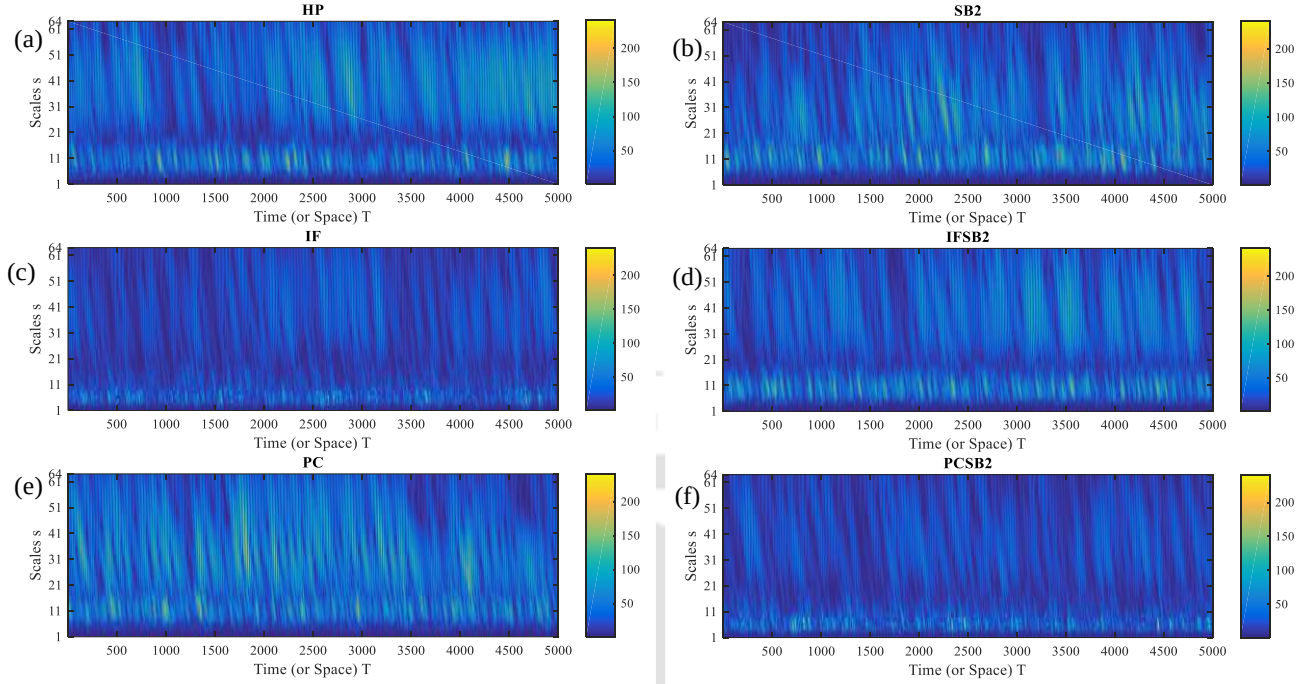


Figure 7.1: CWT spectrum - scale versus space for (a) HP, (b) SB2, (c) IF, (d) IFSB2, (e) PC, and (f) PCSB2 faults using db9 wavelet

Apart from the selection of wavelet, it is essential to select the scales 's' which are best suited for the application. For this purpose, two methods are proposed based on the energy of the wavelet and the principal component analysis. They are described in the subsequent sub-sections.

7.2.2 Scale Selection and Feature Extraction

Selection of the scale of the wavelet transform is significant to capture the needed information from the original signal. In this section, two independent methods based on energy criterion and the principal component analysis are proposed to select the best scales. The description of the methods is presented in the subsequent sections.

7.2.2.1 Energy Based Scale Selection

The energy associated with each scale can be calculated using the coefficients of CWT corresponding to that particular scale (Qi and Neupauer, 2008). In the current study, the coefficients of 64 scales are extracted. The energy of i^{th} scale is given by,

$$\varepsilon(i) = \sum_{p=1}^n |C_{i,p}|^2 \quad (4.20)$$

Where, $\varepsilon(i)$ is the energy of the i^{th} scale of the selected wavelet. $C_{i,p}$ is the coefficient value of the i^{th} scale at the p^{th} location and $p = 1, 2 \dots n$. High energy value indicates high time-frequency activity in the signal. Therefore, energy is used as a criterion for the selection of the best scales. The energy value for each of the scales is calculated, and the scales are corresponding to the highest energies are selected. Following steps are involved in the present methodology.

- Step 1:** From the raw time-domain data of the accelerometers and the line current probes, CWT coefficients are extracted for scales from 1 to 64.
- Step 2:** The energy content of each of the scales is calculated as given in Eq. (7.3) for a particular data set.
- Step 3:** The energy values calculated in Step 2 are arranged in descending order, and the top three scales are selected and frozen. These selected scales are different for each of the accelerometers (or accelerations) in x, y, z directions and each of the three line currents.
- Step 4:** Once the scales are frozen, the CWT coefficients corresponding to these scales are extracted for all the datasets.

Step 5: From these coefficients of each of the finalized scales, the statistical features are extracted.

Table 7.2 gives the scales corresponding to the top three maximum energy values for different faults. In literature, traditionally only the coefficients of one scale were selected. However, this leads to loss of a lot of important fault information at other frequencies. Therefore, in this method, the top three energies are selected. This method of scales selection is termed in this research as – CWT-BE-MSVM. This means, the CWT features are selected using the best energy (BE) criterion and the multi-class SVM (MSVM) is used for the fault segregation.

7.2.3 Principal Component Analysis (PCA) Based Feature Selection

The PCA analyzes a data table representing the observations described by several dependent variables, which are generally inter-correlated. In the previous method, the energy of the scales was used as a criterion to select them. However, it may be possible that the signal activity is distributed among the scales and therefore, a method needs to be devised for selecting the best information from the given CWT coefficients.

One of the popular methods to do a multivariate statistical analysis is the principal component analysis (PCA) which evaluates a matrix representing the data illustrated by numerous dependent variables, which may be correlated. The goal of the PCA is to extract the significant information from the matrix and to re-express this data as a set of new orthogonal variables called the principal components (Abdi and Williams, 2010). The PCA performs three essential tasks, (1) extracts most crucial information from the given matrix of data, (2) compresses the data to only retain the significant information and, (3) simplifies the description of the dataset. Traditionally, the PCA method is used for dimensionality reduction. In the present study, this

concept has been extended to the selection of useful information from all the extracted scales.

This has not been attempted in the literature before.

Table 7.2: Scales giving the top three energies for various fault conditions in x , y , z directions of the accelerometer 1 (A1x, A1y, A1z, respectively), the accelerometer 2 (A2x, A2y, A2z, respectively) and the line phases (C1, C2, C3)

	A1x	A1y	A1z	A2x	A2y	A2z	C1	C2	C3
HP	12, 59, 63	44, 30, 50	63, 34, 45	63, 52, 24	55, 54, 53	63, 62, 61	61, 62, 63	61, 62, 63	61, 62, 63
SB2	12, 13, 58	32, 46, 31	63, 32, 33	63, 51, 24	10, 48, 9	62, 63, 60	61, 62, 63	61, 62, 63	61, 62, 63
SB5	12, 58, 63	33, 35, 36	63, 34, 36	63, 50, 49	9, 10, 48	61, 62, 63	61, 62, 63	61, 62, 63	61, 62, 63
DB1	14, 13, 12	50, 46, 52	36, 60, 63	52, 51, 40	10, 48, 50	54, 63, 62	61, 62, 63	61, 62, 63	61, 62, 63
DB5	63, 12, 13	50, 46, 49	63, 34, 36	63, 50, 48	47, 48, 58	60, 62, 63	61, 62, 63	61, 62, 63	61, 62, 63
IF	38, 14, 40	12, 13, 14	63, 12, 13	37, 38, 39	63, 40, 42	61, 62, 63	61, 62, 63	61, 62, 63	61, 62, 63
IFSB2	14, 40, 46	40, 54, 58	12, 63, 11	46, 54, 38	42, 60, 61	15, 17, 36	61, 62, 63	61, 62, 63	61, 62, 63
IFSB5	44, 40, 43	63, 42, 40	4, 12, 63	40, 42, 44	63, 42, 43	63, 62, 2	61, 62, 63	61, 62, 63	61, 62, 63
IFDB1	43, 24, 23	43, 44, 42	12, 4, 11	57, 50, 58	46, 44, 45	63, 62, 60	61, 62, 63	61, 62, 63	61, 62, 63
IFDB5	15, 14, 13	44, 58, 63	63, 12, 11	57, 56, 58	63, 40, 44	63, 62, 61	61, 62, 63	61, 62, 63	61, 62, 63
PC	8, 7, 6	16, 46, 17	18, 16, 17	14, 13, 15	54, 55, 56	63, 62, 54	61, 62, 63	61, 62, 63	61, 62, 63
PCSB2	8, 7, 6	16, 46, 17	18, 16, 17	14, 15, 13	54, 55, 56	63, 14, 62	61, 62, 63	61, 62, 63	61, 62, 63
PCSB5	8, 7, 6	16, 17, 15	63, 18, 16	14, 15, 16	54, 55, 53	63, 14, 62	61, 62, 63	61, 62, 63	61, 62, 63
PCDB1	8, 7, 6	16, 17, 15	18, 17, 16	13, 14, 15	54, 55, 56	63, 14, 24	61, 62, 63	61, 62, 63	61, 62, 63
PCDB5	8, 7, 6	46, 16, 45	18, 17, 16	14, 15, 16	55, 54, 56	63, 14, 62	61, 62, 63	61, 62, 63	61, 62, 63

The steps involved in selection include,

- Step 1:** From the raw time-domain data of the accelerometers and line current probes, CWT coefficients are extracted for scales from 1 to 64.
- Step 2:** The principal component analysis is performed on this data for each accelerometer and the line current phase, separately, with each of the scale data as one direction.
- Step 3:** The top three principal components are chosen, and others are discarded.
- Step 4:** The features are extracted from these three principal components.
- Step 5:** Here the top three principal components are chosen because they constitute more than 70% of the data variance. The plots of the principal components for various faults are shown in Figure 7.2.

This method of scales selection is termed in this research as – CWT-PCA-MSVM. This means that the CWT features are selected using the principal component analysis (PCA) and the multi-class SVM (MSVM) is used for the fault segregation.

7.2.4 Feature and Mother Wavelet Selection

A systematic choice of feature(s) and the wavelet function(s) is essential to obtain the desired performance of the classifier. For the feature selection, the features are extracted from the db9 wavelet. Also, the CWT-BE-MSVM approach is used for obtaining the fault classification performances. The statistical features considered are: μ (mean), σ (the standard deviation), σ_R (reciprocal of the standard deviation), S (entropy), χ (skewness), κ (kurtosis), the 5th, 6th and 7th statistical moments (μ_5 , μ_6 , μ_7 , respectively), RMS (root mean square), σ^2 (variance), RSS (root sum of squares), SS (sum of squares), shape factor (w), impulsion index (I), crest factor

(P), tolerance index (T), power to average (PA). All these statistical features have their standard mathematical definitions. The results of the analysis are presented in Figure 7.3.

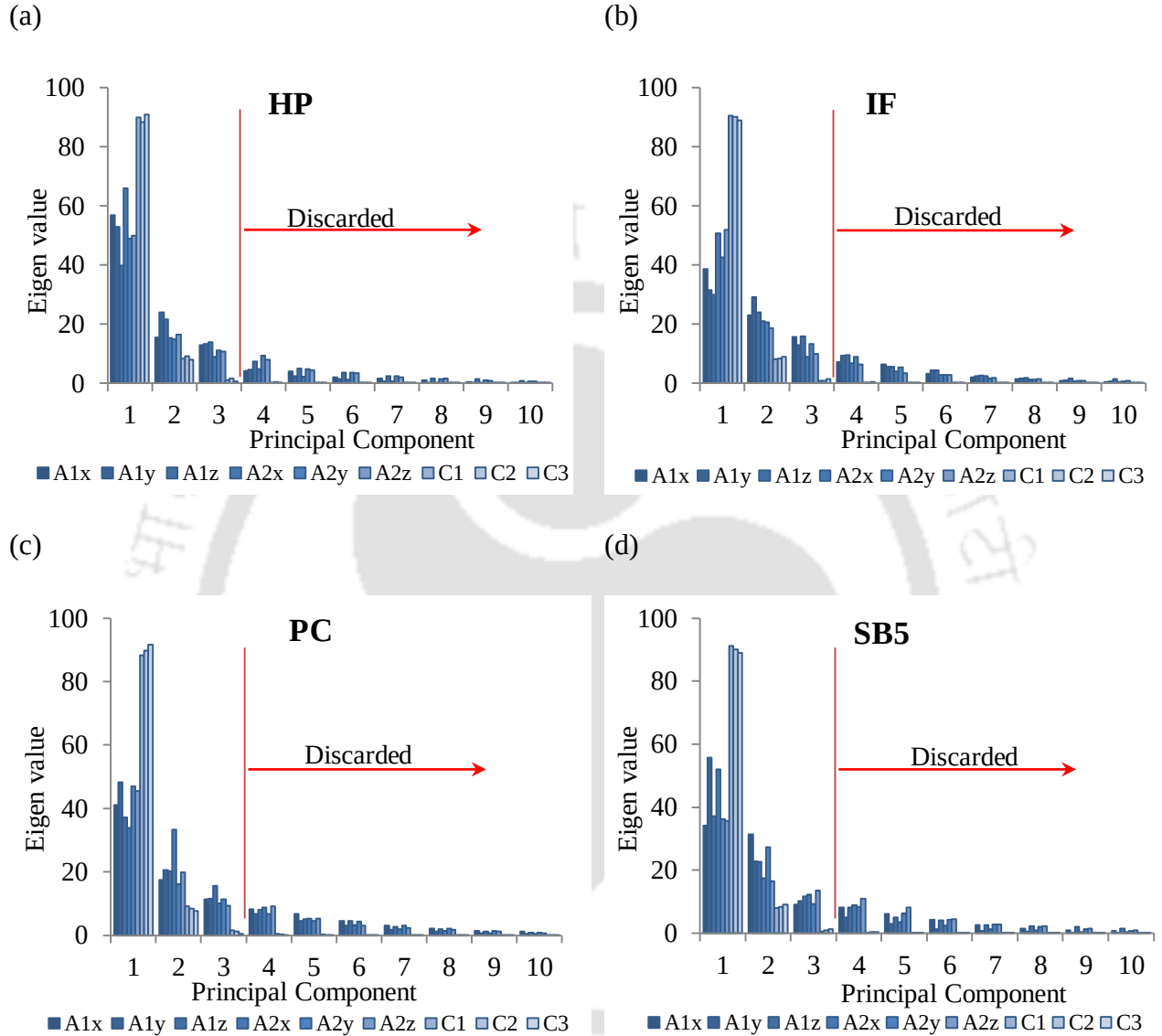


Figure 7.2: Eigenvalues versus principal component for, (a) HP, (b) IF, (c) PC and (d) SB5 faults

From the results, it can be seen that the features σ , σ_R , RMS and RSS give a classification accuracy of more than 99%. Binary combinations of these four features are used to train the classification algorithm. A combination of σ and the RSS features improved the classification

accuracy slightly. Further, upon adding a third feature to this combination, there is no significant improvement in the classification performance. Hence, this combination of features is chosen.

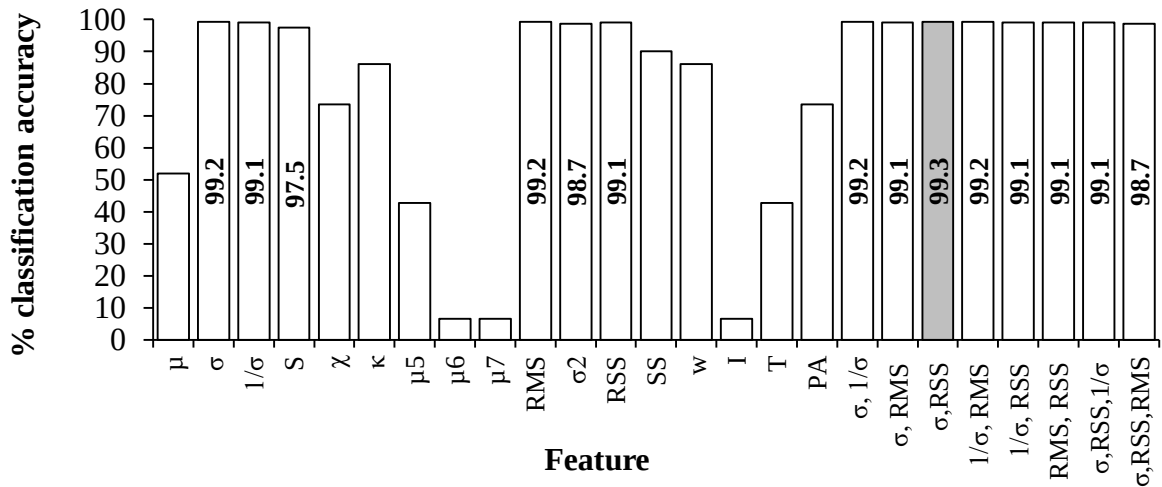


Figure 7.3: Feature selection using AF fault set and db9 wavelet

This finalized feature combination is used to select the best wavelet. The wavelet families that are considered in this work are shown in Table 7.1. The CWT-BE-MSVM methodology in conjunction with the wrapper model is used to select the wavelets while classifying the AF faults set. The classification ability of all the wavelet functions is checked at two speeds, viz. 35 Hz and 50 Hz. A threshold of 99.5% accuracy is set on the average classification accuracy obtained from both the speeds mentioned above. The wavelets giving an accuracy of over 99.5% are selected. Further, these wavelets are trained and tested at all the speeds (30 Hz to 65 Hz in steps of 5 Hz) and whichever wavelet gives an accuracy of over 99.5% is finalized. The comparison of classification performance of all the wavelets presented in Table 7.1 is given in Figure 7.4. Figure 7.4(a) gives the classification performance of the bi-orthogonal family of wavelets. Of all the wavelets considered in this family, bior1.5, bior3.3 and bior6.8

wavelets gave an average classification of above 99.5% and therefore, these three wavelets are selected for further study. Similarly, from Figure 7.4(b) to Figure 7.4(g), it can be inferred that rbior1.1 and rbior2.2 from the reverse bi-orthogonal family, coif3 from the Coiflets family, Haar wavelets (db1), db8, db9 from the Daubechies family, gaus1 to gaus6 in the Gaussian wavelet family, Morlet (morl), sym4, sym7, sym8 from the Symlets family have classification accuracy equal to or above the threshold limit of 99.5% and therefore, these wavelets have been selected.

The classification ability of these selected wavelets is further investigated at all the operating speeds of the CP. It can be seen from the overall classification accuracies that not all the wavelets maintain their performance at each of the speeds. Therefore, only those wavelets that give an accuracy of above 99.5% are finalized. The finalized wavelets are gaus1, gaus2, gaus5, morl, and bior3.3. These finalized wavelets would be further used in the subsequent analysis in this research. Appendix F gives the CWT plots of all faults using gaus1 wavelet.

7.3 SVM Based Fault Prediction Performance

In this section, results of the multi-class classification of various faults are presented. First, different categories of fault classifications are given followed by, the significance of the considered fault classification category and their results of classification. The fault classification cases considered are given in Table 7.3.

7.3.1 Case A: Same Speed Classification

In this case, in continuation to the results presented in Chapters 5 and 6, the fault classifications are attempted using training and testing of the data at the same speed first.

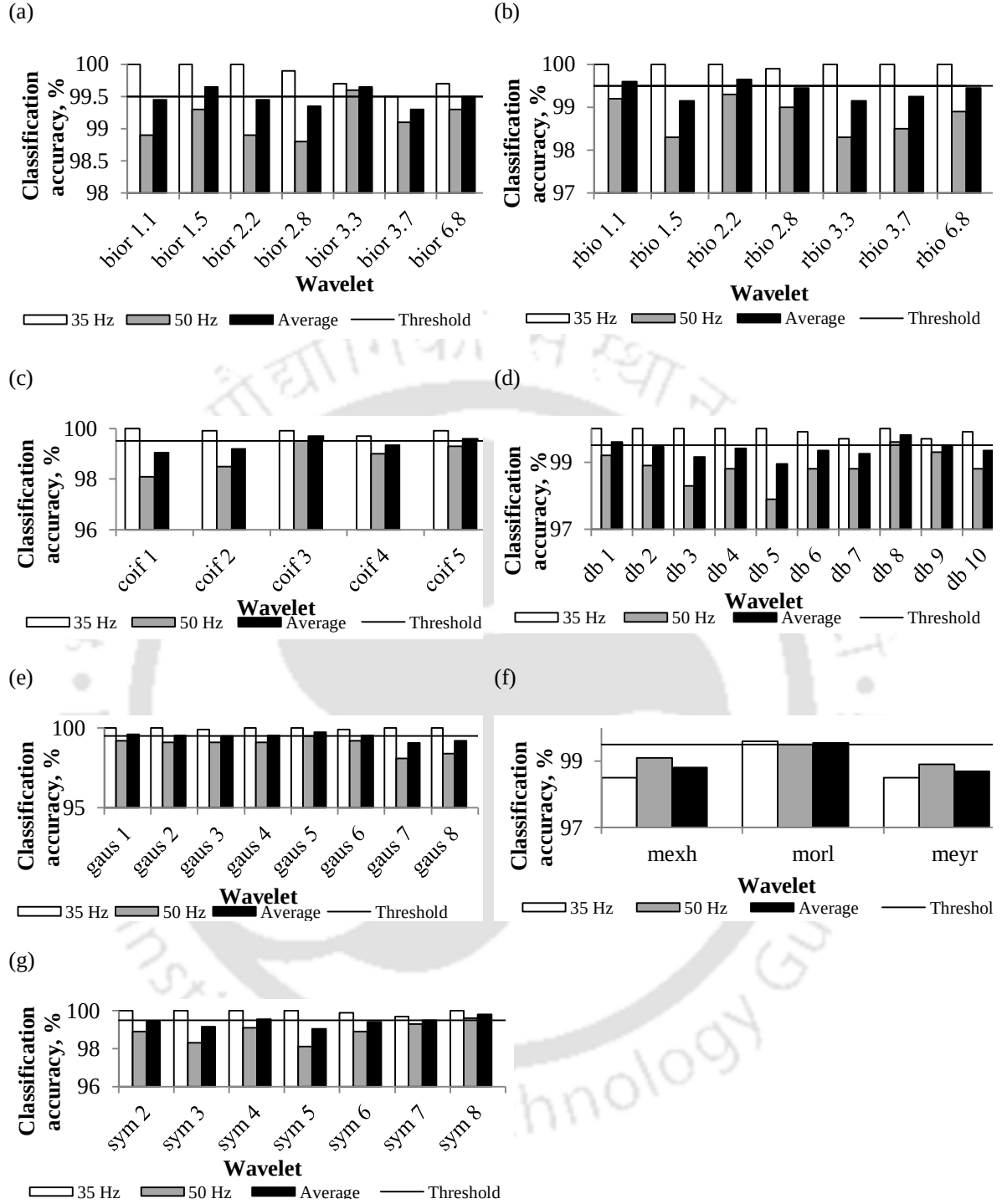


Figure 7.4: The Classification accuracies of various wavelet families at 35 Hz, 50 Hz, (a) bi-orthogonal family, (b) reverse bi-orthogonal family, (c) coiflet, (d) Daubechies, (e) Gaussian, (f) Mexican hat, Morlet, Meyer and (g) symlets. The horizontal line on the figure represents the threshold limit of 99.5% classification accuracy set for each wavelet.

The faults belonging to different fault families including the blockage fault family (BF), the impeller faults and the blockage fault family (IFBF), the dry run fault family and a combination of all the faults are attempted to be classified.

Table 7.3: Fault classification cases considered

Case	Faults for training	Faults for testing	Training speed (Hz)	Testing speed
A	BF, HP	BF, HP	30 – 65	Same as training
	IFBF, HP	IFBF, HP		
	PCBF, HP	PCBF, HP		
	Dry runs, HP	Dry runs, HP		
B	AF	AF	30 – 65	Intermediate speed
	AF	AF		
C	AF	AF	30 – 65	Noisy signal at same training speed

Based on the discussion presented in the previous section, the finalized mother wavelets are gaus1, gaus2, gaus5, morl, and bior3.3. Therefore, the sequence of steps followed in this segment is,

1. Using CWT-BE-MSVM approach and the finalized mother wavelets, the fault classification of each of the fault families is performed, and the overall fault classification accuracy is estimated.
2. Whichever mother wavelet shows the maximum overall classification accuracy is selected for further investigation.

3. The finalized mother wavelet is used to perform the fault diagnosis using the CWT-BE-MSVM and the CWT-PCA-MSVM. Further, the performances of both these methodologies are compared.

7.3.1.1 BF fault set classification – identification of blockage fault severity

In this classification, the objective would be to identify the severity of the suction or discharge blockage in a CP. As blockages can be progressive, it is crucial to identify the severity of the fault in its formative stages itself so that the corrective maintenance may be immediately initiated. Figure 7.5 shows the performance of the finalized wavelets in segregating the BF faults. The highest overall classification accuracy is demonstrated by the morl wavelet, which would, therefore, be used to compare both the developed methodologies. The results of the classification are presented in Table 7.4. The overall classification accuracy obtained using both the methods are comparable. At 50 Hz CP speed, the classification accuracy of the DB4 fault is very low thereby decreasing the average performance of the classifiers. The CWT-BE-MSVM method gives an accuracy of more than 99.5% at 35, 40, 45 and 55 Hz. Also, the CWT-PCA-MSVM gives almost a 100% classification at 45 Hz of CP test speed.

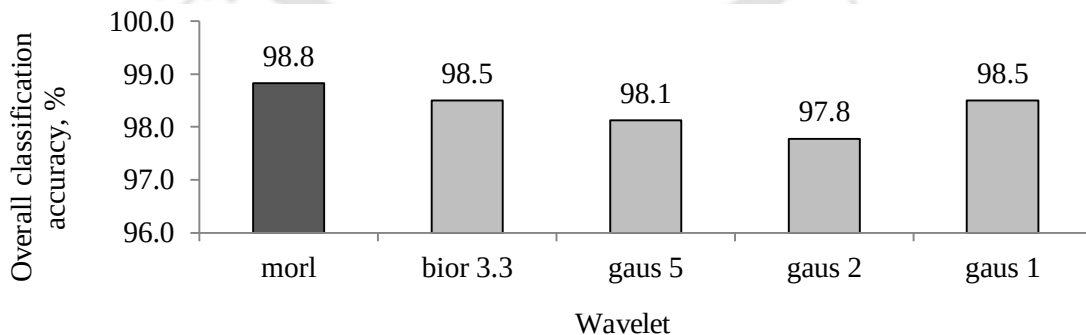


Figure 7.5: Overall classification accuracies of the selected wavelets for BF faults case

Table 7.4: BF fault classification using morl mother wavelet

Train/ Test Speed (Hz)	Method	HP	SB1	SB2	SB3	SB4	DB1	DB2	DB3	DB4	DB5	C_a
30 Hz	CWT-BE-MSVM	98	92	98	98	100	100	100	100	100	96	98.2
	CWT-PCA-MSVM	100	98	98	100	94	100	100	98	100	96	98.4
35 Hz	CWT-BE-MSVM	100	100	100	100	98	100	100	100	100	98	99.6
	CWT-PCA-MSVM	100	100	100	98	100	100	100	96	100	92	98.6
40 Hz	CWT-BE-MSVM	100	100	98	100	100	100	100	100	100	96	99.4
	CWT-PCA-MSVM	92	94	98	100	98	100	100	100	100	96	97.8
45 Hz	CWT-BE-MSVM	100	96	100	100	100	100	100	100	100	100	99.6
	CWT-PCA-MSVM	98	100	100	100	100	100	100	100	100	100	99.8
50 Hz	CWT-BE-MSVM	98	94	100	100	100	100	100	96	82	88	95.8
	CWT-PCA-MSVM	98	92	100	100	100	100	100	100	90	94	97.4
55 Hz	CWT-BE-MSVM	100	98	100	100	100	100	100	100	100	100	99.8
	CWT-PCA-MSVM	100	100	100	100	100	100	100	96	94	96	98.6
60 Hz	CWT-BE-MSVM	96	96	100	100	100	100	100	100	100	100	99.2
	CWT-PCA-MSVM	94	100	100	100	100	100	94	100	100	100	98.8
65 Hz	CWT-BE-MSVM	100	98	100	100	100	94	100	100	98	100	99
	CWT-PCA-MSVM	100	100	100	100	100	100	100	96	100	98	99.4
Overall	CWT-BE-MSVM											98.8
	CWT-PCA-MSVM											98.6

7.3.1.2 IFBF fault set classification – identification of severity of impeller defects in combination with blockage faults

In this classification, the objective is to identify the severity of the suction blockage or the discharge blockage in a CP with impeller defects. This is important because the co-existence

of the mechanical and the hydraulic faults on the CP can accelerate the failure rate and thus, must be curbed immediately.

Figure 7.6 shows the overall classification obtained using the finalized wavelets. bior3.3 wavelet is chosen for the comparison of both the developed methodologies as it shows the highest overall classification accuracy. Table 7.5 gives the comparison of the classification performances for both the methods. The HP and the IFSB4 faults give a 100% classification accuracy using any of the two methods at all the test speeds. The CWT-BE-MSVM method gives a classification accuracy of over 99% at the speeds 55, 60 and 65 Hz. Also, the overall classification accuracy of the CWT-BE-MSVM method is 98.9%, while that of the CWT-PCA-MSVM method is approximately 1% lesser. In general, the classification of the IFDB faults seems to show comparatively lower performance than the other faults. However, the fault severity using both the methods could be identified with satisfactory accuracy.

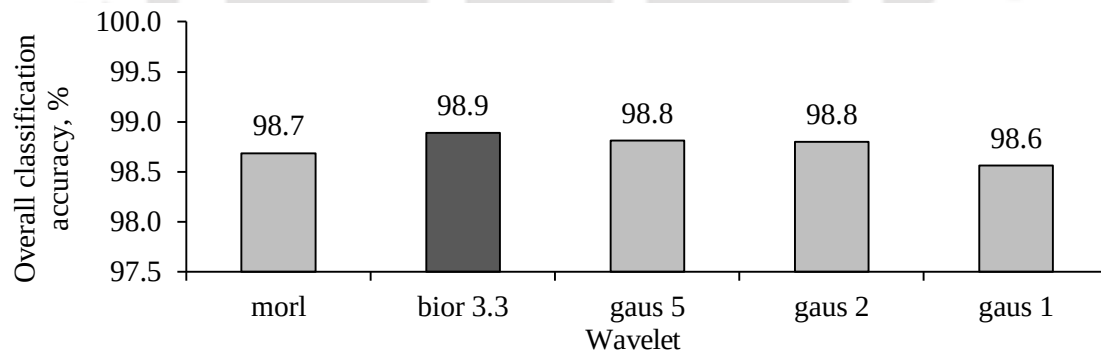


Figure 7.6: Overall classification accuracies of the selected wavelets for IFBF faults case

7.3.1.3 PCBF fault set classification – identification of severity of the pitted cover plate defects in combination with the blockage faults

The objective of this classification would be to identify the severity of the faults belonging to the PCBF family.

Table 7.5: IFBF fault classification using the bior3.3 mother wavelet

Train/ Test Speed (Hz)	Method	HP	IF	IFSB1	IFSB2	IFSB3	IFSB4	IFDB1	IFDB2	IFDB3	IFDB4	IFDB5	C_a
30 Hz	CWT-BE-MSVM	100	100	100	100	100	100	100	98	96	92	96	98.4
	CWT-PCA-MSVM	100	100	100	100	98	100	100	100	96	96	98	98.9
35 Hz	CWT-BE-MSVM	100	100	100	100	100	100	100	100	96	94	98	98.9
	CWT-PCA-MSVM	100	100	100	100	100	100	100	100	98	100	86	98.5
40 Hz	CWT-BE-MSVM	100	100	100	98	100	100	100	100	92	98	100	98.9
	CWT-PCA-MSVM	100	100	100	100	100	100	100	94	96	84	88	96.5
45 Hz	CWT-BE-MSVM	100	100	96	98	100	100	100	98	94	96	100	98.4
	CWT-PCA-MSVM	100	100	98	100	98	100	100	90	94	100	98	98.0
50 Hz	CWT-BE-MSVM	100	100	100	100	100	100	100	100	100	100	98	99.8
	CWT-PCA-MSVM	100	100	100	100	100	100	100	98	100	98	88	98.5
55 Hz	CWT-BE-MSVM	100	100	98	100	100	100	98	96	96	98	98	98.5
	CWT-PCA-MSVM	100	98	100	100	100	100	100	100	100	92	84	97.6
60 Hz	CWT-BE-MSVM	100	98	98	100	100	100	100	100	100	96	98	99.1
	CWT-PCA-MSVM	100	98	100	100	100	100	100	100	100	92	84	97.6
65 Hz	CWT-BE-MSVM	100	98	96	100	100	100	100	98	98	100	100	99.1
	CWT-PCA-MSVM	100	100	100	100	100	100	80	94	100	92	96	96.5
Overall	CWT-BE-MSVM												98.9
	CWT-PCA-MSVM												97.8

Figure 7.7 shows the overall prediction performance shown by the different finalized wavelets. In this case, both bior3.3 and gaus1 wavelets are showing the same overall classification accuracy. gaus1 wavelet is used for the comparison of the classification performance of both the methods. The results of the classification are given in Table 7.6. From the table, it can be seen that the HP and the PCSB4 faults give a 100% classification across all the speeds using any of the two methods. Also, both the methods show a 100% fault prediction at 45 Hz test speed. The classifier shows the least performance at 30 Hz speed with difficulty in the classification of the PCDB faults. The overall classification accuracies

obtained using the CWT-BE-MSVM and the CWT-PCA-MSVM are 98.4% and 97.2%, respectively.

All the finalized wavelets show classification of more than 99.6%. However, here also, the gaus1 wavelet gives the highest overall classification accuracy of 100%. This implies that there is 100% fault segregation of the dry run faults at all the test speeds. This may be attributed to the significant difference between the fault manifestations and thus the feature values for the different dry run. The gaus1 wavelet would be used to compare the classification performance of both the developed methodologies. The results of this classification are shown in Table 7.7.

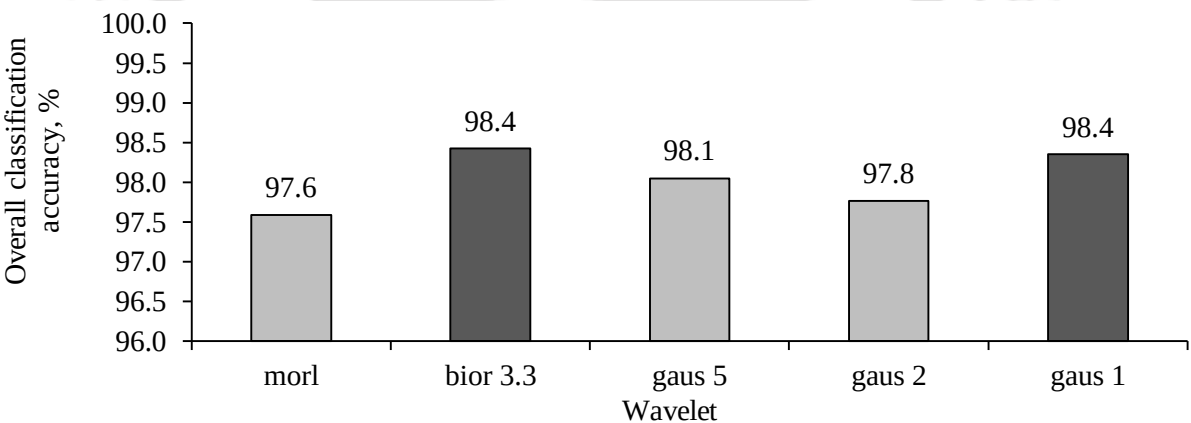


Figure 7.7: Overall classification accuracies of the selected wavelets for the PCBF faults case

7.3.1.4 Dry run fault set classification

Figure 7.8 shows the classification ability of the finalized wavelets to segregate the faults belonging to the DR fault set using the CWT-BE-MSVM method.

Table 7.6: PCBF fault classification using the gaus1 mother wavelet

Train/ Test Speed (Hz)	Method	HP	PC	PCSB1	PCSB2	PCSB3	PCSB4	PCDB1	PCDB2	PCDB3	PCDB4	PCDB5	C_a
30 Hz	CWT-BE-MSVM	100	100	98	100	98	100	96	100	98	82	86	96.2
	CWT-PCA-MSVM	100	100	98	100	92	100	86	100	84	80	74	92.2
35 Hz	CWT-BE-MSVM	100	100	98	98	100	100	94	98	100	100	100	98.9
	CWT-PCA-MSVM	100	100	98	96	100	100	88	94	98	100	100	97.6
40 Hz	CWT-BE-MSVM	100	100	100	96	100	100	98	100	94	98	100	98.7
	CWT-PCA-MSVM	100	96	98	100	100	100	98	100	98	100	100	99.1
45 Hz	CWT-BE-MSVM	100	100	100	100	100	100	100	100	100	100	100	100.0
	CWT-PCA-MSVM	100	100	100	100	100	100	100	100	100	100	100	100.0
50 Hz	CWT-BE-MSVM	100	100	92	98	100	100	94	92	100	100	100	97.8
	CWT-PCA-MSVM	100	100	94	100	100	100	94	98	86	100	100	97.5
55 Hz	CWT-BE-MSVM	100	96	96	98	100	100	98	90	96	100	100	97.6
	CWT-PCA-MSVM	100	98	84	88	100	100	92	98	96	100	100	96.0
60 Hz	CWT-BE-MSVM	100	98	98	100	100	100	96	98	100	100	100	99.1
	CWT-PCA-MSVM	100	100	98	100	100	100	100	100	100	100	100	99.8
65 Hz	CWT-BE-MSVM	100	92	94	98	100	100	100	100	100	100	100	98.5
	CWT-PCA-MSVM	100	88	68	90	100	100	100	100	100	100	100	95.1
Overall	CWT-BE-MSVM												98.4
	CWT-PCA-MSVM												97.2

The CWT-PCA-MSVM shows an overall classification accuracy of only 99.3% while the other method gives a 100% classification accuracy. At 45 Hz, the PCSB5 fault gave a classification accuracy of only 84% using the CWT-PCA-MSVM, which brought down the classification performance of the method significantly.

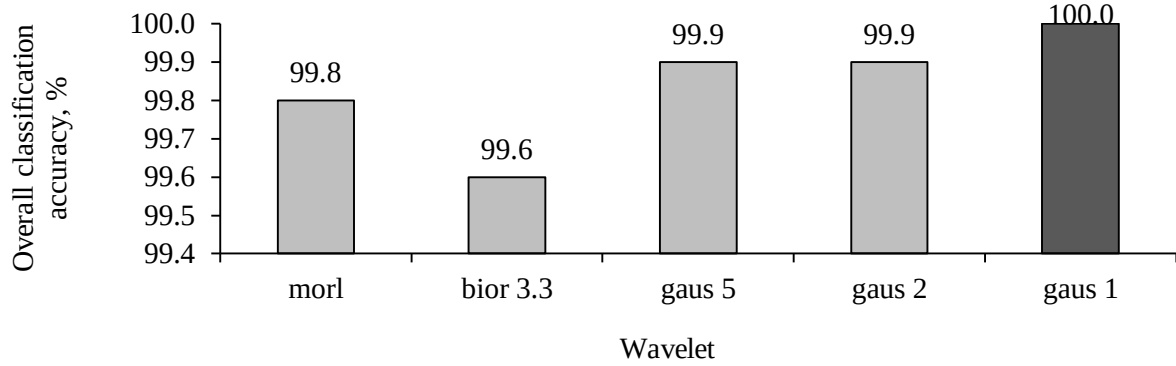


Figure 7.8: Overall classification accuracies of the selected wavelets for the dry run faults case.

7.3.1.5 AF fault set classification

Figure 7.4 shows the classification ability of the finalized wavelets to segregate the faults belonging to the AF fault set. Of all the finalized wavelets, the gaus1 wavelet shows a slightly better classification performance, and so, this wavelet would be used to compare the performance of the two developed methodologies, namely, the CWT-BE-MSVM and the CWT-PCA-MSVM. The results of the classification using the two methods are presented in Table 7.8. From the table, it can be seen that both the methodologies gave satisfactory classification performance at all the tested speeds.

In general, the classification with both the methodologies gives the same results approximately. However, at 65 Hz speed, there is a considerable drop in the classification performance with the CWT-PCA-MSVM. This is because the PC and the PCSB2 faults are showing very less classification accuracies (88% and 78% respectively) at that speed. The dry run faults (SB5, IFSB5, and PCSB5) and the DB1, IFDB1 faults show a 100% classification accuracy at all the test speeds. At the speeds of 35 Hz, 45 Hz, and 55 Hz, almost 100% classification of all the faults using any of the two methodologies is obtained. The overall

classification accuracies of the CWT-BE-MSVM and the CWT-PCA-MSVM using the gaus1 wavelet are found to be 99.6% and 99.5% respectively.

Table 7.7: Dry run fault classification using the gaus1 mother wavelet

Train/ Test Speed (Hz)	Method	HP	SB5	IFSB5	PCSB5	C_a
30 Hz	CWT-BE-MSVM	100	100	100	100	100
	CWT-PCA-MSVM	100	100	98	100	99.5
35 Hz	CWT-BE-MSVM	100	100	100	100	100
	CWT-PCA-MSVM	100	100	100	100	100
40 Hz	CWT-BE-MSVM	100	100	100	100	100
	CWT-PCA-MSVM	98	100	100	100	99.5
45 Hz	CWT-BE-MSVM	100	100	100	100	100
	CWT-PCA-MSVM	100	100	98	84	95.5
50 Hz	CWT-BE-MSVM	100	100	100	100	100
	CWT-PCA-MSVM	100	100	100	100	100
55 Hz	CWT-BE-MSVM	100	100	100	100	100
	CWT-PCA-MSVM	100	100	100	100	100
60 Hz	CWT-BE-MSVM	100	100	100	100	100
	CWT-PCA-MSVM	100	100	98	100	99.5
65 Hz	CWT-BE-MSVM	100	100	100	100	100
	CWT-PCA-MSVM	100	100	100	100	100
Overall	CWT-BE-MSVM					100
	CWT-PCA-MSVM					99.3

This implies that both the methodologies serve as useful tools to identify the type of the fault in the CP. It is interesting to note that even in the previous fault classification results presented for the dry run fault set, the dry run faults got segregated with 100% accuracy independent of the test speed. Compared to the dry run and the AF fault classifications, the classification of

the BF, IFBF and PCBF faults show relatively lower percentage of classification accuracy. This may be attributed to the complexity involved in identifying the faults with close severity levels as they tend to give similar vibration patterns and thus similar features.

7.3.2 Case B: Intermediate Speed Fault Classification

To estimate the ability of the algorithm to classify faults when fault history data is unavailable, a classification is attempted at the intermediate speed. In this case, the algorithm is trained with the fault data at the two operating CP speeds and is tested with data at an intermediate CP speed. To check the ability of the algorithm to distinguish various faults, the AF fault family is attempted to be classified. The results of the fault classification are presented in Table 7.9. In the table, the first column gives the details of the training and the testing speeds, where 30, 40/ 35 stands for training with the data at 30 Hz and 40 Hz and training with the data at 35 Hz. The reason for doing a time-frequency analysis study is to improve the performance of the classifier. When the Case A classification results from Chapter 6 and Chapter 7 are compared, it is clear that the classification accuracies improved significantly by doing a CWT transformation. However, while trying to classify the faults at an intermediate speed, the CWT transformation, and the developed methodology does not perform at par with their counterparts in the time-domain or the frequency domain. This is because the CWT coefficients are shifting their values with the varying speeds. Hence, the CWT transformation is not suitable for the CP fault classification at intermediate speeds.

Table 7.8: AF fault classification using the guas1 mother wavelet

Train/ Test Speed (Hz)	Method	HP	SB2	SB5	DB1	DB5	IF	IFSB2	IFSB5	IFDB1	IFDB5	PC	PCSB2	PCSB5	PCDB1	PCDB5	C_a
30 Hz	CWT-BE-MSVM	100	100	100	100	98	100	100	100	100	100	100	100	100	96	96	99.3
	CWT-PCA-MSVM	94	100	100	100	96	100	100	100	100	100	100	100	100	100	100	99.3
35 Hz	CWT-BE-MSVM	100	100	100	100	100	100	100	100	100	100	98	100	100	100	100	99.9
	CWT-PCA-MSVM	100	100	100	100	100	100	98	100	100	100	100	100	100	100	100	99.9
40 Hz	CWT-BE-MSVM	100	98	100	100	98	100	100	100	100	100	100	98	100	98	100	99.5
	CWT-PCA-MSVM	100	98	100	100	98	100	100	100	100	100	100	100	100	98	100	99.6
45 Hz	CWT-BE-MSVM	100	100	100	100	100	100	100	100	100	98	100	100	100	100	100	99.9
	CWT-PCA-MSVM	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
50 Hz	CWT-BE-MSVM	94	98	100	100	98	100	100	100	100	100	100	100	100	98	100	99.2
	CWT-PCA-MSVM	100	100	100	100	100	100	100	100	100	100	100	100	100	98	100	99.9
55 Hz	CWT-BE-MSVM	100	100	100	100	100	98	100	100	100	100	98	98	100	100	100	99.6
	CWT-PCA-MSVM	100	100	100	100	98	100	100	100	100	100	100	100	100	100	100	99.9
60 Hz	CWT-BE-MSVM	100	100	100	100	100	100	100	100	100	100	98	100	100	100	100	99.9
	CWT-PCA-MSVM	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
65 Hz	CWT-BE-MSVM	100	100	100	100	100	100	100	100	100	100	98	98	100	100	100	99.7
	CWT-PCA-MSVM	100	100	100	100	100	100	100	100	100	100	88	78	100	100	100	97.7
Overall	CWT-BE-MSVM																99.6
	CWT-PCA-MSVM																99.5

Table 7.9: Case B: the AF fault classification performance at intermediate speeds

Train/ Test Speed (Hz)	% fault classification accuracy															C_a
	HP	SB2	SB5	DB1	DB5	IF	IFSB2	IFSB5	IFDB1	IFDB5	PC	PCSB2	PCSB5	PCDB1	PCDB5	
30,40/35	32	16	100	100	70	68	100	100	84	100	66	100	100	76	90	80.2
40,50/45	10	92	92	74	52	10	100	100	98	92	66	100	100	100	100	79.1
50,60/55	10	86	100	100	10	10	2	92	96	10	40	60	100	68	100	58.9

7.3.3 Case C: Fault Identification with Noisy Test Data at the Same Test Speed

To check the robustness of the algorithm in handling noisy data, here also the corrupted time-domain data is used to extract the CWT coefficients. These coefficients are further processed, and statistical features are extracted. These features are used to test the algorithm, which is already trained with the baseline data. The algorithm is tested for the AF fault family classification using the gauss1 mother wavelet and the CWT-BE-MSVM approach. The algorithm is trained with the baseline data and tested with 5% AWGN (noise), 10% AWGN and 25% AWGN data, independently. 5% AWGN case gives an overall classification of 100%. To get a bird's eye view, the results of the classifications are compared with that of the no-noise condition and are shown in Figure 7.9. It is interesting to observe that, with the addition of white Gaussian noise, the classification accuracy has not dropped significantly. The overall classification accuracies obtained for 5% AWGN, 10% AWGN and 25% AWGN are 100%, 99.9%, and 99.4%, respectively. From the results shown for the baseline and the corrupt data classification, it seems as if the addition of noise has improved the classification accuracy. However, from an engineer's perspective, there is not much difference between 99.5% and 100% classification accuracy.

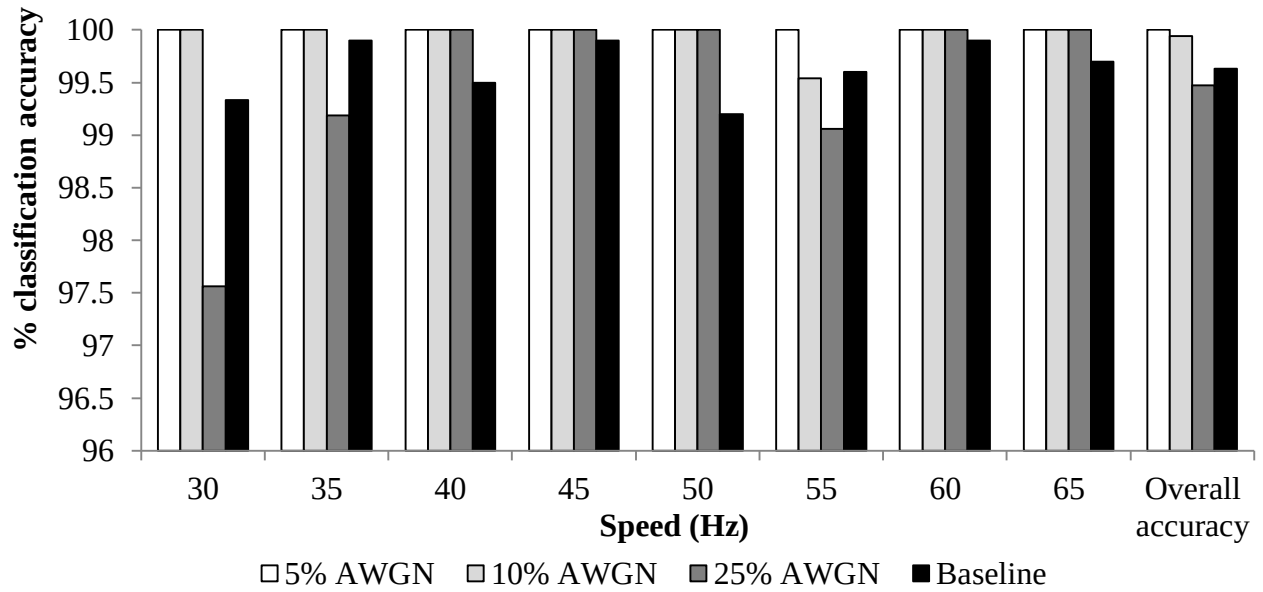


Figure 7.9: Comparison of the fault classification accuracies obtained for the AF fault classification using 5%, 10%, 25% AWGN and the baseline data

The standard deviation feature plot for the baseline, 5% AWGN, 10% AWGN and 25% AWGN is as shown in Figure 7.10.

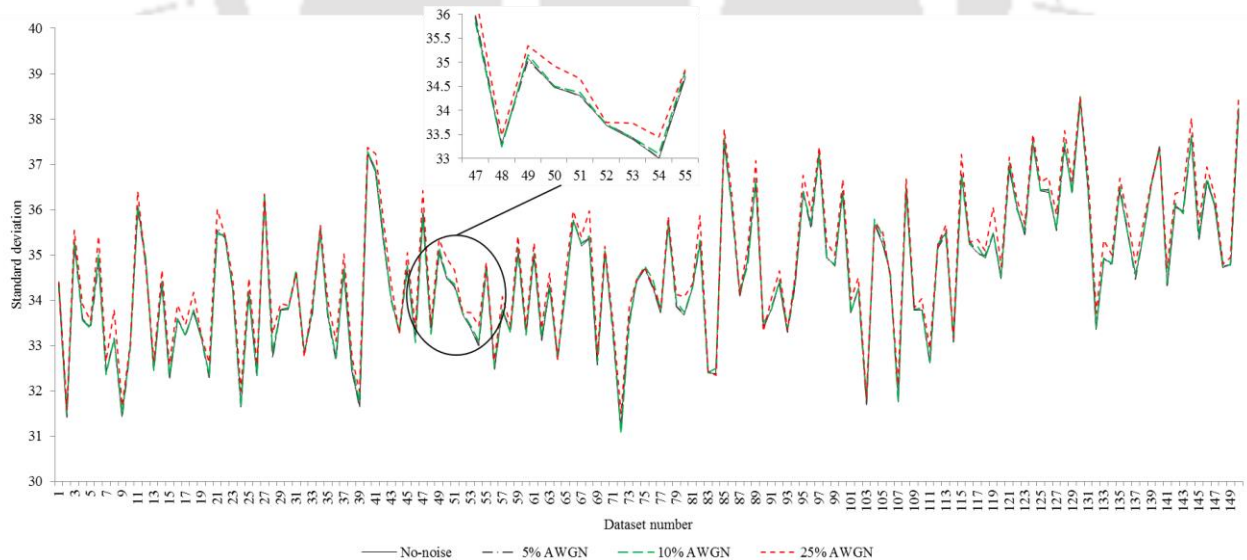


Figure 7.10: Comparison of the standard deviation feature for no-noise, 5% AWGN, 10% AWGN, 25% AWGN for the HP fault at 30 Hz speed

It can be observed that the addition of noise of up to 10% does not shift the feature values much. However, with the addition of 25% AWGN noise, there is a perceivable amount of shift in the feature values. Nevertheless, the selected features are versatile and do not differ considerably with the addition of noise. Thus, it explains the high fault classification percentage.

7.4 Summary

Based on the results of the frequency-domain analysis given in Chapter 6, it was evident that there was some loss of information when simple frequency domain features were used. Hence, it was concluded that a time-frequency analysis approach might help in solving this problem. As the next step, a time-frequency transformation using continuous wavelet transforms was recommended. In this chapter, the application of the features extracted from the continuous wavelet transformation to perform the CP fault classification was presented. Two methodologies based on the energy approach and the principal component analysis were proposed to select the best scales, and the wrapper model was used to select the best features.

Of both the methodologies proposed, the energy method performed slightly better than its counterpart. However, since only three principal components were considered, there was almost 30% loss of information. Probably considering more principal components would improve the classification performance of the PCA based methodology as well. However, such a consideration would make the methodology computationally expensive. As expected, the classification results clearly show that there is an improvement in the performance of the classifier when compared to that of the frequency domain by using the CWT based features for the same speed fault classifications. However, when the algorithm was checked for the

intermediate speed, the classification performance was not on par with the time-domain for the frequency-domain based features which was attributed to the shift in the features with the changing speeds. The algorithm could very easily handle the time-domain data corruption with up to 25% additive white Gaussian noise without a significant drop in the classification accuracy which was because the features were insensitive to the noise addition.

Though CWT analysis seems useful, its inability to detect the faults at an intermediate speed is not clearly understood. Hence, in the subsequent chapter, the features extracted from another variant of the wavelet transform called the wavelet packet transform (WPT) is attempted to be used. The CWT has perfect local property in the time-space and the frequency space. However, to split the high-frequency band where modulation information of machine faults is expected to exist, WPT may be useful. In the next chapter a comparison of wavelet based methodologies to that of works done in literature is also presented.

CHAPTER 8: Multi-Fault Diagnosis of Centrifugal Pumps using Wavelet Packet Transform Analysis

8.1 Introduction

In the previous chapter, the multi-fault diagnosis of the CPs was attempted using the continuous wavelet transform approach which was a time-frequency domain analysis based on wavelet transforms. In this chapter, another wavelet transform technique called the wavelet packet transform (WPT) is used to extract the useful features to solve the CP multi-fault diagnosis problem.

Wavelet packets are particular linear combinations of wavelets. They form the bases which retain properties such as the orthogonality, smoothness, and localization of the parent wavelets. Coefficients in the linear combinations are calculated using a recursive algorithm, making each newly computed sequence a root of its own analysis tree.

The improved signal decomposition capability of the WPT in high-frequency ranges makes it an appealing alternative for detecting faults which have transient components in high-frequency ranges. Thus, WPT is an excellent tool for analyzing time-varying signals as the same frequency bandwidth can provide required resolution in both high and low-frequency ranges.

In the previous chapter, it was found that the feature characteristics were shifting with the operating speed of the CP. Thus, CWT based features were not suitable for intermediate speed fault classifications. However, it cannot be generalized that any time-frequency based

approach has the same limitation. Therefore, in this chapter another variant of the wavelet transform, the WPT is used. The WPT is applied to all the extracted time-domain signals from the CP (vibrations as well as motor line current), and the wavelet packet coefficients are extracted. The decomposition is done up to two levels. Later, the best nodes are selected using two methods, the best energy (BE) approach and the principal component analysis (PCA) approach. The best suited statistical features are selected using the wrapper model. Various parent wavelets are examined, and their performances are compared. By using the finalized feature(s) and wavelet(s) combinations, the algorithm is trained and tested for various fault classification cases at the same speed and intermediate speeds. Also, the robustness of the algorithm is found out by testing the already trained SVM algorithm with corrupted data. To understand the operating speed dependency of the algorithm, it is trained and tested at a range of CP operating speeds (as the fault manifestations change very intricately with the CP speed).

8.2 Fault Feature Extraction

In this section, the details of the wavelet packet transform (WPT), the methods applied for node selection and the feature selection approach are explained.

8.2.1 Wavelet Packet Transform

In this section, a brief description of the wavelet packet transform is presented. The detailed account on the WPT can be seen in refs. (Walczak and Massart, 1997; Wu and Liu, 2009). Projection of a time domain signal onto wavelet basis functions is called 'wavelet transform.' It is a time-frequency signal analysis method. In some aspects, it is similar to Fourier transform. That is, in Fourier transform sine and cosine functions are used as the basis functions, these functions are localized in the frequency domain but are global in the time

domain. However, in the wavelet transform, functions localized in both time and frequency domains called the ‘wavelets’ are used. The wavelet packet function $W_{p,q}^n$ is defined as,

$$W_{p,q}^n(t) = 2^{\frac{p}{2}} W^n(2^p t - q) \quad (5.1)$$

Where, the integers p and q are the index scale and translation operations, t is the time. The index n is an operation modulation parameter or oscillation parameter. First, two wavelet packet functions are the scaling and mother wavelet functions, given as,

$$W_{0,0}^0 = \phi(t) ; W_{0,0}^1 = \psi(t) \quad (5.2)$$

When $n = 2, 3, \dots$, the functions can be defined by the following equations,

$$\left. \begin{aligned} W_{0,0}^{2n}(t) &= \sqrt{2} \sum_q h(q) W_{1,q}^n(2t - q) \\ W_{0,0}^{2n+1}(t) &= \sqrt{2} \sum_q g(q) W_{1,q}^n(2t - q) \end{aligned} \right\} \quad (5.3)$$

Where, $h(q)$ and $g(q)$ are the quadrature mirror filters associated with the predefined scaling function and mother wavelet function. After the filters are defined, recursive decomposition is applied to the signal, to obtain its hierarchical multi-resolution representation. The choice of the mother wavelet and the corresponding scaling function determines how intricately the signal details are captured in the time-frequency domain.

Each decomposition layer is given a ‘level’ number, with original signal at level 0. For an i^{th} level decomposition, there are 2^i bases. From the data acquisition sampling rate mentioned in Chapter 2, it can be derived that the Nyquist frequency of the current study is 2500 Hz. The decomposition of the signal in the current study is depicted in Figure 8.1. With the increase in

the level of decomposition, the signal resolution improves, but the signal energy gets more distributed. Hence, in this study, the signal is decomposed up to level 2, i.e., with 4 bases.

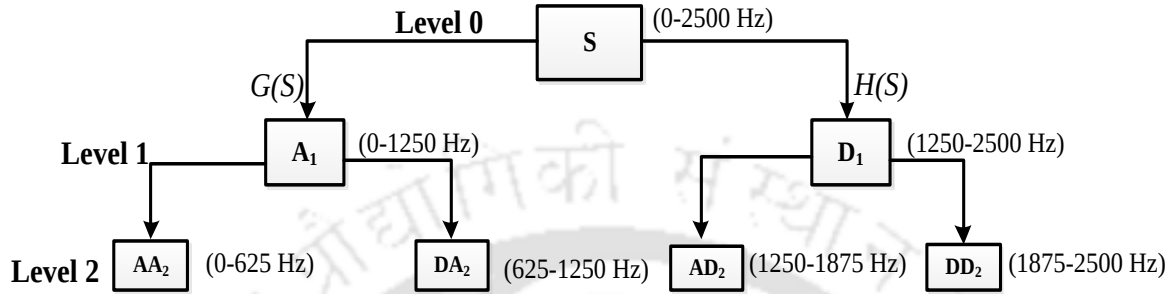


Figure 8.1: WPT decomposition

In this research, an attempt is made to investigate the wavelets that are the most suitable for the CP fault classification. A total of seven wavelet families are considered, and their details are tabulated in Table 8.1.

8.2.2 Scale Selection and Feature Extraction

The selection of the best bases from the wavelet decomposition is crucial to capture the right fault information and to eliminate the unwanted information. In this section, two bases selection methods based on node energy and principal component analysis are presented for the selection of the best bases.

8.2.2.1 Energy Based Scale Selection

One of the important physical variables in the signal analysis is the signal energy. This signal energy is distributed among decomposed WPT bases, based on the time-frequency activity of the fault. It is vital to eliminate low energy, less pertinent information and capture the fault characteristics from the signals.

Table 8.1: Families and order of wavelets considered in the present work

No	Name of Wavelet	Order of Wavelets
1	Haar	db1
2	Daubechies	db2, db3, db4, db5, db6, db7, db8, db9, db10
3	Symlets	sym2, sym3, sym4, sym5, sym6, sym7, sym8
4	Coiflets	coif1, coif2, coif3, coif4, coif5
5	Discrete approximation of Meyer	dmey
6	Biorthogonal	bior1.1, bior1.5, bior2.2, bior2.8, bior3.3, bior3.7, bior6.8
7	Reverse Biorthogonal	rbior1.1, rbior1.5, rbior2.2, rbior2.8, rbior3.3, rbior3.7, rbior6.8

The following steps are used to select the best bases using energy criterion,

- Step 1:** The fault signal in time-domain is decomposed using WPT up to level 2, using different wavelets.
- Step 2:** The energy of each of the bases (nodes) is calculated from the wavelet coefficients.
- Step 3:** Top n energy levels are kept, and their corresponding basis coefficients are stacked in the form of a vector.
- Step 4:** Statistical features are extracted from this vector.
- Step 5:** Accuracy based method (wrapper model) is used to finalize the features as well as the value of n .

The combination of the db9 wavelet and AF fault set is chosen to carry out the steps mentioned above, and finalize the features and the value of n . The results of the analysis are given in Table 8.2.

Table 8.2: The feature and energy level selection for WPT-BE-MSVM

% Classification accuracy					
Feature	Energy 1	Energy 2	Energy 3	Energy 4	Average
μ	97.1	98.2	99.2	99.4	98.5
Σ	95.6	96.4	98.7	98.5	97.3
$1/\sigma$	45.9	16.1	98.9	98.9	65.0
S	97.1	97.2	96.4	97.6	97.1
X	51.2	71.5	76.8	79.3	69.7
K	40.8	56.4	67.3	67.3	58.0
μ_5	65.9	73.7	66.8	60.2	66.7
μ_6	61.5	66.2	10.7	6.7	36.3
μ_7	46.5	56.3	6.7	6.7	29.1
RMS	98.4	97.7	99.5	98.7	98.6
σ^2	95.1	96.3	98.5	98.7	97.2
RSS	97.9	97.7	99.5	98.7	98.5
SS	95.8	97.1	94.4	94	95.3
w	43.5	18.2	10.7	6.7	19.8
I	51.2	66.2	16.1	40.8	43.6
P	56.3	61.5	30.2	6.7	38.7
T	76.8	60.2	16.1	10.7	40.9
PA	79.3	59.2	40.8	33.3	53.2

In the table, Energy 1 corresponds to $n = 1$, Energy 2 corresponds to $n = 2$, and so on. That means when $n = 4$, the coefficients of all the four bases of the level 2 decomposition are stacked, and no data is discarded. The percentage classification accuracy obtained for different statistical feature and n combinations are presented in the table. Statistical features considered are: μ (mean), σ (standard deviation), $1/\sigma$ (reciprocal of standard deviation), S (entropy), χ (skewness), κ (kurtosis), 5th, 6th and 7th statistical moments (μ_5 , μ_6 , μ_7 , respectively), RMS (root mean square), σ^2 (variance), RSS (root sum of squares), SS (sum of squares), shape

factor (w), impulsion index (I), crest factor (P), tolerance index (T), and power to average (PA).

In Table 8.2, it can be observed that the features such as μ , σ , σ^2 , e , RMS, RSS, and SS give the best average classification. With these features, Energy 3 gives the best prediction performance. To select the best feature combinations, the performance of different binary permutations and combinations of these features are done and are tabulated in Table 8.3. The classifications are attempted at 30, 50 and 65 Hz speed. From the table, it is evident that at all the speeds the binary combinations of μ , σ , RMS and S are giving accuracies of more than 99.5%. Hence, a combination of these features is finalized for this method. This methodology is termed as best energy wavelet packet transform multiclass support vector machine (WPT-BE-MSVM). A feature scatters plot of the μ , σ , RMS, and S features for various CP faults are shown in Figure 8.2. In the figure, the three axes are μ , σ and S and the variation in RMS are color coded. From the figure, it can be observed that all the faults are forming separate clusters on the plot.

8.2.3 Principal Component Analysis (PCA) Based Feature Selection

The PCA computes the most meaningful basis to re-express garbled data. In this case, the number of bases (four) is the dimension of the data. One of the strong assumptions of PCA is that it considers the significant variations in data to have essential dynamics attached to it. Hence, the order of principal components determines the order of decreasing covariance.

Table 8.3: Binary feature permutations and combinations performance for WPT-BE-MSVM

Feature	Combination feature	35 Hz	50 Hz	65 Hz	Average
μ	σ	99	100	99.7	99.6
	e	99.3	99.8	99.8	99.6
	RMS	99.2	99.7	99.7	99.5
	σ^2	98.2	99	99	98.7
	RSS	98.5	99.5	99.3	99.1
	SS	91.2	6.7	6.7	34.9
Σ	e	99.2	100	99.7	99.6
	RMS	99.2	99.8	99.7	99.6
	σ^2	98	99	98.4	98.5
	RSS	99	100	99.3	99.4
	SS	91.2	6.7	61.7	34.9
	RMS	99.5	100	99.7	99.7
S	σ^2	98.4	99.2	99	98.9
	RSS	99	100	99.3	99.4
	SS	91.2	6.7	6.7	34.9
	σ^2	98.2	99.2	99	98.8
RMS	RSS	99	100	99.3	99.4
	SS	91.2	6.7	6.7	34.9
	RSS	99	100	99.2	99.4
σ^2	SS	91.2	6.7	6.7	34.9
	RSS	91.2	6.7	6.7	34.9

The PCA method for the basis selection is carried out in following steps,

- Step 1:** The fault signal in time-domain is decomposed using WPT up to level 2, using different wavelets.
- Step 2:** The principal component scores are calculated for 4 decomposed bases.
- Step 3:** The first two principal components are retained.

Step 4: Statistical features are extracted for each of these two principal components.

Step 5: Accuracy based method (wrapper model) is used to finalize the features.

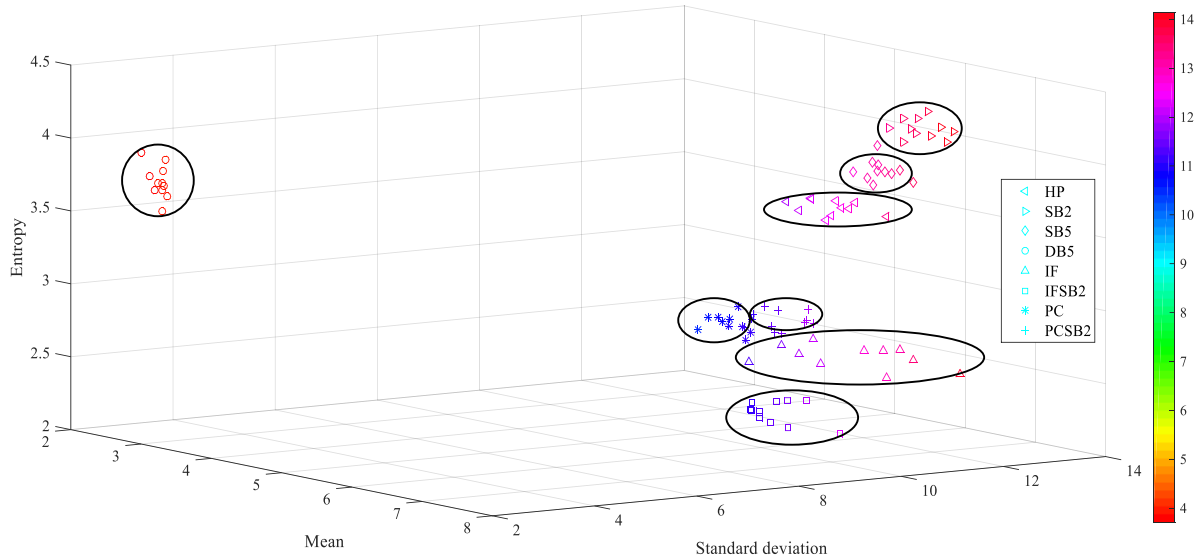


Figure 8.2: Feature scatter plot for HP, SB2, SB5, DB5, IF, IFSB2, PC and PCSB2 faults. Plot showing 4-D of mean (x-axis), standard deviation (y-axis), entropy (z-axis) and RMS (colour bar)

Here also db9 wavelet and AF fault set are used for the feature selection. The results of this selection are tabulated in Table 8.4. It can be seen that features σ and RSS give classification accuracy of more than 99.5%. Hence, their combination is finalized for classification. This methodology is termed as PCA wavelet packet transform multiclass support vector machine (WPT-PCA-MSVM). Therefore, the desired features for WPT-BE-MSVM are μ , σ , RMS, and e , and that for WPT-PCA-MSVM are σ and RSS. In the next section, the suitable wavelet is selected based on the maximum classification accuracy it demonstrates.

8.2.4 Feature and Mother Wavelet Selection

After feature selection, the next significant step is to identify the best-suited mother wavelet(s). Like in Chapter 7, in this chapter also AF fault family is used for the wavelet selection.

Table 8.4: Feature selection for the WPT-PCA-MSVM

Feature	35 Hz	50 Hz	65 Hz	Average
σ	99.3	99.7	99.8	99.6
S	95.5	95.8	93.8	94.9
σ^2	98.9	99.3	99.7	99.3
RSS	99.7	99.7	99.8	99.7

This multi-class classification set is used to study the performance of various wavelet families described in Table 8.1. The WPT-BE-MSVM methodology is used to demonstrate this performance. Since the fault manifestation changes with operating speed (Wolfram et al., 2001b), the performance of the wavelet is checked at a low speed (35 Hz), medium speed (50 Hz) and high speed (65 Hz). The *average* classification performance of the wavelet at these three speeds is said to describe its capability in the CP fault classification.

Figure 8.3 shows the classification performance of Haar and Daubechies family of wavelets. It can be seen that all the wavelets show an average classification of more than 99%. However, db6 and db8 wavelets show classification of close to 100%. A similar exercise was done for other wavelet families as well. The average classification shown by them is presented in Figure 8.4. From the figure, it can be seen that sym4, sym7, sym8, rbio1.5, rbio2.8, and rbio6.8 give a near 100% fault classification. db6, sym4, and rbior6.8 perform at a similar level.

Moreover, from the analysis presented in Chapter 7, it was clear that all the good performing wavelets almost have a similar trend of classification. Hence, for simplicity, in the subsequent studies, db6 wavelet would be employed for all fault classifications. The WPT spectrum for all faults using db6 wavelets at 30 Hz speed is presented in Appendix G.

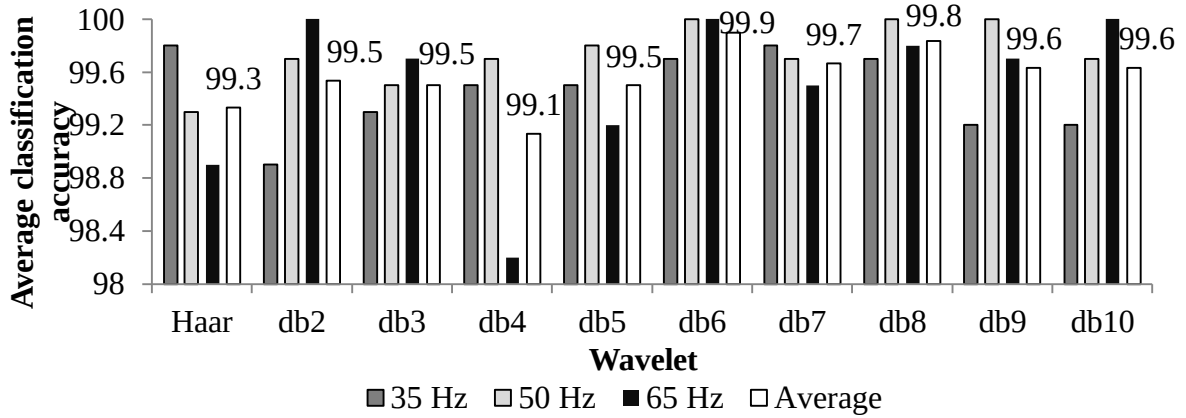


Figure 8.3: Fault classification ability of different wavelets of Haar and Daubechies family at 35, 50, and 65 Hz

8.3 SVM Based Fault Prediction Performance

In this section, the results of classification of various fault families are presented. First, different categories of fault classifications are given. Later, the significance of the considered fault classification category is illustrated, and their results of classification are presented. The fault classification cases considered are given in Table 8.5. Case A considers the fault classification of BF, IFBF, PCBF, Dry run and AF fault families when the algorithm is trained and tested at the same speed. Case B and Case C only consider the AF fault family and their fault classification at an intermediate speed and with noisy data, respectively.

As mentioned in Section 8.2.2, there are two fault classification methodologies defined for the selection of the most representative coefficients and thereby the features of the WPT

decomposition. These methods are WPT-BE-MSVM and WPT-PCA-MSVM. It must be noted that both these methods are independent. The flowchart of the methodology used in the present work is given in Figure 8.5. As a recapitulation, BE stands for the best energy, PCA stands for the principal component analysis and MSVM stands for the multiclass support vector machine.

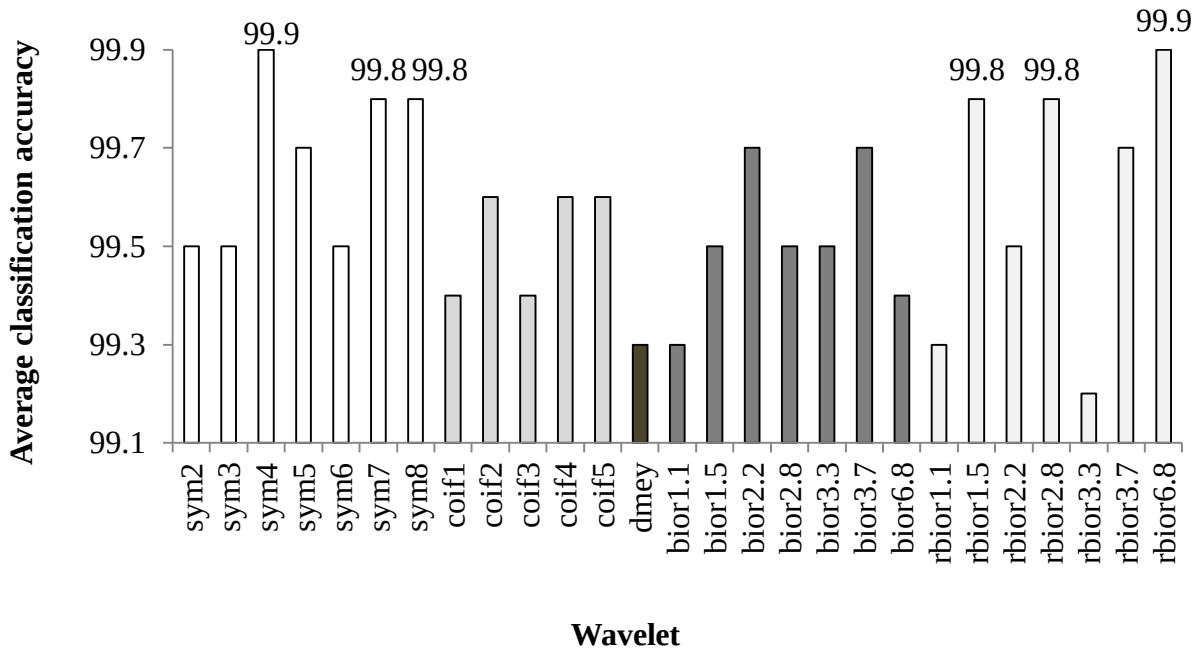


Figure 8.4: Average fault classification ability of different wavelets belonging to Symlets, Coiflets, Dmey, biorthogonal and reverse biorthogonal families

8.3.1 Case A: Same Speed Classification

This is the most frequently occurring case of fault identification. According to this case, the operating condition at which fault history data is available is the same as that of the current operating condition of the machine. The classification is aimed at identifying different faults as well as their severities. The finalized wavelet (db6) is used for the fault classifications. The

classification accuracies obtained using WPT-BE-MSVM and WPT-PCA-MSVM is compared.

Table 8.5: Fault classification cases considered

Case	Faults for training	Faults for testing	Training speed (Hz)	Testing speed
A	BF, HP	BF, HP	30 – 65	Same as training
	IFBF, HP	IFBF, HP		
	PCBF, HP	PCBF, HP		
	Dry runs, HP	Dry runs, HP		
B	AF	AF	30 – 65	Intermediate speed
	AF	AF		
C	AF	AF	30 – 65	Noisy signal at same training speed

8.3.1.1 BF fault set classification – identification of blockage fault severity

This classification would assist a plant maintenance engineer to answer two crucial questions, (1) the location of flow restriction and (2) the amount of flow restriction. This is important to schedule the system maintenance depending on the severity of the fault. The results of the fault classification are given in Table 8.6. At both low and high CP speeds DB3, DB4 and DB5 classes have shown relatively lower classification which may be because of the overlapping signature they produce at these speeds. On the other hand, faults such as SB3, SB4, DB1, and DB2 were identified at all the CP speeds with 100% accuracy.

In Figure 8.6, the average classification accuracy at each speed for the two methodologies WPT-BE-MSVM and WPT-PCA-MSVM is compared and presented. It can be observed from the figure that the WPT-BE-MSVM has a better overall classification accuracy than its

counterpart. However, at lower CP speeds (35, 40 Hz) the WPT-PCA-MSVM gave a better prediction performance. At 45 and 50 Hz speeds, the WPT-BE-MSVM outperformed the WPT-PCA-MSVM; this may be because of the loss of information from the eliminated principal components.

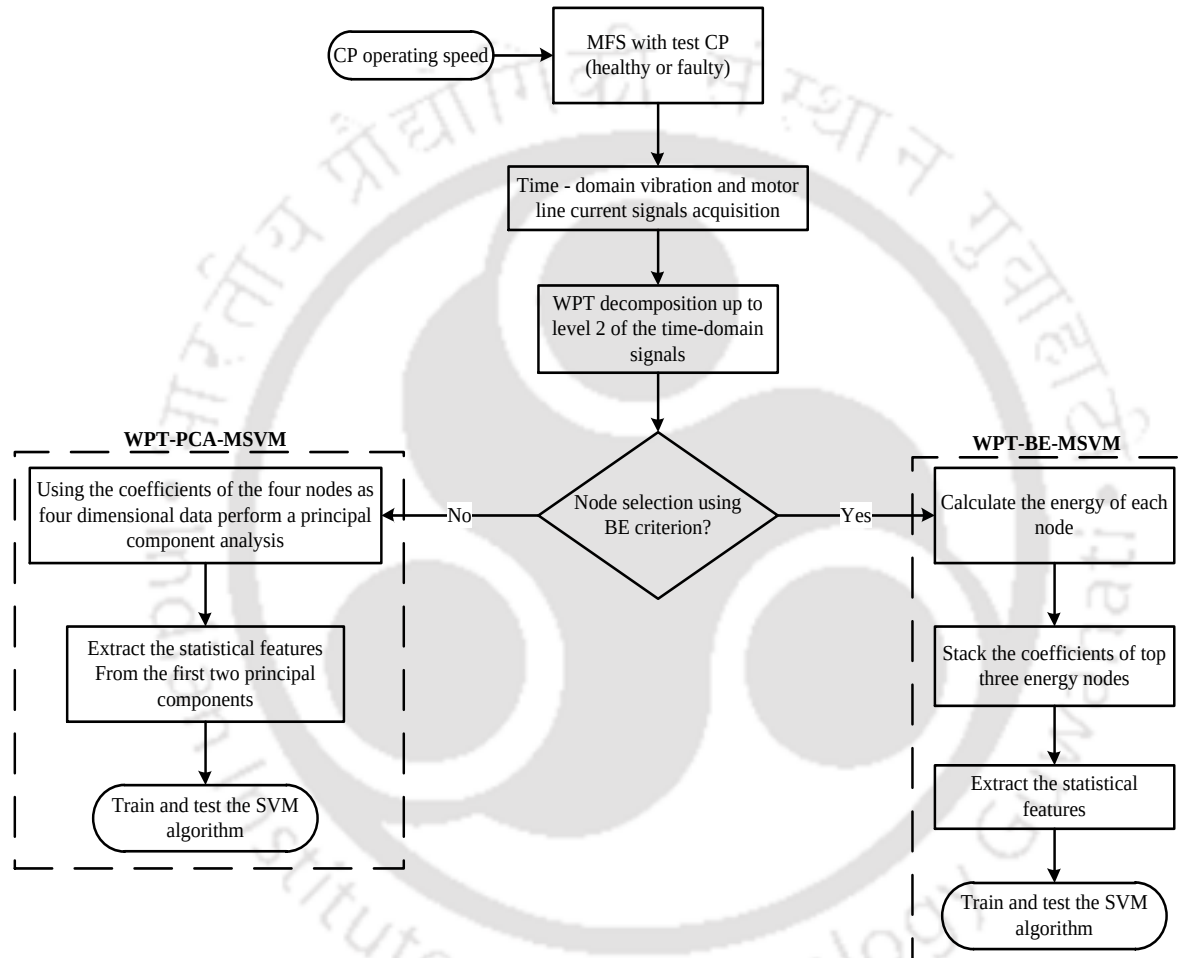


Figure 8.5: WPT based fault classification methodology

The overall classification accuracy of WPT-BE-MSVM and WPT-PCA-MSVM methodologies is 98.5 and 97.7%, respectively. The WPT-BE-MSVM works best at 50 Hz speed with an efficiency of 99.8%, and the WPT-PCA-MSVM works best at 35 and 40 Hz speed with an average accuracy of 99.3%.

Table 8.6: The BF fault family classification using WPT-BE-MSVM and WPT-PCA-MSVM

Train and Test Speeds (Hz)	Method	HP	SB1	SB2	SB3	SB4	DB1	DB2	DB3	DB4	DB5	C_a
30 Hz	WPT-BE-MSVM	100	95.1	100	100	100	100	100	95.1	97.6	100	98.8
	WPT-PCA-MSVM	97.6	100	100	100	97.6	100	100	95.1	97.6	92.7	98
35 Hz	WPT-BE-MSVM	92.7	100	97.6	100	100	100	100	95.1	100	100	98.5
	WPT-PCA-MSVM	97.6	100	100	100	100	100	100	97.6	100	97.6	99.3
40 Hz	WPT-BE-MSVM	100	97.6	100	100	100	100	100	95.1	100	100	99.3
	WPT-PCA-MSVM	92	94	98	100	98	100	100	100	100	96	97.8
45 Hz	WPT-BE-MSVM	97.6	100	100	100	100	100	100	97.6	95.1	100	99
	WPT-PCA-MSVM	95.1	97.6	100	100	100	100	100	82.9	95.1	92.7	96.3
50 Hz	WPT-BE-MSVM	97.6	100	100	100	100	100	100	100	100	100	99.8
	WPT-PCA-MSVM	92.7	100	100	100	100	97.6	100	87.8	95.1	97.6	97.1
55 Hz	WPT-BE-MSVM	100	100	100	100	100	100	100	100	97.6	90.2	98.8
	WPT-PCA-MSVM	97.6	100	100	100	100	95.1	95.1	100	100	95.1	98.3
60 Hz	WPT-BE-MSVM	100	100	100	100	100	100	100	100	80.5	92.7	97.3
	WPT-PCA-MSVM	100	100	100	100	100	90.2	97.6	97.6	82.9	87.8	95.6
65 Hz	WPT-BE-MSVM	100	100	100	100	100	97.6	100	100	85.4	90.2	97.3
	WPT-PCA-MSVM	100	100	100	100	97.6	100	100	100	92.7	90.2	98
Overall	WPT-BE-MSVM											98.5
	WPT-PCA-MSVM											97.7

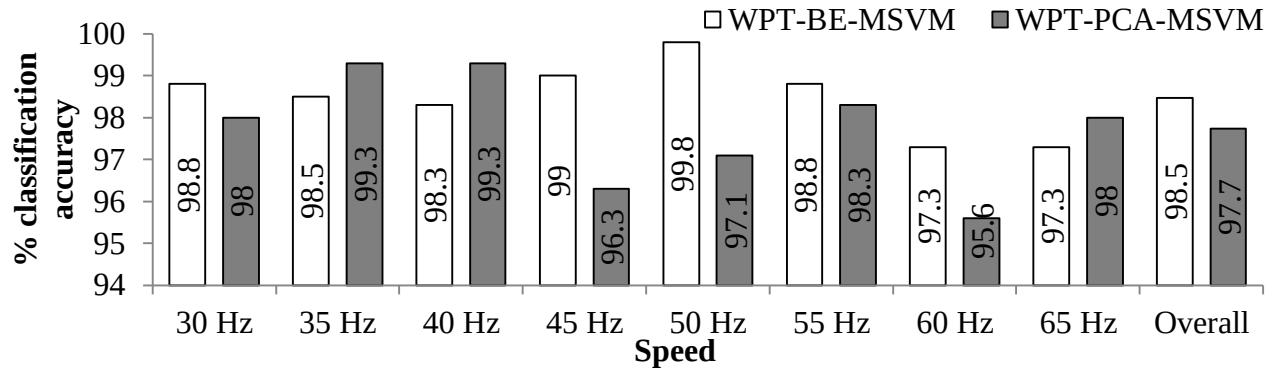


Figure 8.6: Percentage classification accuracy versus CP operating speed. Algorithm trained and tested at the same speed. The performance comparison of WPT-BE-MSVM and WPT-PCA-MSVM for BF fault case.

8.3.1.2 IFBF fault set classification – identification of severity of impeller defects in combination with blockage faults

In this segment, the objective is to classify varying severity of faults belonging to IFBF fault family which includes impeller defect fault existing independently and also in distinct combinations with the suction and discharge blockages. As explained in Chapter 2, the coexistence of mechanical and hydraulic faults on the CP may accelerate its failure. Thus, it is essential to identify them at their low severity itself to curb its progression. The results of fault classification are presented in Table 8.7. Also, the comparison of both the developed methods is presented in Figure 8.7.

In this case, as it can be seen, both the fault classification methodologies perform at the same level, giving an overall classification accuracy of 97%, approximately. Similar to the BF fault family study, in this study also at lower speeds (30, 35 Hz) the WPT-PCA-MSVM based method is better, and at higher speeds (60, 65 Hz) the WPT-BE-MSVM method is better. HP, IFSB3, and IFSB4 are accurately classified at all the test speeds of the CP. At 65 Hz, the

IFDB1 and IFDB2 faults have overlapping feature representation and thus show a low classifiability. Compared to the BF fault family classification the IFBF fault family segregation seems more difficult due to the complex fluid-structure interaction manifestations due to the presence of mechanical faults.

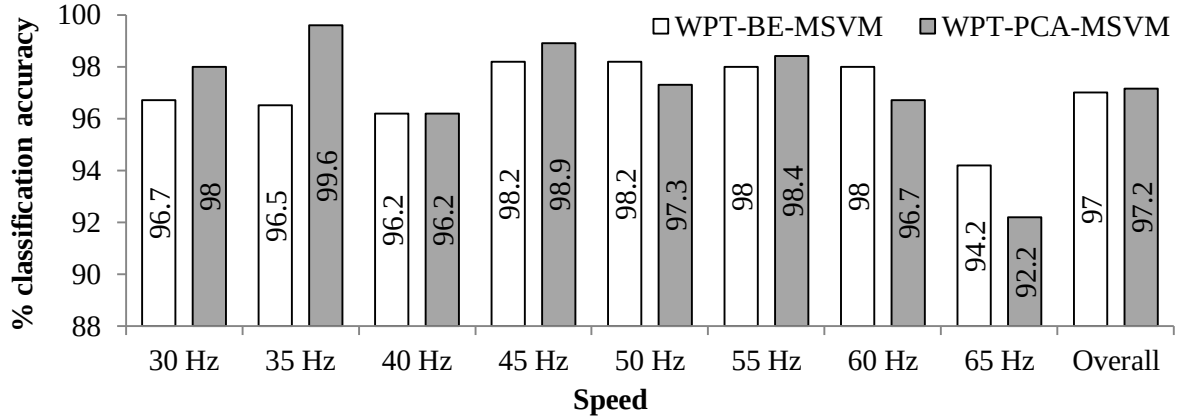


Figure 8.7: Percentage classification accuracy versus CP operating speed. Algorithm trained and tested at the same speed. Comparison of performance of WPT-BE-MSVM and WPT-PCA-MSVM for IFBF fault case

8.3.1.3 PCBF fault set classification – identification of severity of pitted cover plate defects in combination with blockage faults

In this segment, the faults belonging to PCBF family are segregated. The PCBF fault family constitutes the pitted cover plate faults and also their combination with suction/ discharge blockages. The results of the classification are given in Table 8.8 and Figure 8.8. At all the speeds the WPT-BE-MSVM method performed better (with an overall classification accuracy of 96.3%) than the WPT-PCA-MSVM (with an overall classification accuracy of 94.4%). The HP and PCSB4 faults are classified with 100% accuracy at all the speeds of the CP. The PCDB faults of various severities were not classified with high accuracies owing to the closeness of the features they generate.

Table 8.7: IFBF fault family classification using WPT-BE-MSVM and WPT-PCA-MSVM

Train and Test Speeds (Hz)	Method	HP	IF	IFSB1	IFSB2	IFSB3	IFSB4	IFDB1	IFDB2	IFDB3	IFDB4	IFDB5	C_a
30 Hz	WPT-BE-MSVM	100	100	100	100	100	100	100	100	85.4	87.8	90.2	96.7
	WPT-PCA-MSVM	100	100	100	100	100	100	100	100	85.4	97.6	95.1	98
35 Hz	WPT-BE-MSVM	100	100	100	100	100	100	100	97.6	100	82.9	80.5	96.5
	WPT-PCA-MSVM	100	100	100	97.6	100	100	100	100	100	100	97.6	99.6
40 Hz	WPT-BE-MSVM	100	97.6	100	100	100	100	100	100	95.1	87.8	78	96.2
	WPT-PCA-MSVM	100	100	100	100	100	97.6	100	100	95.1	87.8	78.1	96.2
45 Hz	WPT-BE-MSVM	100	100	100	97.6	100	100	100	97.6	95.1	95.1	95.1	98.2
	WPT-PCA-MSVM	100	100	100	95.1	100	100	100	100	97.6	97.6	97.6	98.9
50 Hz	WPT-BE-MSVM	100	100	100	100	100	100	100	97.6	90.2	100	92.7	98.2
	WPT-PCA-MSVM	100	100	100	100	100	100	100	97.6	85.4	97.6	90.2	97.3
55 Hz	WPT-BE-MSVM	100	100	100	100	100	100	100	100	100	87.8	90.2	98
	WPT-PCA-MSVM	100	100	100	100	100	100	92.7	100	97.6	97.6	95.1	98.4
60 Hz	WPT-BE-MSVM	100	100	97.6	100	100	100	97.6	92.7	100	97.6	92.7	98
	WPT-PCA-MSVM	100	100	97.6	100	92.7	95.1	92.7	90.2	97.6	100	97.6	96.7
65 Hz	WPT-BE-MSVM	100	97.6	100	100	100	100	68.3	73.2	100	97.6	100	94.2
	WPT-PCA-MSVM	100	97.6	97.6	100	100	100	58.5	61	100	100	100	92.2
Overall	WPT-BE-MSVM												97
	WPT-PCA-MSVM												97.2

Table 8.8: PCBF fault family classification using WPT-BE-MSVM and WPT-PCA-MSVM

Train and Test Speeds (Hz)	Method	HP	PC	PCSB1	PCSB2	PCSB3	PCSB4	PCDB1	PCDB2	PCDB3	PCDB4	PCDB5	C_a
30 Hz	WPT-BE-MSVM	100	97.6	100	97.6	92.7	100	100	92.7	75.6	80.5	68.3	91.4
	WPT-PCA-MSVM	100	100	92.7	92.7	85.4	97.6	97.6	90.2	82.9	61	61	87.4
35 Hz	WPT-BE-MSVM	100	92.7	95.1	95.1	100	100	92.7	87.8	97.6	87.8	100	95.3
	WPT-PCA-MSVM	100	100	90.2	100	100	97.6	92.7	90.2	85.4	90.2	100	95.1
40 Hz	WPT-BE-MSVM	100	95.1	95.1	95.1	100	100	95.1	90.2	95.1	90.2	100	96
	WPT-PCA-MSVM	100	85.4	95.1	82.9	97.6	100	90.2	78	78	95	100	91.1
45 Hz	WPT-BE-MSVM	100	100	100	100	100	100	95.1	100	100	100	100	99.6
	WPT-PCA-MSVM	100	95.1	100	95.1	100	100	95.1	92.7	100	100	100	98
50 Hz	WPT-BE-MSVM	100	100	97.6	100	100	100	97.6	85.4	90.2	97.6	100	97.1
	WPT-PCA-MSVM	100	95.1	100	95.1	100	100	97.6	78	92.7	97.6	100	96
55 Hz	WPT-BE-MSVM	100	92.7	97.6	92.7	100	100	90.2	82.9	90.2	100	100	95.1
	WPT-PCA-MSVM	100	97.6	100	92.7	100	100	85.4	75.6	90.2	100	100	94.7
60 Hz	WPT-BE-MSVM	100	100	100	97.6	100	100	95.1	82.9	92.7	100	100	97.1
	WPT-PCA-MSVM	100	97.6	100	100	100	100	92.7	73.2	95.1	100	100	96.2
65 Hz	WPT-BE-MSVM	100	97.6	97.6	97.6	100	100	97.6	100	100	100	100	99.1
	WPT-PCA-MSVM	100	87.8	85.4	100	100	100	95.1	100	100	100	92.7	96.5
Overall	WPT-BE-MSVM												96.3
	WPT-PCA-MSVM												94.4

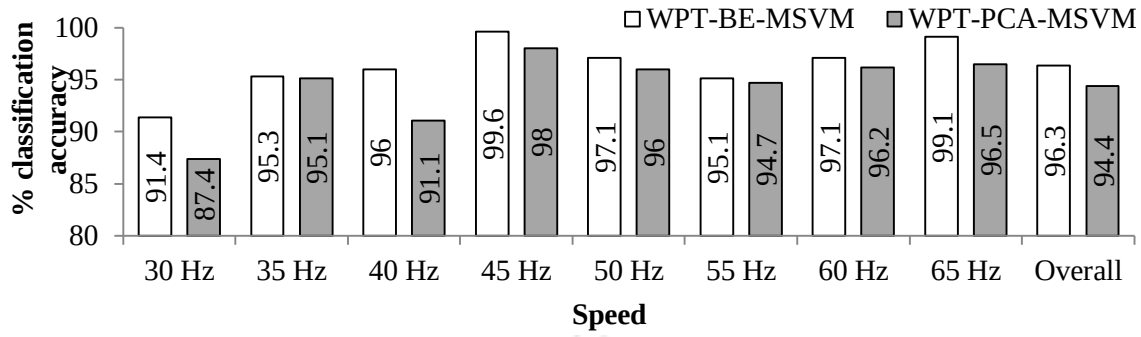


Figure 8.8: Percentage classification accuracy versus CP operating speed. Algorithm trained and tested at the same speed. Comparison of performance of WPT-BE-MSVM and WPT-PCA-MSVM for PCBF fault case

8.3.1.4 Dry run fault set classification

In this classification study, dry run faults are classified. The dry run condition is considered on various CP configurations. The results of the classification are presented in Table 8.9. The overall classification accuracies are given in Figure 8.9. It can be seen that both the methodologies show similar performance and the faults are classified with almost 100% overall accuracy. However, WPT-BE-MSVM demonstrated a clear prediction at all the speeds.

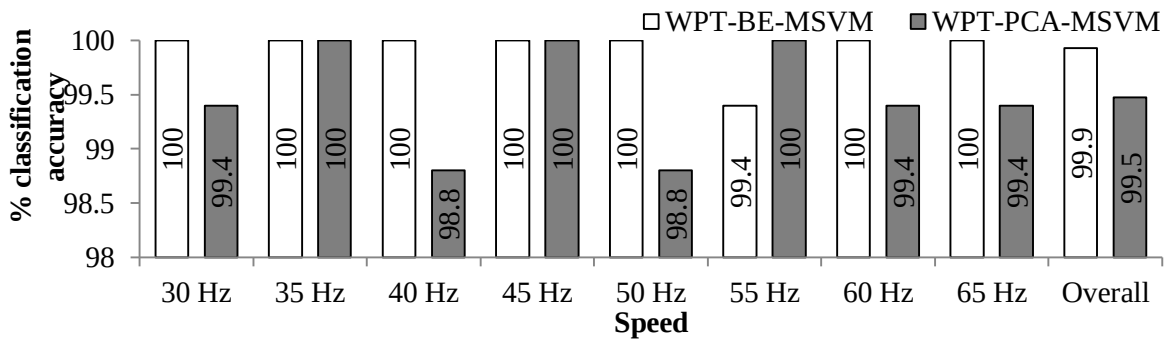


Figure 8.9: Percentage classification accuracy versus CP operating speed. Algorithm trained and tested at the same speed. The performance comparison of WPT-BE-MSVM and WPT-PCA-MSVM for DR fault case.

Table 8.9: Dry run fault family classification using WPT-BE-MSVM and WPT-PCA-MSVM

Train / Test Speeds (Hz)	Method	HP	SB5	IFSB5	PCSB5	C_a
30 Hz	WPT-BE-MSVM	100	100	100	100	100
	WPT-PCA-MSVM	97.6	100	100	100	99.4
35 Hz	WPT-BE-MSVM	100	100	100	100	100
	WPT-PCA-MSVM	100	100	100	100	100
40 Hz	WPT-BE-MSVM	100	100	100	100	100
	WPT-PCA-MSVM	95.1	100	100	100	98.8
45 Hz	WPT-BE-MSVM	100	100	100	100	100
	WPT-PCA-MSVM	100	100	100	100	100
50 Hz	WPT-BE-MSVM	100	100	100	100	100
	WPT-PCA-MSVM	97.6	100	100	97.6	98.8
55 Hz	WPT-BE-MSVM	100	100	100	97.6	99.4
	WPT-PCA-MSVM	100	100	100	100	100
60 Hz	WPT-BE-MSVM	100	100	100	100	100
	WPT-PCA-MSVM	100	97.6	100	100	99.4
65 Hz	WPT-BE-MSVM	100	100	100	100	100
	WPT-PCA-MSVM	97.6	100	100	100	99.4
Overall	WPT-BE-MSVM					100
	WPT-PCA-MSVM					99.5

8.3.1.5 AF fault set classification

As previously described, the AF fault family constitutes of all the faults at their low or medium severity and high severity (whenever applicable). This classification ensures that the classifier is capable of detection of faults belonging to different fault families (having distinct reasons of formation). Table 8.10 and Figure 8.10 show the performance comparison of both the methodologies to classify AF faults. It can be seen that both the methodologies (WPT-BE-MSVM and WPT-PCA-MSVM) can classify the faults at all the operating speeds of the CP with overall accuracies of 99.7% and 99.6%, respectively. Faults such as, SB2, SB5, DB1, IFSB5, IFDB1, IFDB5, PCSB2, PCSB5 and PCDB5 show 100% classification at all the test

speeds. WPT-BE-MSVM gives 100% fault classification at 50, 60 and 65 Hz, and, WPT-PCA-MSVM shows 100% classification at 60 Hz speed. From this Case A analysis, it is clear that distinguishing faults belonging to the same family and with closely varying severity is much more challenging than identifying faults belonging to different families (AF fault family and dry run case). Also, the fault classification becomes more involved when there is pseudo flow recirculation (arising due to mechanical CP faults/ manufacturing defects).

To check the capability of the algorithm in handling unavailability of fault history data, an intermediate speed fault classification is carried out. As a first step the algorithm is trained at two distinct speeds, and then it is tested at an intermediate speed, which is given as the Case B fault classification. The results of this analysis are given in the subsequent section.

8.3.2 Case B: Intermediate Speed Classification

In case the fault data is not available at the given operating speed of the CP but is available at higher and lower speeds, the Case B methodology may be used. In this method, however, cross-validation technique is not employed for the optimum parameter selection. Instead, from given ranges of parameters, the best combination of them is chosen by a systematic search. Case B classification is done on AF fault set to have enough representation of each fault family. The algorithm is trained at two speeds (e.g., 40 Hz and 50 Hz) and is tested at an intermediate speed (e.g., 45 Hz). While performing the experiments (details in Chapter 2), it was observed that the CP fault manifestation changes tremendously with its operating speed. Therefore, to accurately identify faults at intermediate speeds means that the chosen features must be insensitive to the CP operating speeds. In all the analyses presented under Case A, it was observed that both the developed methodologies and their associated features were not

showing any operating speed dependency. However, it cannot be said that this would ensure identification of faults at intermediate speeds as well. From Chapter 7 (CWT analyses) it was clear that though the classifier performed very well with same speed fault classification, it showed low classification ability at intermediate speeds. Therefore, it would be interesting to see if whether or not WPT (which is another time-frequency based representation of signal) would be successful in fault identification at the intermediate speeds.

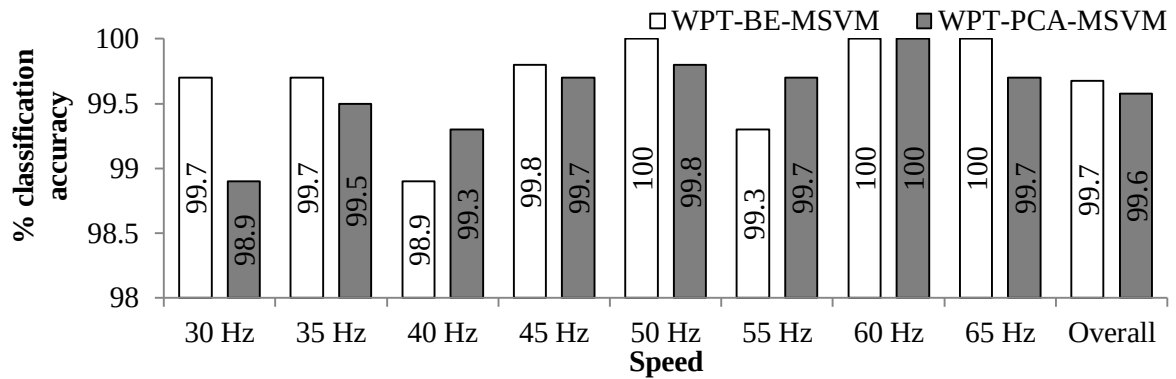


Figure 8.10: Percentage classification accuracy versus CP operating speed. Algorithm trained and tested at the same speed. The fault classification performance comparison of WPT-BE-MSVM and WPT-PCA-MSVM for AF fault case.

The comparison of the performance of both the methodologies is presented in Figure 8.11. In figure (30, 40/ 35) stands for training speeds of 30 Hz and 40 Hz and a test speed of 35 Hz, similarly others. Clearly, the WPT-BE-MSVM method outperformed the WPT-PCA-MSVM. The overall prediction accuracy of the WPT-BE-MSVM is 92.3%. If the results of this classification are compared to that of analysis in Case A, it is evident that the overall prediction accuracy decreased in this case which may be due to the change in fault manifestation with the changing operating conditions of the CP. However, the results of this classification are promising and better than that of the CWT based results given in Chapter 7. Comparison of performance of WPT-BE-MSVM and WPT-PCA-MSVM for AF fault case. To check the

robustness of the algorithm to handle noise in the testing data, the trained algorithm is tested with noisy test data. The analysis and its results are presented in the next section.

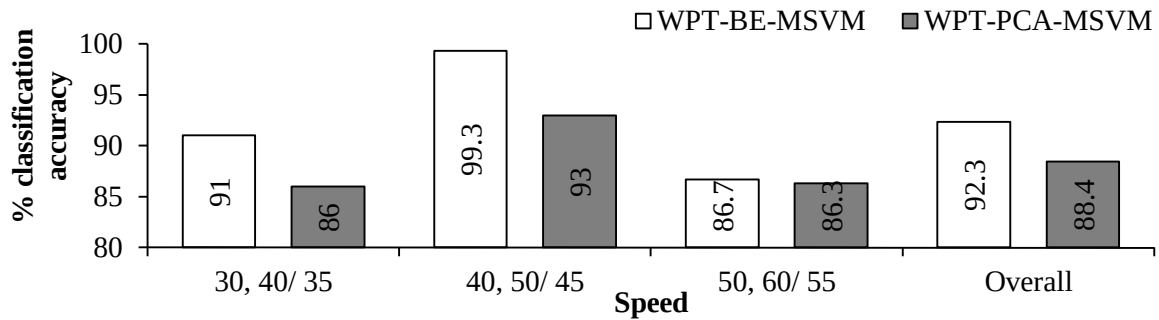


Figure 8.11: Percentage classification accuracy versus CP operating speed. Algorithm trained and tested at different speeds (30, 40/35 stand for algorithm training speeds of 30, 40 Hz and algorithm test speed of 35 Hz).

8.3.3 Case C: Fault Identification with Noisy Test Data at the Same Test Speed

In the next multiclass classification, the SVM algorithm is trained with features extracted from uncorrupted (baseline) signal and tested separately with the features extracted from time-domain signal corrupted with 5% AWGN, 10% AWGN, and 25% AWGN. The results of the classification are tabulated in Table 8.11. In order to understand the versatility of the algorithm, a multi-class fault classification of AF fault family is attempted. Also, WPT-BE-MSVM methodology has been used to perform the classification. Here it must be noted that a CV technique has not been used for the selection of best kernel and SVM parameters. Alternatively, the best parameters are selected based on a systematic search. That is, the parameters are given a range of values, and the combinations of the parameters that give the best classification performance are presented. From the table, it can be seen that the overall classification accuracies obtained while testing with 5% AWGN, 10% AWGN and 25% AWGN are 99.9%, 98.7%, and 98.5%, respectively.

Table 8.10: AF fault family classification using WPT-BE-MSVM and WPT-PCA-MSVM

Train and Test Speeds (Hz)	Method	HP	SB2	SB5	DB1	DB5	IF	IFSB2	IFSB5	IFDB1	IFDB5	PC	PCSB2	PCSB5	PCDB1	PCDB5	C_a
30 Hz	WPT-BE-MSVM	100	100	100	100	97.6	100	100	100	100	100	97.6	100	100	100	100	99.7
	WPT-PCA-MSVM	92.7	100	100	100	97.6	100	100	100	100	100	100	95.1	100	100	97.6	98.9
35 Hz	WPT-BE-MSVM	97.6	100	100	100	100	100	100	100	100	100	97.6	100	100	100	100	99.7
	WPT-PCA-MSVM	97.6	100	100	100	100	100	100	100	100	100	95.1	100	100	100	100	99.5
40 Hz	WPT-BE-MSVM	100	100	100	100	95.1	100	90.2	100	100	100	100	100	100	100	100	98.9
	WPT-PCA-MSVM	100	100	100	100	100	100	100	100	100	100	90.2	100	100	100	100	99.3
45 Hz	WPT-BE-MSVM	100	100	100	100	100	100	100	100	100	100	100	100	100	97.6	100	99.8
	WPT-PCA-MSVM	100	100	100	100	100	100	97.6	100	100	100	100	100	100	97.6	100	99.7
50 Hz	WPT-BE-MSVM	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	WPT-PCA-MSVM	100	100	100	100	100	100	100	100	100	100	97.6	100	100	100	100	99.8
55 Hz	WPT-BE-MSVM	100	100	100	100	100	100	100	100	100	100	95.1	100	100	95.1	100	99.3
	WPT-PCA-MSVM	100	100	100	100	100	100	100	100	100	100	97.6	100	100	97.6	100	99.7
60 Hz	WPT-BE-MSVM	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	WPT-PCA-MSVM	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
65 Hz	WPT-BE-MSVM	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	WPT-PCA-MSVM	100	99.7	100	100	100	100	100	100	100	100	99.7	100	100	100	100	99.7
Overall	WPT-BE-MSVM																99.7
	WPT-PCA-MSVM																99.6

Here, it seems that the classification accuracy has improved with noise addition, but in an engineering perspective, there is no significant difference between 99.7 and 99.9% classification. Therefore, it can be said that adding noise of up to 25% does not deteriorate the classifier's performance. However, with the addition of noise, there is an indicative decline in the performance of the classifier.

8.4 Comparison of the developed wavelet based methodologies to similar work from literature

In Chapter 7 and Chapter 8 CP fault diagnosis using features extracted from wavelet coefficients using CWT and WPT respectively, is presented. In order to understand the novelty of the current approach when compared to the other approaches (based on wavelets) discussed in literature, a comparison of studies is presented in Table 8.12. From the table it is clear that vibration signals are a preferred choice when it comes to fault diagnosis using wavelets. However, combination of signals proved to be beneficial in solving the fault diagnosis problems. Though almost all the methodologies in literature using various classifiers showed good prediction performance, the faults considered were also not too involved. In the current research a combination of faults (mechanical and hydraulic), varying severity of hydraulic faults, and different speeds of operation of the CP are considered. Apart from this and intermediate fault classification, classification when tested with noisy data is also considered. In addition, the classifiers using SVM and Wavelets combination haven't performed too well in two out of three cases. While in the current research even with high complicated sets of faults, the classifiers performed with nearly 100% accuracy. Therefore, it can be said that the present study may find applications for practical CP fault diagnosis.

Table 8.11: Comparison of fault classification accuracies obtained for AF fault classification using 5%, 10% and 25% AWGN

Train/ Test Speed (Hz)	AWGN	HP	SB2	SB5	DB1	DB5	IF	IFSB2	IFSB5	IFDB1	IFDB5	PC	PCSB2	PCSB5	PCDB1	PCDB5	C_a
30	5	100	99.2	100	100	94.1	100	100	100	99.2	100	100	100	100	100	100	99.5
	10	95	98.3	100	100	40.3	100	100	100	100	100	100	95	100	100	100	95.2
	25	98.3	95	100	100	40.3	100	100	100	87.4	100	100	100	100	99.2	100	94.7
35	5	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	10	100	100	100	100	76.5	100	100	99.2	100	100	100	98.3	100	100	99.2	98.2
	25	98.3	100	100	100	87.4	100	100	100	90.3	100	100	95	100	100	99.2	98
40	5	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	10	100	100	100	100	100	100	100	100	100	100	100	100	100	99.2	97.4	99.8
	25	100	100	100	100	98.3	100	100	98.3	100	87.4	100	95	100	100	100	98.6
45	5	100	100	100	100	100	98.3	100	100	100	100	100	100	100	100	100	99.9
	10	99.2	100	100	100	87.4	100	99.2	100	100	100	100	99.2	100	100	74	97.3
	25	100	100	100	100	90.3	100	100	100	100	76.5	100	97.5	100	100	100	97.6
50	5	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	10	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	25	100	100	100	100	100	100	100	100	100	100	99.2	100	100	100	100	99.9
55	5	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	10	100	100	100	100	100	100	100	100	100	100	97.5	100	100	98.3	100	99.7
	25	99.2	100	100	99.2	100	100	100	100	100	97.5	100	100	100	100	100	99.7
60	5	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	10	100	100	100	99.2	100	100	100	100	100	100	98.3	100	100	99.2	100	99.8
	25	100	99.2	100	100	98.3	100	100	99.2	100	100	100	100	100	97.5	100	99.6
65	5	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	10	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	25	100	99.2	100	100	98.3	100	100	100	100	100	100	100	100	100	100	99.8

Table 8.12: Comparison of fault prediction performance using wavelets to that presented in literature

Reference	Fault Type	Signal used	Sinal domain	Classification algorithm	Data sampling rate/ sample length	Features considered	Classification performance	Remarks
Nasiri et al., 2011	Cavitation with varying severities	Vibration	DFT	NN	10kHz	Kurtosis, crest factor	n/a	1. Varying severities of cavitation classified 2. Single speed fault detection
Muralidharan and Sugumaran, 2012	Bearing faults, Impeller faults, both bearing and impeller faults, cavitation	Vibration	DWT	Naïve Bayes, Bayes Net	24kHz/ 1024	DWT coefficients	100%	
Sakthivel et al., 2012a	Bearing faults, seal defects, impeller defects, impeller and bearing faults, cavitation	Vibration	Time domain, Wavelets	Gene expression programming, SVM, wavelets – gene expression programming	24kHz/ 1024	mean, standard error, median, standard deviation, sample variance, kurtosis, skewness, range, minimum, maximum, and sum	99.93%	
Muralidharan and Sugumaran, 2013c, 2013a	Bearing faults, Impeller faults, both bearing and impeller faults, cavitation	Vibration	DWT	Decision tree, J48	24kHz/ 1024	DWT coefficients	99.84%	1. Combination of mechanical faults considered 2. Single speed fault detection
Muralidharan and Sugumaran, 2013b	Bearing faults, Impeller faults, both bearing and impeller faults, cavitation	Vibration	DWT	Roughset fuzzy	24kHz/ 1024	DWT coefficients	99.84%	
Muralidharan et al., 2014	Same as above	Vibration	CWT	SVM	24kHz/1024	CWT coefficients	99.84%	
Ebrahimi and Javidan, 2017	Impeller faults and seal defects	Vibration	DWT	SVM	n/a	Mean, standard deviation, sample variance, RMS, crest factor, skewness, slippage, kurtosis, fifth and sixth central moments	96.67%	
Kumar and Kumar, 2017	Broken impeller, clogged impeller, bearing inner race defects and bearing outer race defects	Vibration	Time-domain and CWT	SVM	70kHz/ 7000	Kurtosis, skewness, mean, RMS, variance, peak, impulse factor, shape factor, crest factor, scale corresponding to highest energy, ration of peak energy to total energy	96.66%	
Current research	Suction/ discharge blockages (with varying severity), impeller defects, impeller defects together with blockages, pitted cover plate faults, pitted cover plate together blockages, dry run	Vibration and motor line current	CWT	SVM	5 kHz/ 5000	Mean, Standard deviation, reciprocal of standard deviation, entropy, 5 th , 6 th , 7 th moments, variance, RMS, RSS, SS	99.6% (Same train ,test speed)	1. Two signal types used 2. Multiple speed classification 3. Combination of mechanical and hydraulic faults considered 4. Varying severity of faults considered 5. Faults classified at intermediate speeds 6. Effect of data resolution presented 7. Effect of additive noise on the classification performance of the algorithm presented
							72.7% (Intermediate speed)	
			WPT				100% (Noise)	
							99.7% (Same train, test speed)	
							97.3% (Intermediate speed)	
							100% (Noise)	

8.5 Summary

In this chapter, the multi-fault classification is performed to detect the mechanical, hydraulic, and combined mechanical and hydraulic CP faults. The time-domain vibration and motor line current data is converted to the time-frequency domain using WPT. The multiclass SVM is used to perform the fault classifications. To select the most appropriate nodes (or coefficients) from the wavelet packet tree, two methodologies, namely, WPT-BE-MSVM and WPT-PCA-MSVM are developed. Where WPT-BE-MSVM relies on the calculation of nodal energies and then selects the top three nodes giving the highest energies. On the other hand, WPT-PCA-MSVM uses the philosophy of principal component analysis to select the most appropriate representation of the data. Many statistical features are extracted from the finalized nodes (and their corresponding coefficients). The best features are then selected using an accuracy-based approach, i.e., the wrapper model. For WPT-BE-MSVM methodology, mean, standard deviation, root mean square, and entropy features combination gave a good performance. Whereas, for WPT-PCA-MSVM, feature combination of standard deviation and root sum of squares is finalized. Later, the performance of various wavelet families is compared based on the prediction accuracy they demonstrate. Various wavelets showed very promising classification of CP faults. Hence, a wide range of wavelets can be used with the developed approach. However, the results of only one among the many good performing wavelets, Daubechies 6 (db6) is presented. The algorithm using both the methodologies performed satisfactorily in classifying faults at the same speed, intermediate speed and also with noisy data. The WPT based methods overcome the limitation posed by for CWT based analysis for an intermediate speed fault classification. Hence, this methodology can be very well adapted for an on-line CP fault detection.



CHAPTER 9: Multi-Fault Diagnosis of Centrifugal Pumps: A Comparative Analysis

9.1 Introduction

In Chapters 4 to 8, the multiclass CP fault classification accuracies obtained using features extracted from the time-domain, frequency-domain and the wavelet transforms have been presented. Each of these domains has its own share of advantages and limitations based on the nature of the faults and the type of signature the faults generate. To consolidate the results and to ease the comprehension of the same, in this chapter the results as well as the salient features of each of these analyses are compared and presented.

In all the three domains, the fault classification has been attempted at same CP operating speed and at an intermediate speed. To have an apple to apple comparison only the results obtained from low sampling rate data are presented. In addition, overall classification accuracies are given whenever possible so that a general picture of the fault classification can be understood.

9.2 Feature(s) used for the Fault Classification

For the CP fault classification, experiments are performed on the healthy as well as faulty CPs. The vibration from CP and line current data from the motor are extracted from the experimentation. This acquired data is in time-domain. This data cannot be directly used to train the machine learning algorithm, because it is of high dimensionality and it also has a lot of unnecessary/ redundant information. Hence, it needs to be properly processed so that only the relevant information is retained. One such technique of data pre-processing is the extraction of statistical features from the data which can be used for fault classifications.

Statistical features play a very important role in the classification accuracy obtained and the time taken for the fault identification. In the present work, features such as, the mean, variance, standard deviation, skewness, kurtosis, 5th statistical moment, 6th statistical moment, 7th statistical moment, root mean square, root sum of squares, sum of squares, shape factor, impulsion index, crest factor, tolerance index, power to average ratio, entropy, and reciprocal of standard deviation have been extracted from the domain of interest. In all the three domains of analysis, the wrapper model has been used for the feature selection. To recapitulate, the wrapper model-based feature selection relies on the accuracy of classification obtained using individual features. Whichever feature gives acceptable classification accuracy is shortlisted and the combinations of such features are tried out until there is no further improvement in the classification performance. It is important to note that though mathematically these features have the same expression in all the domains their values and their significance is very different.

Table 9.1 shows the list of domains used for CP fault diagnosis, their corresponding signal processing techniques and the finalized features obtained using the wrapper model. For the feature selection, a multi-class CP fault classification has been considered involving faults belonging to various fault families. It can be seen from the table that the time-domain based analyses require minimum signal processing efforts on the other hand the time-frequency domain based analyses have the most complicated method of signal pre-processing. The standard deviation feature works well in all the domains. The minimum number of features required for fault classification is two (i.e., for time-domain, time-frequency domain (CWT-BE-MSVM, CWT-PCA-MSVM and WPT-PCA-MSVM)). WPT-BE-MSVM uses a maximum number of features (i.e., four). However, one of the advantages of using the SVM

classifier is that it can handle a large number of features. But, extracting more features may prove to be costly and requires higher computational time.

Table 9.1: Features used for CP multiclass fault classification in time-domain, frequency-domain and time-frequency domain

S. N.	Domain of analysis	Method of signal processing	Finalized features
1	Time-domain	Extracted from raw time-domain data captured	Mean, reciprocal of standard deviation
2	Frequency-domain (power spectrum)	Extracted after power spectrum transformation of the time-domain signal	Mean, standard deviation, reciprocal of standard deviation
3	Time-frequency domain (Continuous wavelet transform (CWT))	Time-domain signal transformed using CWT, best scales and corresponding coefficients selected using best energy or principal component analysis methods, the statistical features extracted from the finalized coefficients	Standard deviation, root sum of squares (CWT-BE-MSVM and CWT-PCA-MSVM)
4	Time-frequency domain (Wavelet packet transform (WPT))	Time-domain signal transformed using WPT, best nodes and corresponding coefficients selected using best energy or principal component analysis methods, the statistical features extracted from the finalized coefficients	Mean, standard deviation, entropy and root mean square (WPT-BE-MSVM); Standard deviation and root sum of squares (WPT-PCA-MSVM)

9.3 CP Fault Classification Performance Using Features Extracted From Various Domains

The multi-class CP fault classification was aimed at segregating various CP faults based on the fault type and the fault location. Also, in the case of a hydraulic fault the identification of fault severity is also attempted. To carry out the fault classifications faults were divided into various fault families. They are BF fault family, IFBF fault family, PCBF fault family, dry run fault family and AF fault family.

9.3.1 BF Fault Family

BF stands for blockage faults family. Blockages could be of two types, suction blockages (SB) or discharge blockages (DB). In BF fault family pure blockage faults (i.e., without any mechanical faults) are considered, which include, SB1, SB2, SB3, SB4, DB1, DB2, DB3, DB4 and DB5⁹ faults. The BF fault classification is aimed at, (1) identification of the absence/presence of the fault and, (2) establishing its location and severity.

In this section, the fault classification performance of various domain-based analyses carried out in identifying BF fault family is presented. It is to be noted that in all the classifications, a randomly selected 50% of the data is used for training the algorithm and the reminder 50% of non-overlapping set of data is used for testing. In all cases, the grid search method and the cross-validation techniques are used for selection of optimum SVM parameters. One-versus-one approach is used for the multiclass classification. Table 9.2 presents the comparison of the classification performance obtained using best-performing features extracted in similar domains.

⁹ SBk – (k/6)th suction flow restriction, DBk – (k/6)th discharge flow restriction

In Table 9.2 the highest classification accuracies obtained for each of the faults are shown in bold. These classification accuracies are the overall accuracies corresponding to each of the faults. That means they are average classification accuracies obtained for individual faults over the entire operating speed range considered.

The classification of HP and SB1 faults is very complicated as SB1 is of low severity and while experimentation there was no visible flow anomaly observed when SB1 blockage was induced. The bubble formation started after SB2 level and before the SB3 level. However, using time-domain based features all the SBs and HP condition are classified with almost 100% accuracy. The CWT-BE-MSVM method classified all the DBs with high precision. Overall, the best classification of the BF family of faults was found using the CWT-BE-MSVM method, and the least accuracy was found using frequency-domain approach. Moreover, the frequency-domain based method showed very less classification of high severity DB faults. This is because frequency-domain analysis considers the fault signatures to be stationary. However, fluid flow abnormalities and corresponding bubble formations result in transient signals, which are well captured in time domain or time-frequency domain studies. The other wavelet-based methods including the CWT-PCA-MSVM method and the WPT-BE-MSVM method also performed on par with the CWT-BE-MSVM method.

9.3.2 IFBF Fault Family

IFBF is the abbreviation for impeller faults and blockage faults. This family of faults comprises of blockages given on CP with impeller defects IFSB (impeller defects with suction blockage) and IFDB (impeller defects with discharge blockage), and also impeller defects alone.

Table 9.2: Comparison of BF fault family classification using features extracted from various domains

Domain of analysis	Overall fault classification accuracy of individual faults										Overall
	HP	SB1	SB2	SB3	SB4	DB1	DB2	DB3	DB4	DB5	
Time-domain	99.4	100	100	100	99.7	95.4	97.9	98.8	94.2	93.3	97.9
Frequency-domain	95.8	97.3	99.4	97.3	97.9	100	98.5	97.6	87.8	89	96
CWT-BE-MSVM	99	96.8	99.5	99.8	99.8	99.3	100	99.5	97.5	97.3	98.8
CWT-PCA-MSVM	97.8	98	99.5	99.8	99	100	99.3	98.3	98	96.5	98.6
WPT-BE-MSVM	98.5	99.1	99.7	100	100	99.7	100	97.9	94.5	96.6	98.6
WPT-PCA-MSVM	96.6	99	99.8	100	99.2	97.9	99.1	95.1	95.4	93.7	97.6

This classification aims at, (1) identification of absence or presence of impeller defects, (2) detection of the combined existence of blockage faults and impeller defects, and (3) estimation of the location and severity of flow restriction (if any). The faults considered in this family are: HP, IF, IFSB1, IFSB2, IFSB3, IFSB4, IFDB1, IFDB2, IFDB3, IFDB4 and IFDB5¹⁰. This classification case is essential because experimentally it was observed that mechanical CP faults are capable of inducing flow recirculation (pseudo-recirculation) even without any external flow restrictions. Therefore, the presence of a combination of mechanical and hydraulic faults in a CP may accelerate its failure. Also, rotating mechanical faults, like the impeller defects have characteristic fault frequencies and are identifiable in frequency-domain. However, the recirculation (and bubble formation) has a transient behavior. Hence, it is a complex classification problem. The comparison of the classification performance in segregating the IFBF fault family is presented in Table 9.3.

From Table 9.3, it can be seen that HP condition is classified with 100% accuracy in all the domains. This is because IF and its allied faults produce very different fault signatures when compared to the HP condition. IFDB faults show low classification while IFSB faults show almost 100% in all the domains. It is interesting to observe that the frequency-domain based fault classification gave a higher overall fault classification accuracy when compared to the time-domain. This may be attributed to the fault characteristic frequencies produced by the IF faults. Like in the previous case (BF fault classification case), the highest fault classification is demonstrated by the CWT-BE-MSVM method.

¹⁰ IF – impeller faults, IFSB k – CP with impeller faults and $(k/6)^{\text{th}}$ suction flow restriction, IFDB k – CP with impeller faults and $(k/6)^{\text{th}}$ discharge flow restriction

9.3.3 PCBF Fault Family

The cover plate faults along with the suction and discharge blockage faults fall under the category of the PCBF fault family. The PCBF, therefore, stands for pitted cover and blockage faults. The faults considered in this family are HP, PC, PCSB1, PCSB2, PCSB3, PCSB4, PCDB1, PCDB2, PCDB3, PCDB4, and PCDB5¹¹.

Unlike impeller defects, the pitted cover fault is not a rotating fault. Therefore, there may not be any apparent characteristic frequencies of this fault. However, due to the presence of undulations on the face of the cover plate of the CP, the flow patterns change, and a recirculating flow may result. This was even evidently seen as bubble formations in the suction and discharge pipes of the CP during experimentation. Therefore, early identification of faults is essential to avoid any premature damage of CPs. Table 9.4 shows the comparison of the classification performance obtained using various domain based methodologies to identify PCBF faults. Here also, the HP condition is identifiable with 100% accuracy using all the developed methodologies. It is also observed from the table that there is difficulty in segregating PCDB faults, especially using time or frequency domain methods. This may be attributed to the need to capture both the time and frequency content of the data simultaneously to identify the faults. On the other hand, PCSB faults are clearly distinguished using all the domains with almost 100% accuracy. For this classification, the best performing methodology again is the CWT-BE-MSVM, and the frequency domain features show the lowest classification performance.

¹¹ PC – pitted cover faults, PCSB_k – CP with pitted cover and (k/6)th suction flow restriction, PCDB_k – CP with pitted cover and (k/6)th discharge flow restriction

9.3.4 Dry Run Fault Family

Experimentally, it was observed that when the 5/6th of suction pipe's cross-sectional area was restricted, there was no fluid pumping. This means that the CP was running *dry*. The faults considered in this family are, SB5, IFSB5, and PCSB5. Table 9.5 shows a comparison of the classification performance demonstrated by various methods. Due to the severe as well as distinct nature of all the considered faults in this family, almost all the methodologies show a perfect fault classification. Here also like in all the previous cases of fault classifications, the CWT-BE-MSVM algorithm demonstrated the highest classification performance.

9.3.5 AF Fault Family

Up till now, all the classifications performed considered faults belonging to a particular fault family meaning that there was a common origin point for all the faults belonging to a particular family and there was varying severity suction or discharge flow restriction on it. However, it may also be possible that multiple faults exist simultaneously or there might be a group of CPs which are being monitored and each CP may have a different fault inside them. To consider such a case, the AF fault family is created. It comprises CP faults belonging to various fault families at a low severity level and a high severity level. A total of 15 faults are considered in this cluster and they are: HP, SB2, SB5, DB1, DB5, IF, IFSB2, IFSB5, IFDB1, IFDB5, PC, PCSB2, PCSB5, PCDB1, and PCDB5. This AF fault family is considered to be versatile as it represents all the fault families especially and the fault formative stages of suction or discharge blockages. Therefore, this fault classification set is used for many other fault classification analyses as well.

9.3.5.1 AF fault family – testing at same speed

In continuation to the analyses presented till now (BF, IFBF, PCBF and dry run fault family classifications), first, it is attempted to classify the CP faults while training and testing the algorithm at the same CP operating speed. Table 9.6 presents the comparison of the AF family fault classification results obtained using various domain-based methodologies developed. The maximum overall classification accuracy obtained is 99.9% using time-domain based features. It can be seen from the table that all the methodologies presented to give a classification accuracy of more than 99.1% for this fault set. Also, many faults show an overall 100% classification using all the methods. From the comparison of analyses presented in Table 9.2 to Table 9.6, it is clear that identifying faults originating from different causes is comparatively more precise than identifying faults belonging to the same family with varying severities. This is because faults originating from various causes are expected to have very distinct features and therefore are easy to segregate.

9.3.5.2 AF fault family – testing at intermediate speed

In most practical cases where these algorithms are supposed to be applied for CP monitoring, the history of CP data (faulty and/or healthy) is readily available at the current operating state of the CP. However, due to certain modifications in the plant processes, if the current operating speed of the CP needs to be changed, how is the fault diagnosis supposed to be performed without available history of fault data to train the algorithm?

To study such a case, an intermediate speed fault classification is attempted. According to this case, the training of the multiclass SVM algorithm is done using data available at two different speeds, and later this trained algorithm is tested at an intermediate speed. This intermediate

speed represents the current CP operating speed, and the training speeds represent the data that is already available. To carry out this classification AF fault family is used. Figure 9.1 shows a comparison of AF fault classification performed at an intermediate speed. In the figure, '30, 40/ 35' stands for training with features at 30 Hz and 40 Hz, and testing with the features at 35 Hz, similarly others.

Here, it must be noted that only one wavelet-based method for each of CWT and WPT is presented. From all the analyses using features at same training and testing speeds, the CWT-BE-MSVM outperformed all the other methods. However, for an intermediate speed fault classification, the CWT-BE-MSVM gives lowest classification accuracy. This is because there is a shift in features observed with the CP operating speed using the CWT analysis. The time-domain analysis followed by the WPT analysis shows the highest fault classification performance at the intermediate speeds.

9.3.5.3 AF fault family – testing with corrupted data

In the next segment of fault classification, the baseline time-domain data obtained from the CP experimentation is corrupted with additive white Gaussian noise (AWGN). The algorithm is trained with features obtained from the baseline data, while it is tested with features obtained from corrupted data. This fault classification is considered to account for cases in which the classifier is developed based on laboratory data and tested with real-time industry data. The real-time industry data may be corrupted due to various surrounding influences, or there might be sensor related problems.

Table 9.3: Comparison of IFBF fault family classification using features extracted from various domains

Domain of analysis	Overall fault classification accuracy of individual faults											Overall
	HP	IF	IFSB1	IFSB2	IFSB3	IFSB4	IFDB1	IFDB2	IFDB3	IFDB4	IFDB5	
Time	100	99.1	99.1	99.7	98.5	100	91.2	91.8	88.1	93	96.7	96.1
Frequency	100	99.7	99.7	99.7	99.7	98.2	98.8	93.9	87.8	96.7	93.9	97.1
CWT-BE-MSVM	100	99.5	98.5	99.5	100	100	99.8	98.8	96.5	96.8	98.5	98.9
CWT-PCA-MSVM	100	99.5	99.8	100	99.5	100	97.5	97	98	94.3	90.3	97.8
WPT-BE-MSVM	100	99.4	99.7	99.7	100	100	95.7	94.8	95.7	92.1	89.9	97
WPT-PCA-MSVM	100	99.7	99.4	99.1	99.1	99.1	93	93.6	94.8	97.3	93.9	97.2

Table 9.4: Comparison of PCBF fault family classification using features extracted from various domains.

Domain of analysis	Overall fault classification accuracy of individual faults											Overall
	HP	PC	PCSB1	PCSB2	PCSB3	PCSB4	PCDB1	PCDB2	PCDB3	PCDB4	PCDB5	
Time	100	91.2	99.4	94.2	100	100	99.4	98.8	92.4	89.9	99.1	96.8
Frequency	100	97	96.4	95.4	99.4	100	91.5	85.4	89.6	89.3	89.6	93.9
CWT-BE-MSVM	100	98.3	97	98.5	99.8	100	97	97.3	98.5	97.5	98.3	98.4
CWT-PCA-MSVM	100	97.8	92.3	96.8	99	100	94.8	98.8	95.3	97.5	96.8	97.2
WPT-BE-MSVM	100	97	97.9	97	99.1	100	95.4	90.2	92.7	94.5	96	96.3
WPT-PCA-MSVM	100	94.8	95.4	94.8	97.9	99.4	93.3	84.7	90.5	93	94.2	94.4

Table 9.5: Comparison of dry run fault family classification using features extracted from various domains

Domain of analysis	Overall fault classification accuracy of individual faults				Overall
	HP	SB5	IFSB5	PCSB5	
Time	100	98.8	99.1	99.7	99.4
Frequency	99.4	100	100	100	99.9
CWT-BE-MSVM	100	100	100	100	100
CWT-PCA-MSVM	99.8	100	99.3	98	99.3
WPT-BE-MSVM	100	100	100	99.7	99.9
WPT-PCA-MSVM	98.5	99.7	100	99.7	99.5

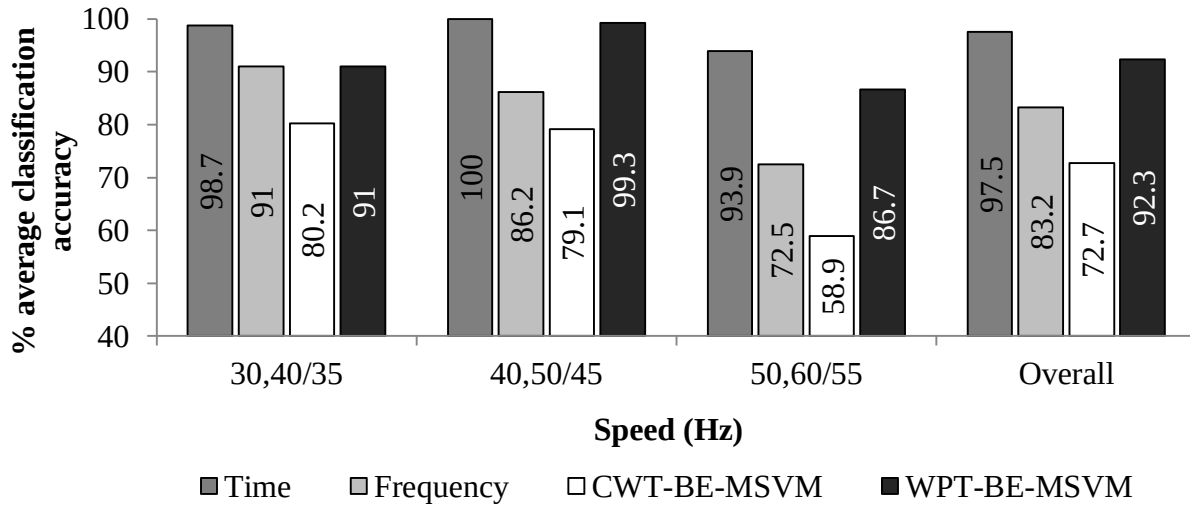


Figure 9.1: Comparison of AF fault family classification using features extracted from various domains at an intermediate speed

Table 9.6: Comparison of AF fault family classification using features extracted from various domains

Domain of analysis	Overall fault classification accuracy of individual faults															Overall
	HP	SB2	SB5	DB1	DB5	IF	IFSB2	IFSB5	IFDB1	IFDB5	PC	PCSB2	PCSB5	PCDB1	PCDB5	
Time	100	100	100	99.9	100	99.7	99.7	100	100	100	100	99.7	100	99.4	100	99.9
Frequency	97	99.4	97	100	100	99.7	100	100	99.7	100	100	97	100	98.2	98.2	99.1
CWT-BE-MSVM	99.3	99.5	100	100	99.3	99.8	100	100	100	99.8	99	99.3	100	99	99.5	99.6
CWT-PCA-MSVM	99.3	99.8	100	100	99	100	99.8	100	100	100	98.5	97.3	100	99.5	100	99.5
WPT-BE-MSVM	99.7	100	100	100	99.1	100	98.8	100	100	100	98.8	100	100	99.1	100	99.7
WPT-PCA-MSVM	98.8	100	100	100	99.7	100	99.7	100	100	100	97.5	99.4	100	99.4	99.7	99.6

To study this case the AF fault family is considered. However, here the training and the testing of the algorithm is carried out at the same CP operating speed. The requirement of this classification to be successful is that the features must stay unaffected by the corruption of the signal.

Table 9.7 shows the comparison of the AF fault family classification using test features extracted in various domains from corrupted testing data. From the table, it can be seen that up to 25% of AWGN time-domain, the CWT and WPT based methods seem unaffected. However, even with the addition of 5% AWGN, there is a drop of 6% classification accuracy using frequency domain based features. This is because the noise added to the time domain signal changes the frequency pattern of the signal. The CWT-BE-MSVM method seems to be the least affected of all the developed methodologies for the addition of white Gaussian noise.

9.3.6 Overall Comparison of Different Fault Classification Cases

In all the chapters (Chapter 5 to Chapter 8) fault classifications have been presented for various cases. Where, Case A talks about the classification using features extracted at same training and testing speeds, Case B represents the AF fault classification at an intermediate speed, and Case C represents the AF fault classification using noisy test data. To understand in a single shot, which methodology is the best suited for which classification and the maximum classification accuracy obtained for each of the methods, a table, Table 9.8 is presented.

For almost all the Case A classifications, the CWT-BE-MSVM methodology gives the best prediction performance, with a least of 98.4 (for the PCBF fault classification) and highest

classification of 100% (for the dry run fault classification). Also, for the AF fault family classification of Case A, the time-domain based method outperformed all the other methods.

Table 9.7: Comparison of AF fault family classification using features extracted in various domains from corrupted testing data

Domain of analysis	Corruption state	Overall accuracy
Time-domain	Baseline	99.9
	5%	100
	10%	99.9
	25%	96.2
Frequency-domain	Baseline	99.1
	3%	96
	5%	93.3
CWT-BE-MSVM	Baseline	99.6
	5%	100
	10%	99.9
	25%	99.5
WPT-BE-MSVM	Baseline	99.7
	5%	99.9
	10%	98.8
	25%	98.5

For Case B, an intermediate speed fault classification of AF faults, the time-domain based method showed the highest overall classification accuracy of 97.5%. The best classification performance for Case C fault classification using corrupted test data was demonstrated again by the CWT-BE-MSVM method. It could handle 25% of AWGN while retaining the classification accuracy to 99.5%.

Table 9.8: Bird's eye view of various fault classification cases

Case	Faults for training and testing	Training speed (Hz)	Remarks	Maximum overall classification accuracy	Method demonstrating maximum classification
A	BF, HP	30 – 65	Train and test speeds are same	98.8	CWT-BE-MSVM
	IFBF, HP			98.9	CWT-BE-MSVM
	PCBF, HP			98.4	CWT-BE-MSVM
	Dry run, HP			100	CWT-BE-MSVM
	AF			99.9	Time-domain
B	AF	30 – 65	Intermediate test speed	97.5	Time-domain
C	AF	30 – 65	Noisy signal for testing	99.5	CWT-BE-MSVM

9.4 Summary

In this chapter, an attempt is made to compare the classification performance of various methodologies developed in the current work. These methods were primarily based on the domain in which the fault features were extracted. Therefore, as a broad classification, there were three domains of analysis, i.e., the time-domain, frequency-domain and time-frequency domain. Each of these domains has its share of advantages and limitations. Hence, by comparison of all these methods, it is attempted to understand which method can be chosen for a particular type of fault classification. The simplest of the all the methodologies is that of the time-domain. The features extraction in time-domain is not at all complicated and is

straightforward. However, the frequency-based method is a bit more complicated than time-domain, and the time frequency-based methods (i.e., the CWT-BE-MSVM, CWT-PCA-MSVM, WPT-BE-MSVM, and WPT-PCA-MSVM) are much more involved. Also, the time-based methods, CWT-BE-MSVM, CWT-PCA-MSVM, and WPT-PCA-MSVM methods require only two features for the representation of the faults.

From the comparison of all the methods, it was understood that for the most of fault classification cases involving faults with closely varying severities (belonging to the same family, e.g., BF, IFBF, PCBF, and dry run) and having training and testing data at the same operating conditions available, the CWT-BE-MSVM outperformed all the other methods for CP fault classifications. However, when faults belong to different fault families (AF faults), the time-domain analysis showed the best classification performance both at same and intermediate test speeds. The CWT-BE-MSVM method also showed the highest immunity to the corruption of the time-domain signal. Here, it must be noted that all the other methodologies based on other domains of the signals have also shown good classification performance record. However, the results of the CWT-BE-MSVM and that of time-domain are highlighted as they showed the highest prediction.



CHAPTER 10: Concluding Remarks

10.1 Introduction

In general, the present day industries are still relying on age-old techniques for plant maintenances. These techniques though mature and reliable consume a lot of time and investments. Therefore, to overcome the drawbacks/ limitations of traditional maintenance strategies, in the present work, a machine learning based condition monitoring technique is proposed to identify multiple mechanical and hydraulic CP faults existing independently or in combinations. The multiclass SVM is employed as the machine learning algorithm. Experimentally acquired vibration signals from the CP and current signatures from the motor are used to identify the CP faults. The fault detections/ diagnoses are done in time-domain, frequency-domain and time-frequency domains. The fault classification performance results of these three domains are compared. In addition, the effect of using both the vibration and current signatures together on the classification performance of the algorithm is studied in time and frequency domains. The fault prediction performances of the algorithm have been checked in both time and frequency domain using data sampled with the low and high sampling rates.

To select the optimal SVM and kernel parameters, a method adapting the grid search technique, the cross-validation technique and the wrapper model has been devised. The wrapper model is simple to comprehend and is time-saving. To study the fault classification in the time-frequency domain, CWT and WPT methods are employed. Here, the effect of the choice of mother wavelets on the classification performance of the algorithm is compared and established. Initially, the fault diagnosis is attempted for the same CP speed case, and then it

is extended to the intermediate CP speed case to take care of the practical problem of unavailability of data at certain CP speeds. The CP fault diagnosis is attempted at various distinct speeds of the CP operation. To check the robustness of the developed methodology, the test data is corrupted. The fault prediction performance using features extracted in various domains of the original experimental signal is noted and reported.

10.2 Major Conclusions from the Current Works

- This investigation used SVM based multi-class classification methodologies which suggest that the features extracted from vibration signatures are advantageous and in the most cases sufficient for the CP fault diagnosis. However, the addition of the features extracted from the motor line current signatures improves the classification performance of the developed algorithm. More so, it is expected that if the motor directly drives the CP, the load variation on the CP due to the faults is more sensitively reflected on the motor current, thus making it an excellent option for the CP fault diagnosis.
- To study the effect of data sampling rate on the classification performance of the algorithm, two sampling rates, high sampling rate (sampling rate: 20 kHz, 2000 sampling points) and low sampling rate (sampling rate: 5 kHz, 5000 sampling points) are extracted. It is observed that the sampling rate has a significant effect on the classification performance of the algorithm. Also, from the time-domain and the frequency-domain studies it is observed that the classification performance of the algorithm using the lower sampling rate is much better than that of the higher sampling rate.
- Both independently existing and co-existing mechanical and hydraulic CP faults are considered during experimentation. Also, varying severities of the hydraulic faults are considered on the CP. Also, it is observed from the experiments that the mechanical and

hydraulic CP faults form a vicious fault cycle (one fault type leading to another and vice versa). Therefore, it is vital to curbing the CP faults at their formative stages.

- CP fault classifications are done using features extracted from three domains (time-domain, frequency-domain and time-frequency domain (CWT and WPT). For the selection of best scales or best nodes from the WT analysis two methods based on the energy criterion and the principal component analysis are presented. The energy-based method derived from the CWT analysis is termed as the CWT-BE-MSVM, and that derived from the WPT is termed as the WPT-BE-MSVM. Similarly, the PCA based method derived from the CWT is called the CWT-PCA-MSVM, and that derived from the WPT is called the WPT-PCA-MSVM. The classifications were attempted at test speed same as train speed and also an intermediate test speed. The same speed fault classification showed a classification accuracy of more than 95% in all the domains. However, the intermediate speed case classification did not show high classification accuracy but is satisfactory.
- The faults considered on the CP were divided into different fault families, the BF (blockage fault family), IFBF (Impeller faults and blockage faults), PCBF (pitted cover and blockage faults), dry run fault family, and AF (all faults family). In all the domains, it is observed that the identification of fault severity is much more complicated than identifying faults belonging to different fault families. This is because faults belong to the same family and having slight variations in the severity produce similar fault signatures, and therefore similar fault features.
- In the present work, features such as, the mean, variance, standard deviation, skewness, kurtosis, 5th statistical moment, 6th statistical moment, 7th statistical moment, root mean

square, root sum of squares, sum of squares, shape factor, impulsion index, crest factor, tolerance index, power to average ratio, entropy, and kurtosis index have been extracted from the domain of interest. Features such as the mean and the reciprocal of standard deviation are used to characterize the considered faults when extracted from time-domain. Similarly, the mean, standard deviation and reciprocal of standard deviation from the frequency-domain; the standard deviation and root sum of squares from the CWT-BE-MSVM and the CWT-PCA-MSVM; the mean, standard deviation, entropy and root mean square from the WPT-BE-MSVM; and the standard deviation and root sum of squares features from WPT-PCA-MSVM are found to be useful.

- Classification accuracy is used as a criterion to select the best mother wavelets. Wavelets such as the gaus1, gaus2, gaus5, morl, and bior3.3 showed good classification performance when used in the continuous wavelet transform; and db6, sym4, and rbior6.8 wavelets showed good classification when used in the wavelet packet transform.
- To validate the developed methodology, it is trained and tested with data acquired from the CP operating at a wide range of distinct speeds. All the methodologies developed based on different domains of considered data demonstrated that there is no or less operating speed dependence of the fault classification performance. However, for an intermediate speed case, it is found that there is some operating speed dependence as the fault manifestation changes significantly with the CP speed. Also, to check the robustness of the algorithm, it is tested with features acquired from the corrupted data. The amount of corruption of data that can be handled clearly depends on the domain in which the data is used.

- From the comparison of all the methods, it is understood that for the most fault classification cases involving faults with closely varying severities (belonging to the same family, e.g., the BF, IFBF, PCBF, and dry run) and having training and testing data at the same operating conditions available, the CWT-BE-MSVM outperformed all the other methods. However, when faults belong to different fault families (e.g., AF faults) the time-domain analysis showed the best classification performance both at the same and intermediate test speeds. The CWT-BE-MSVM method also showed the highest immunity to the corruption of the time-domain signal. Here, it must be noted that all the other methodologies based on other domains of the signals have also shown good classification performance record. However, the results of the CWT-BE-MSVM and that of time-domain are being highlighted as they showed the highest prediction.
- The average CPU time taken to perform the fault diagnosis is around 2-6 seconds for same speed cases and around 3-8 seconds for intermediate speed. This indicates that using the developed methodologies, quick decision regarding the faults can be taken. However, it must be noted that this speed is very much dependent on the computational capabilities available.

10.3 Main Contributions of the Thesis Work

- An SVM based on the CP multiple fault diagnosis approach is developed. This method makes use of the cross-validation along with the grid search and wrapper model to detect various mechanical and hydraulic CP faults.
- Wide varieties of statistical features are considered, and the most effective ones are chosen in each of the considered domains using the wrapper model for an effective CP fault diagnosis.

- The best representative coefficients from the CWT and WPT analyses are chosen using the best energy, and principal component analysis approaches.
- The effect of various mother wavelets based on the CWT and the WPT on fault classification accuracies is studied. It is found that some of the chosen mother wavelets performed very well in the CP fault diagnosis.
- The fault diagnosis is performed based on data acquired at low and high sampling rates. The lower sampling rate analyses demonstrated better classification performance.
- The algorithm is tested for performance in three cases, the same speed train and test case, the intermediate speed test case, and the corrupted data test case. The prediction accuracy slightly lowered in the intermediate load case. Also, frequency domain features are sensitive to noise addition.
- The developed algorithm can classify 33 different mechanical and hydraulic CP faults. The severity of the faults is also predicted successfully.

10.4 Recommendations from the Current Work

A methodology for the fault diagnosis of various mechanical and hydraulic CP faults is developed in the present work. It is expected that this methodology can be extended further to other CP systems/ pump installations in general and for the diagnosis of faults in other rotating machinery. The following are abstracted recommendations derived from the present work:

1. The vibration and motor line current signals are useful in identification of CP faults. Vibration signatures, if acquired from appropriate locations on the CP, can help identify CP faults with very good accuracy. However, it seems that the motor line current signals more useful when the CP is directly driven by the motor (unlike in the present study where the CP is driven using a belt drive).

2. The quality of acquired data (frequency resolution of it) plays a significant role in the fault identification. In addition, the choice of its quality depends upon the nature of the faults under study, the operating conditions of the system. Therefore, it is expected that a user of the current methodology doesn't blindly use the demonstrated values of signal sampling rates and length of the samples. In the present study an improved resolution in the frequency domain helped significantly improve the classification performance. However, the time taken to acquire the fault data also increases significantly. This maybe a concern for practitioners applying this technique for a real time fault diagnosis. In the present work features extracted from 5 kHz/ 5000 samples/dataset (Nyquist frequency: 2500 Hz) seem to give a better classification results than 20kHz/ 2000 samples/dataset (Nyquist frequency: 10000 Hz). However, after closer observation of the data (especially at 65 Hz frequency) it is found that the maximum frequency activity is up to 2800 to 2000 Hz. Hence, a further lower sampling rate may be chosen (giving a Nyquist frequency between 1800 to 2000 Hz) and which may end up giving a further improved performance of the classifier. It must be noted that it is very important to take into account the operating CP speed before deciding on the data sampling rate.
3. The choice of the features also plays a significant role in the classification performance; hence they have to be carefully and systematically selected.
4. In case of CP fault diagnosis, the domain in which CP fault diagnosis is carried out also has a deciding role on the accuracy of the classifier. Hence, it needs to be judiciously chosen based on the type of the faults present and how quickly the diagnosis needs to be performed.

5. The convergence of statistical features may also be chosen to identify the sampling rate of data acquisition.
6. Spectral kurtosis may be extracted instead of kurtosis when dealing with frequency or time-frequency spectra.
7. Taking data from two axis (ignoring the vertical transverse direction) may be enough for the estimation of faults.

10.5 Limitations of the Thesis Work

This work considers the laboratory generated data for the fault diagnosis and the faults were administered artificially. However, in the real-world scenario the faults develop naturally and they may decrease the performance of the classifier. As presented in Chapter 1, there are numerous causes of CP failures, only a few of them have been considered in this work. This work has not been tested on the industry data for validation. The thesis work does not study the mathematical/numerical modeling of the faults which is a very involved subject in itself.

10.6 Scope for Future Work

- In the current work, vibration and motor line current data have been employed. However, due to the transient nature of the flow anomalies, the transient pressure measurements can also be used for the CP fault diagnosis. Apart from the pressure, acoustic signals may also carry significant fault information.
- It would be interesting to apply a laboratory developed algorithm (such as this) on another CP system of different configuration. Since, the statistical features carry the general tendency of the signals, it is expected that they show similar feature variation in specified conditions. If this works well, it will ensure that the methodology becomes applicable to

a wide variety of CP geometries and assemblies. In this work the wrapper model is used for feature selection; however, other methods like the principal component analysis, the linear discriminate analysis, the genetic algorithm, and the distance evaluation technique (Soylemezoglu et al., 2011; Yunlong and Peng, 2012; Sakthivel et al., 2014) can be used, and the performances can be compared.

- Furthermore, it would be interesting to see if the algorithm training data could be acquired from numerical simulations packages (like finite element analysis packages ABAQUS, ANSYS, etc.) and tested with real-time data. This would eliminate the problem of data generation for algorithm training.
- Also, the study of the effect of geometric tolerance variation (for the same configuration of the CP) on the fault classification accuracy would be interesting. Because, if slight geometric variations affect the algorithm it becomes challenging to identify healthy or fault CP states.
- In this study for the intermediate speed fault classifications, a range of 10 Hz has been taken between two training data sets. It would be interesting to see how the reduction or increment of this range would affect the identification of faults at intermediate speeds.
- Not always can the sensors be placed at the desired locations on the CP. Therefore, it would be significant to see if this methodology could be extended in such a way that vibration measurements on the induction motor can be directly used to identify the CP fault state.
- All through this research steady state CP fault identification was done. However, it will be interesting to see if the run-up or run-down data (producing transients) would be useful to identify CP faults.

- This methodology can be extended to developing a system level fault identification technique (ex. Motor and CP together).
- Other emerging machine learning techniques like the deep learning techniques or hybrid machine learning algorithms can be adapted to carry out the fault identifications.
- Measures to implement predictive maintenance can be done by studying the development and progress of the fault



Appendix A: Centrifugal Pump Applications

Table A.1: Various applications in which CPs are used and their salient features

CP	Type of CP	R (kW)	Applications and requirements
Universal CP	Single-stage radial	2 – 200	Low cost
Process CP	Single-stage radial	10 - 300	High reliability, low leakage, safety
	Multistage radial or semi-axial	50 – 1,000	
CPs for central power stations cooling systems	Single-stage radial or semi-axial	500 – 3,000	High N_s vertical CP
CP for boiler feed	Multistage radial	100 – 2,000	Power generation
CP in power station for boiler feed		5000 - 45,000	High speed machines. Designed specially to evade vibrations and cavitation
CP for injection		1000 - 20,000	Water injection to boost oil production
Pipeline CP			Long distance water or oil transport
Mine CP	Radial, semi-axial, axial	500 – 3,000	Mine drying, sandy water causes abrasive wear
Sewage CP		10 – 1,000	Huge stream passages, unaffected by blockage
Marine CP		1 – 1,000	Low space requirement, vertical inline, low NPSH
Submerged CP		5 - 500	Sewage, removal of water
Gas-liquid mixture	Single-stage radial	10 – 5,000	Process industry, oil production
Seal-less CP		5 - 250	Toxic fluids, canned motor or magnetic coupling
Dredge CP	Single-stage radial	200 – 5,000	Large flow passages, protection against abrasion
Slurry CP	Single-stage radial	50 – 1,000	Fortification against abrasion
Rocket CP	Radial or semi-axial	1,000	Extreme tip speed, short CP life
Food industry	All	1 - 50	Extreme cleanliness, protection of process fluid, no ingress of oil
Medical services	All	<0.1	Extreme cleanliness and reliability, protection of process fluid



Appendix B: Time-domain plots of various CP faults high sampling rate

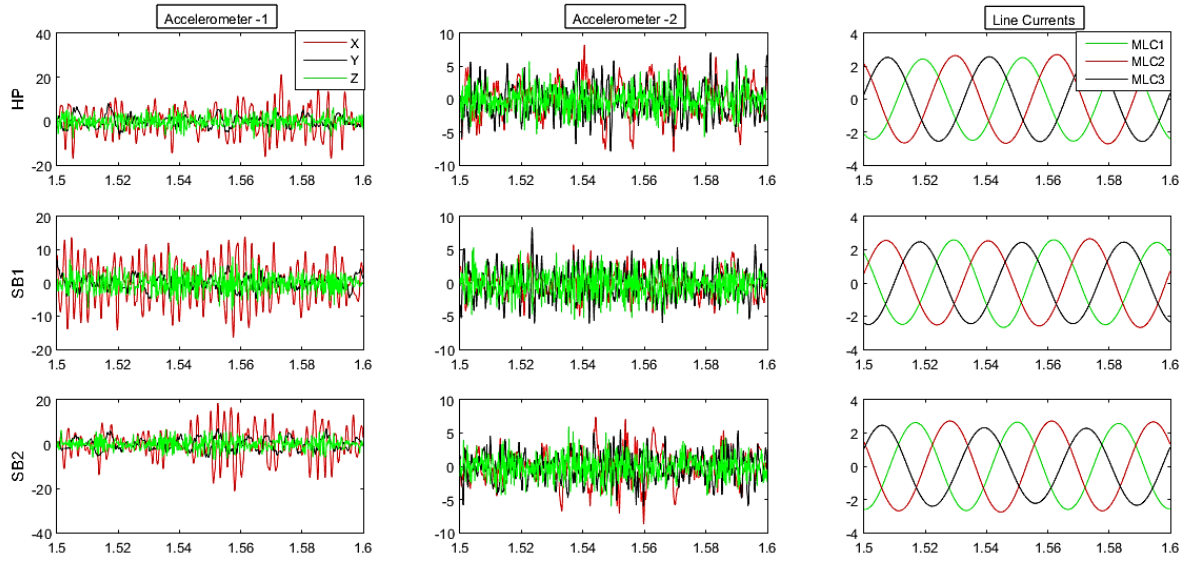


Figure B.1: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for HP, SB1 and SB2 faults.

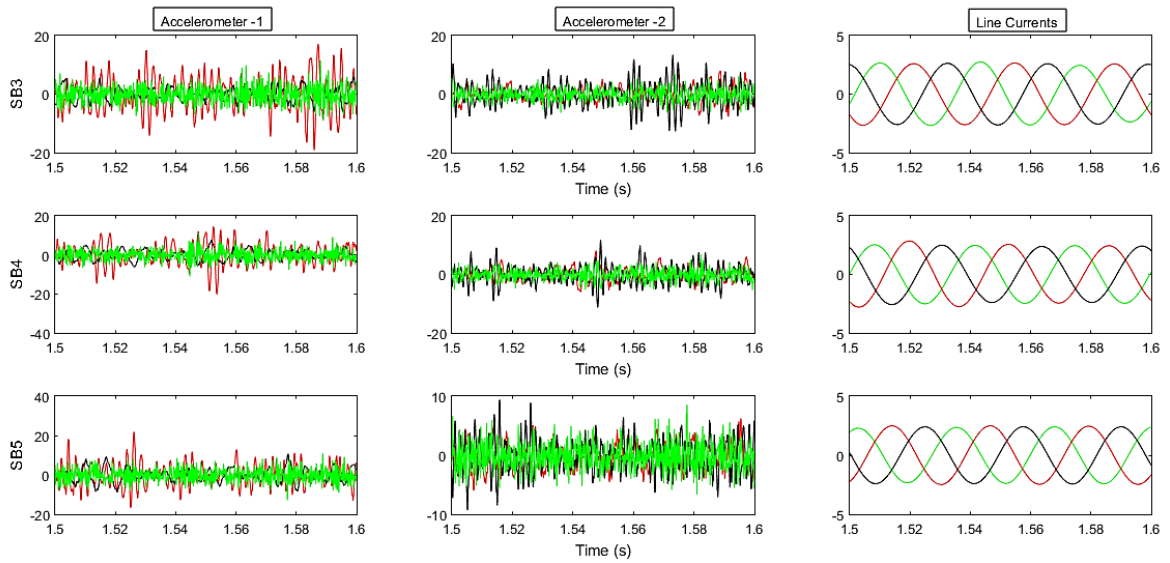


Figure B.2: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for SB3, SB4 and SB5 faults.

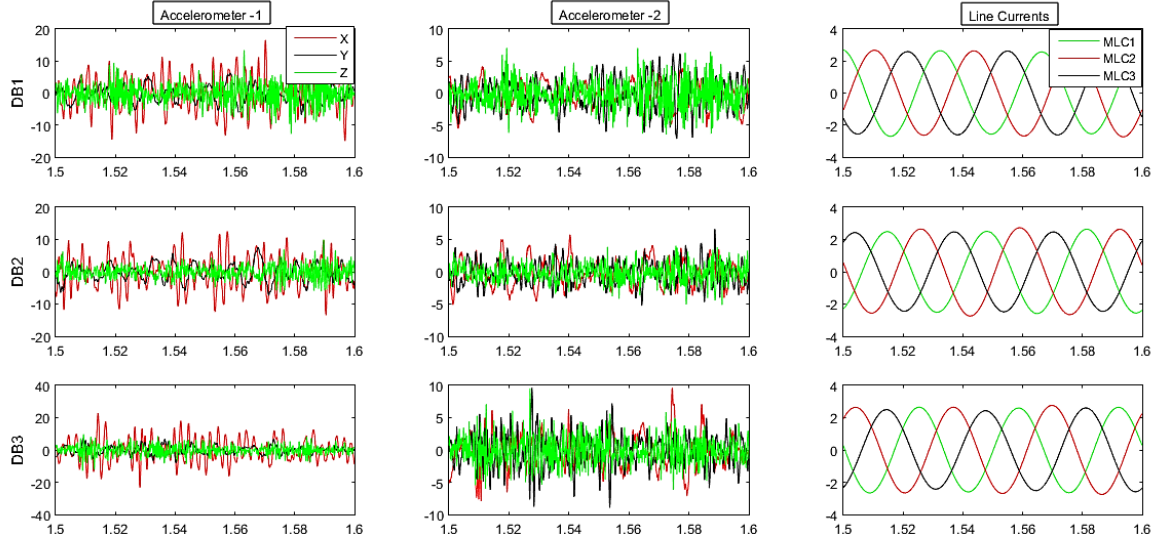


Figure B.3: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for DB1, DB2 faults.

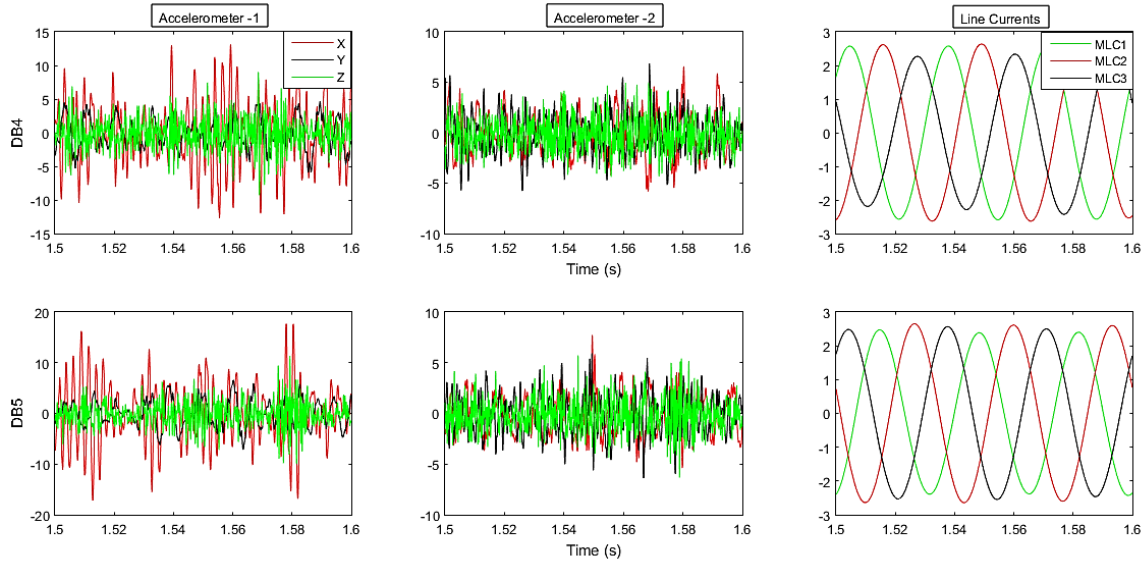


Figure B.4: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for DB3, DB4, DB5 faults.

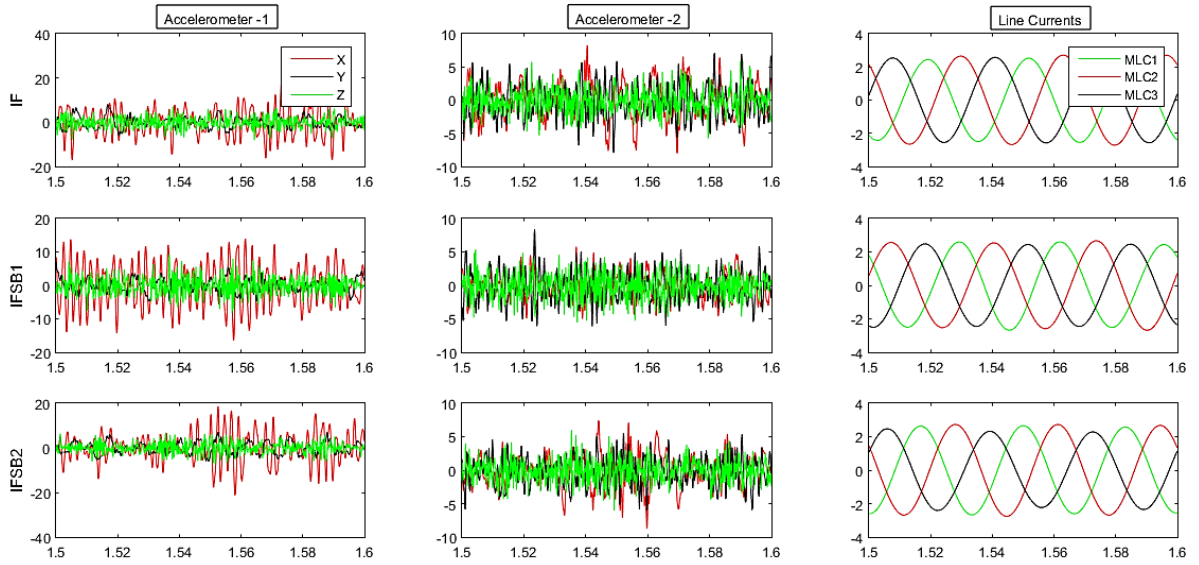


Figure B.5: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for IF, IFSB1, IFSB2 faults

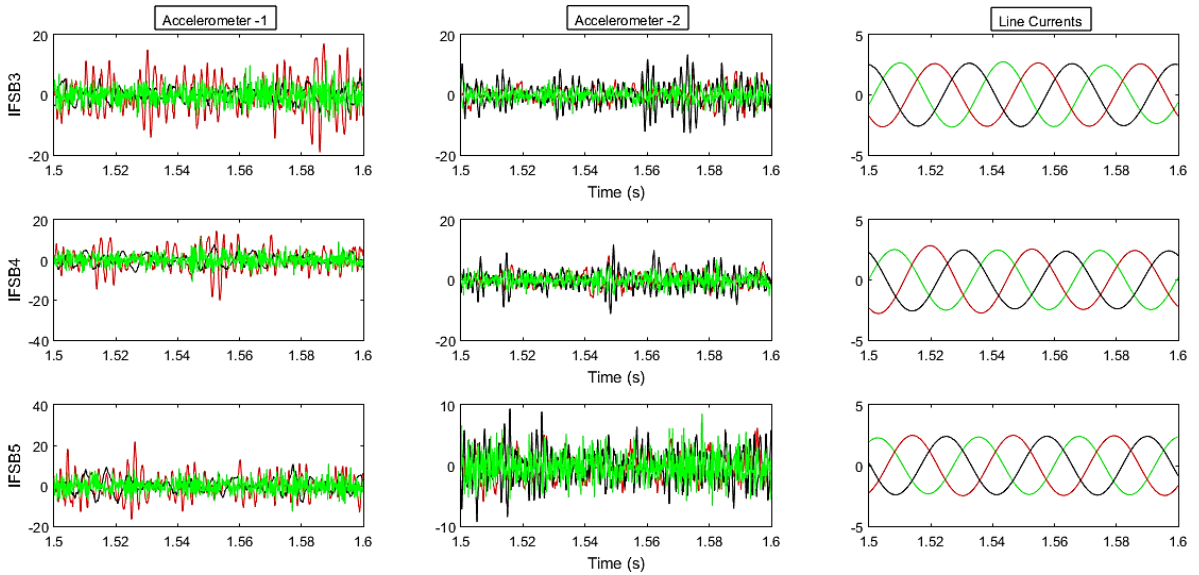


Figure B.6: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for IFSB3, IFSB4 and IFSB5 faults.

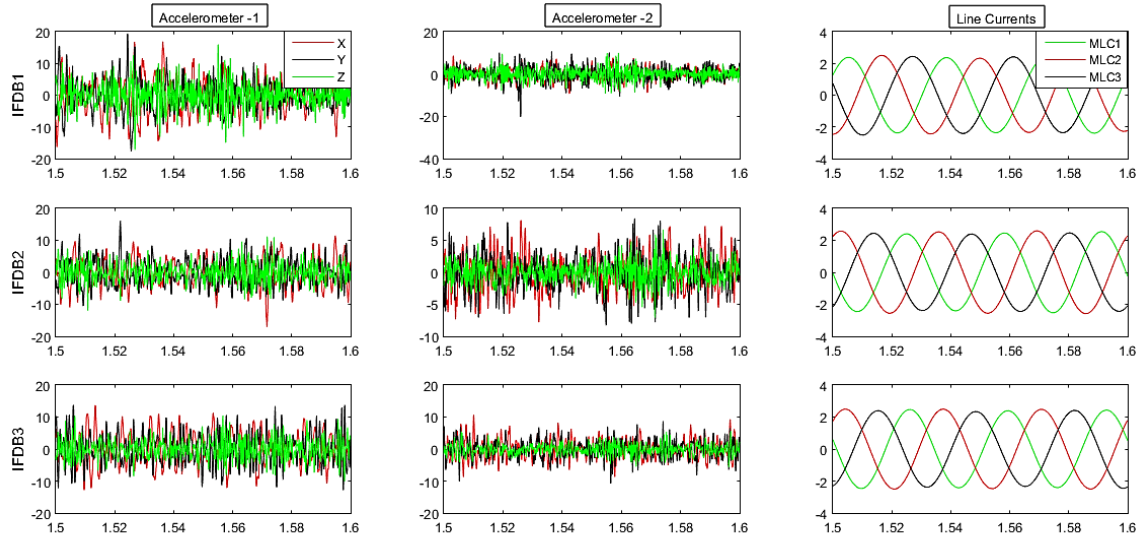


Figure B.7: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for IFDB1, IFDB2 faults.

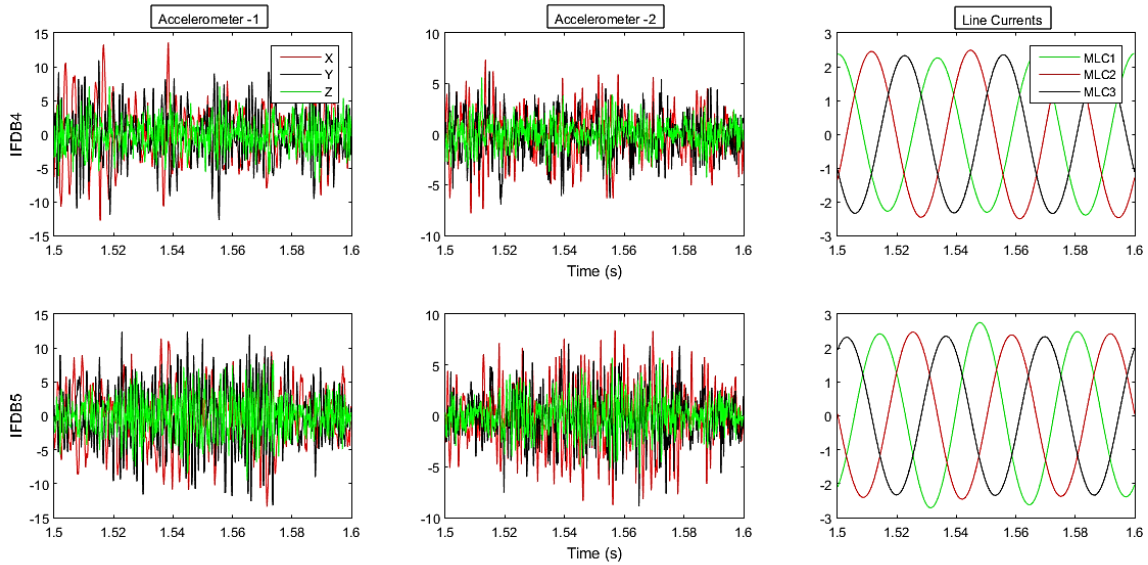


Figure B.8: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for IFDB3, IFDB4, IFDB5 faults.

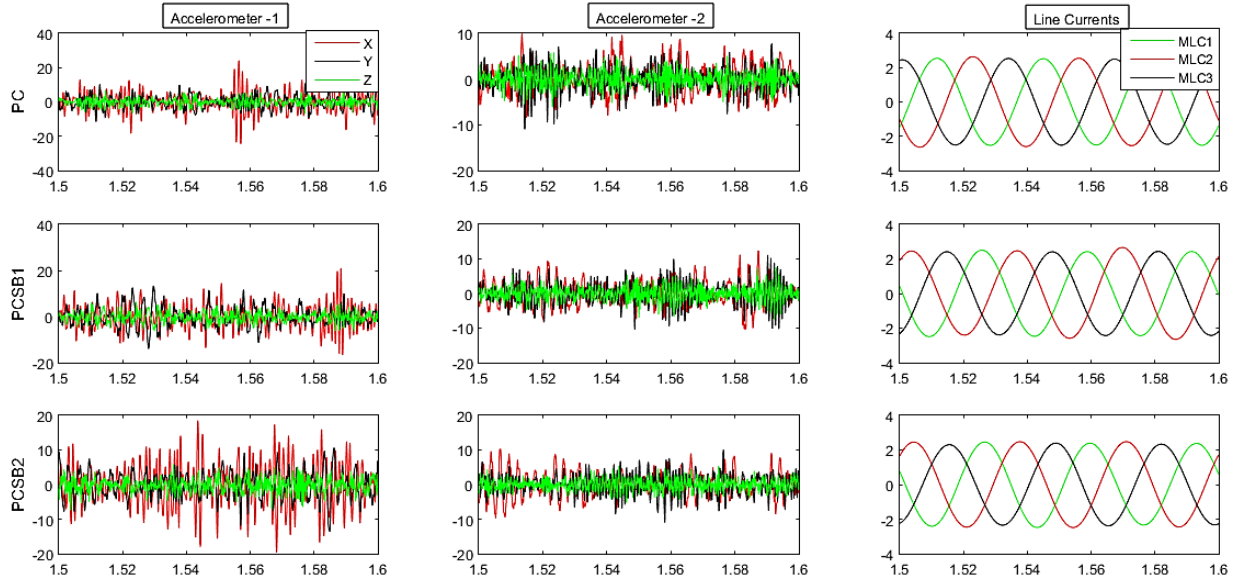


Figure B.9: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for PC, PCSB1, PCSB2 faults.

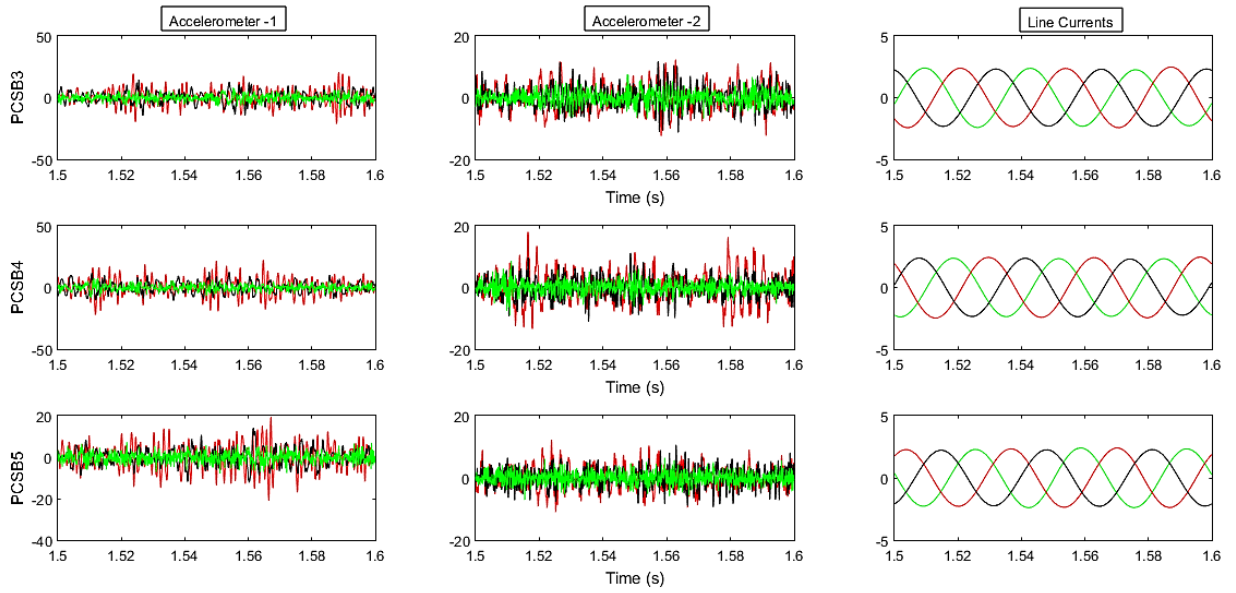


Figure B.10: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for PCSB3, PCSB4 and PCSB5 faults.

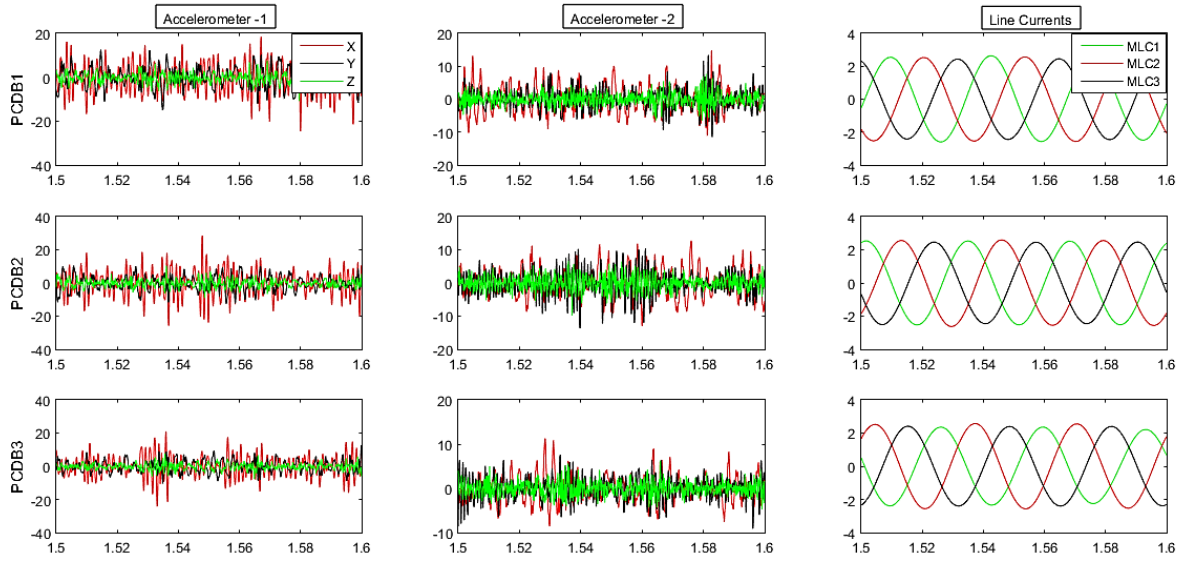


Figure B.11: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for PCDB1, PCDB2 faults.

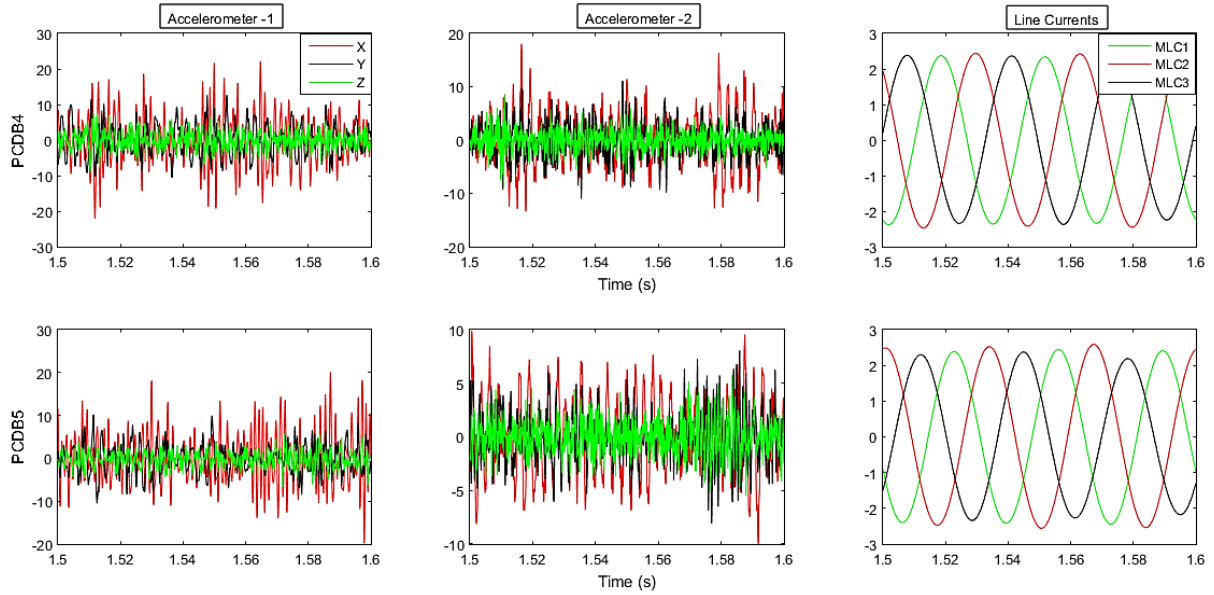


Figure B.12: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for PCDB3, PCDB4, PCDB5 faults.

Appendix C: Time-domain plots of various CP faults low sampling rate

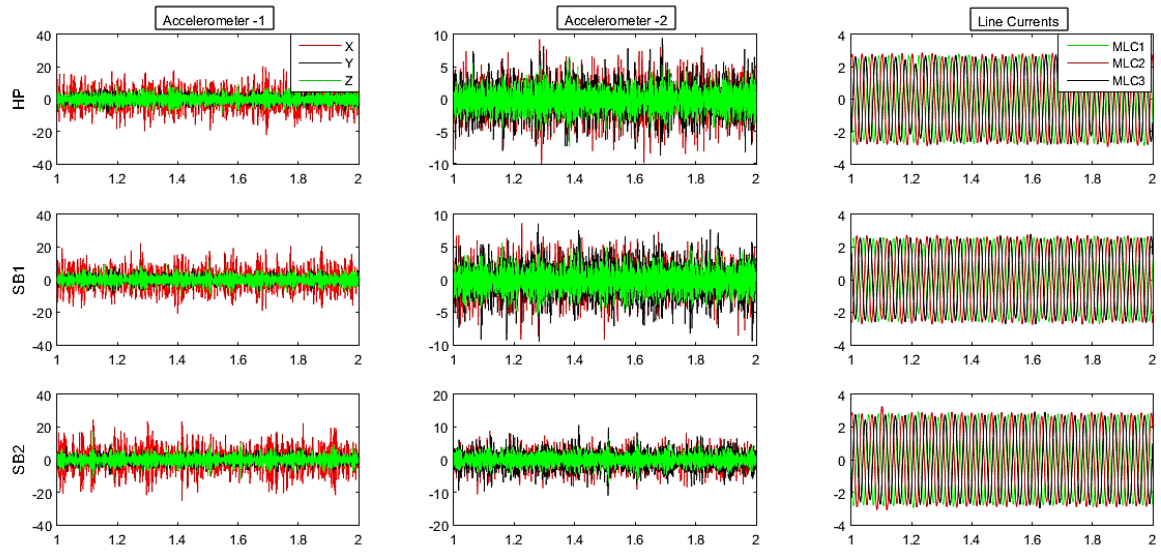


Figure C.1: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for HP, SB1 and SB2 faults.

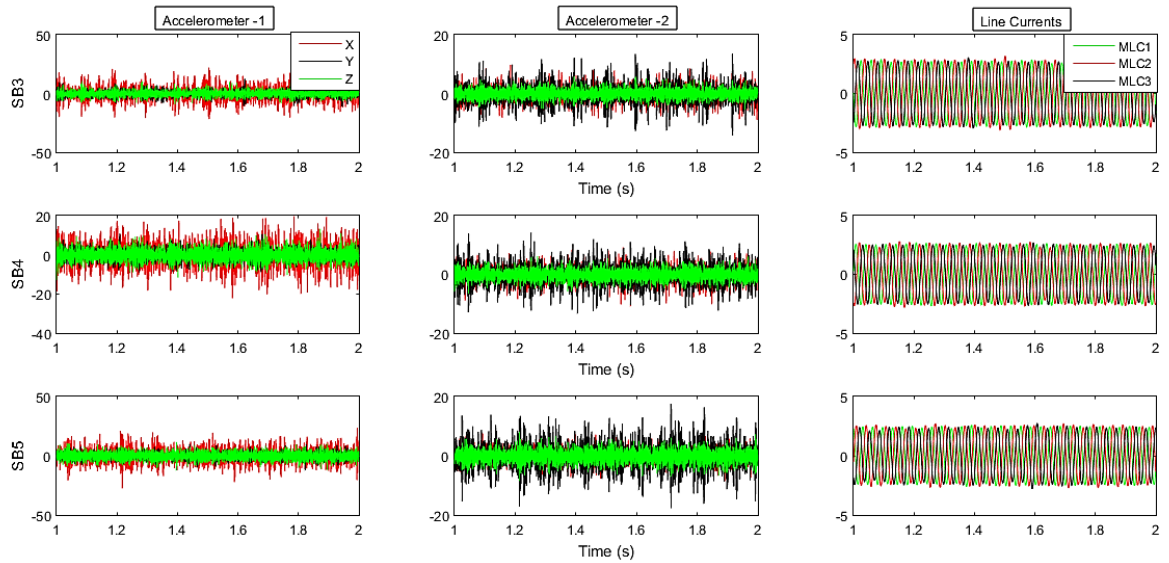


Figure C.2: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for SB3, SB4 and SB5 faults.

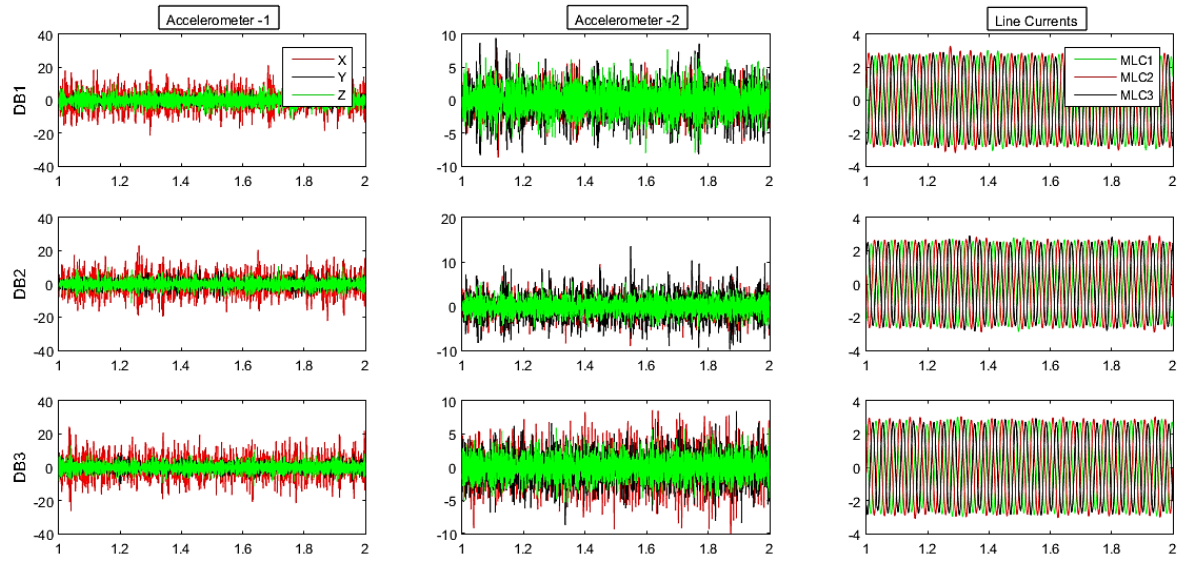


Figure C.3: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for DB1, DB2 faults.

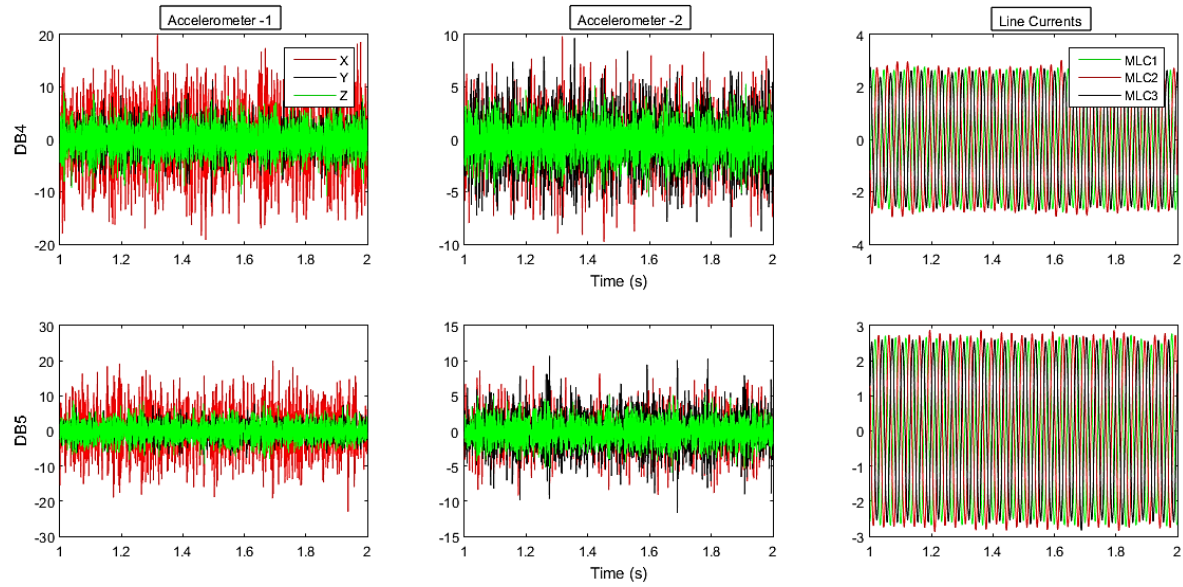


Figure C.4: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for DB3, DB4, DB5 faults.

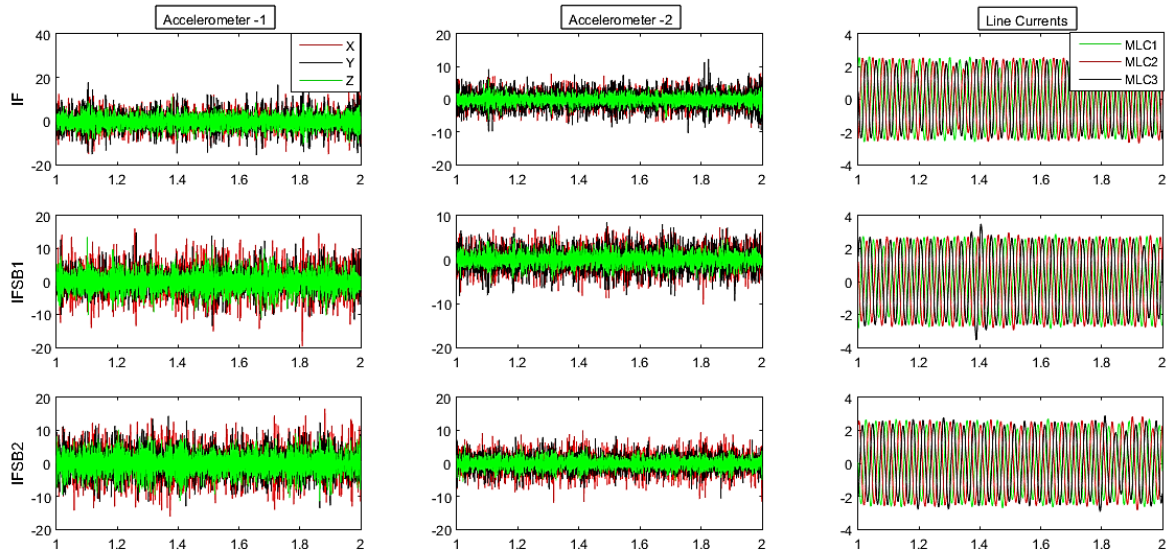


Figure C.5: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for IF, IFSB1, IFSB2 faults.

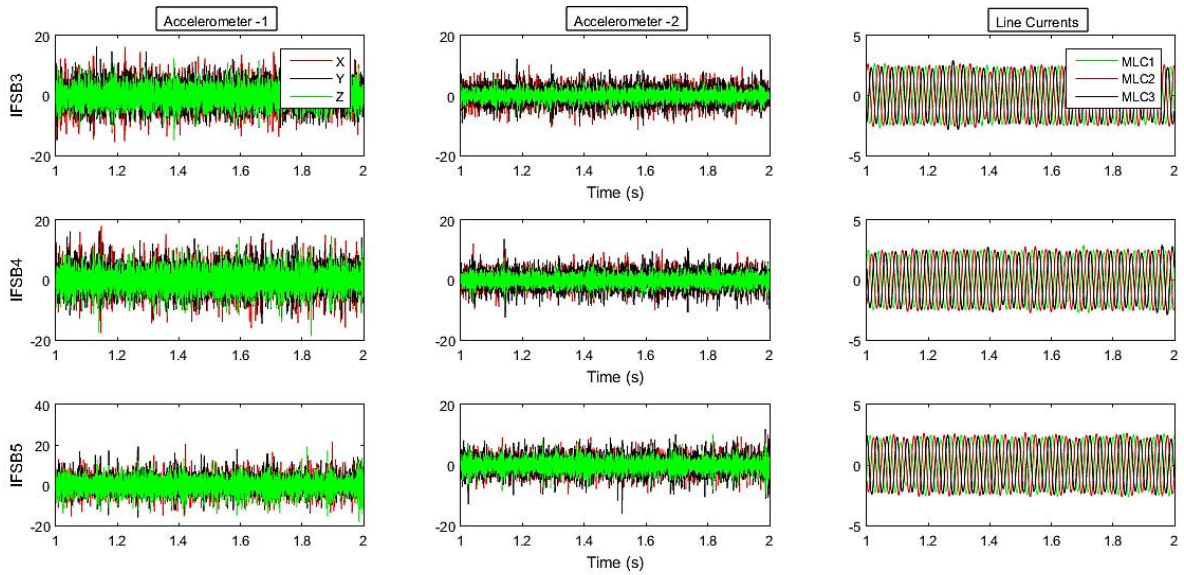


Figure C.6: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for IFSB3, IFSB4 and IFSB5 faults.

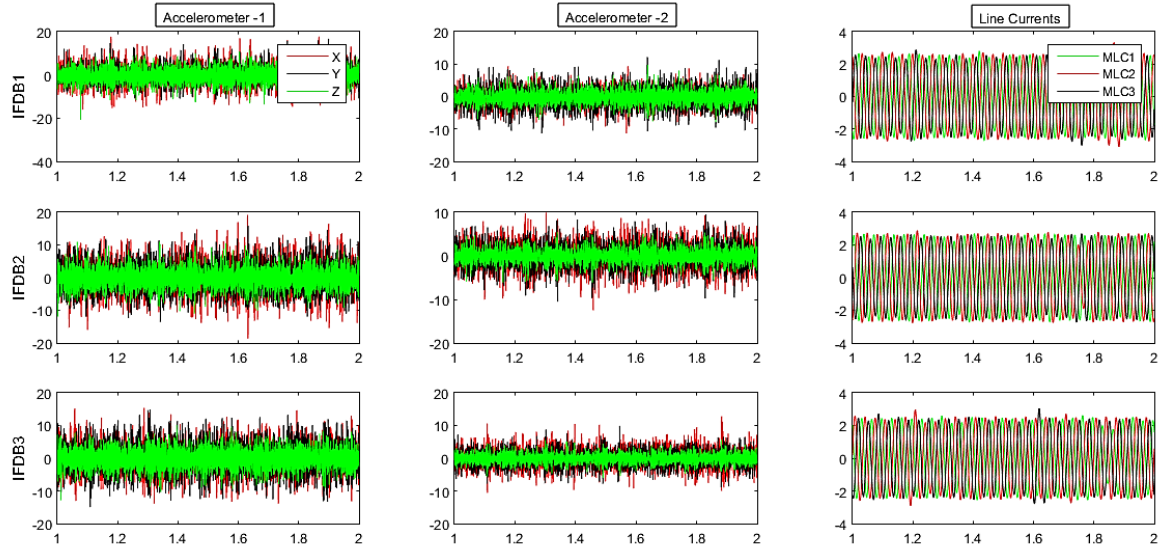


Figure C.7: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for IFDB1, IFDB2 faults.

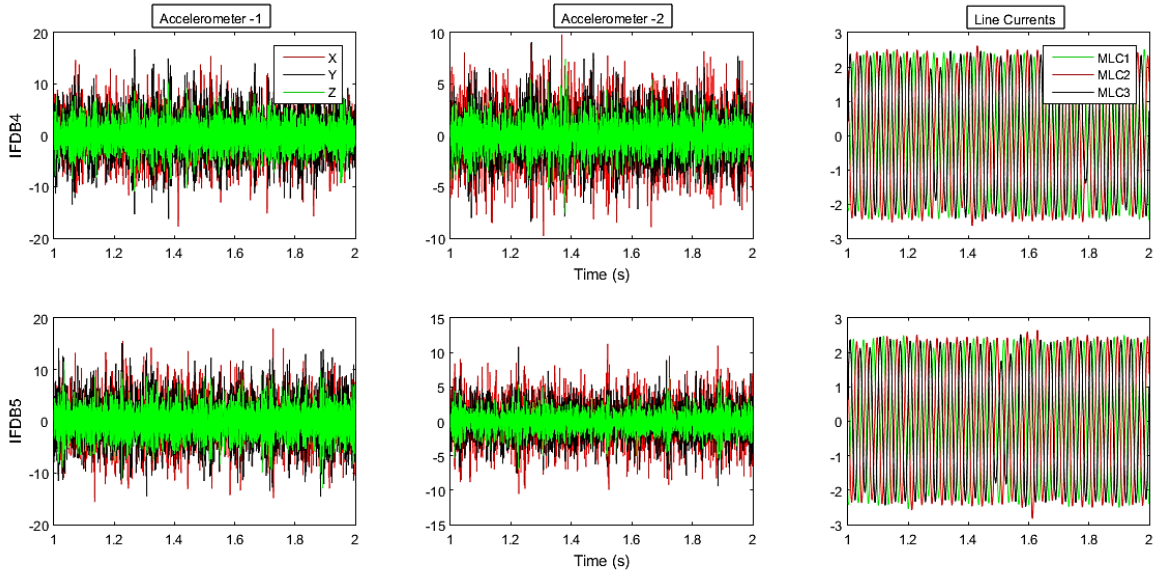


Figure C.8: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for IFDB3, IFDB4, IFDB5 faults.

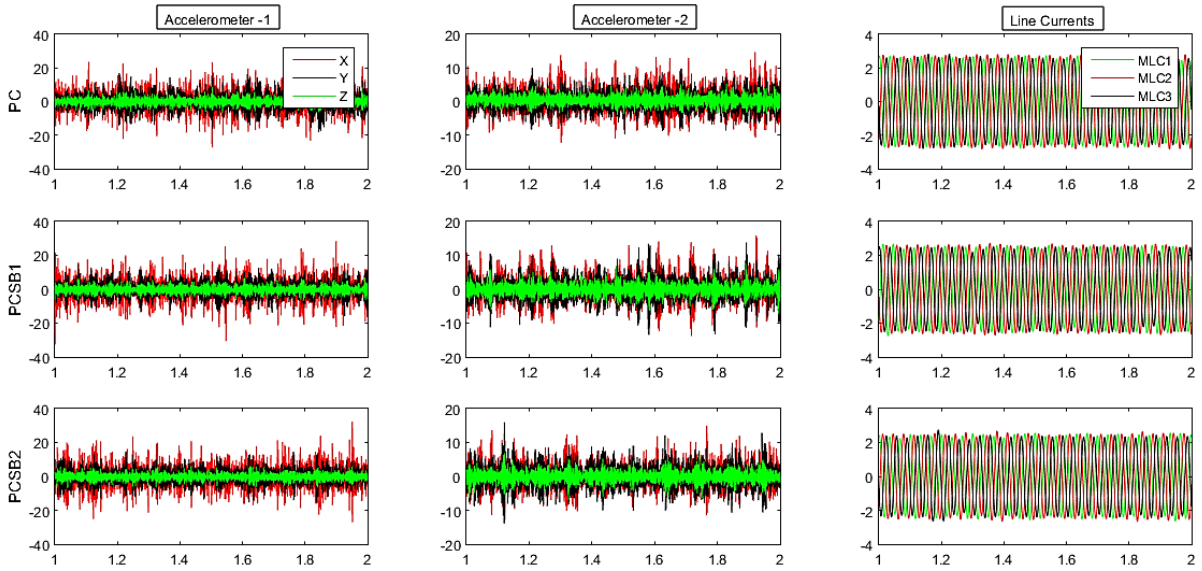


Figure C.9: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for PC, PCSB1, PCSB2 faults.

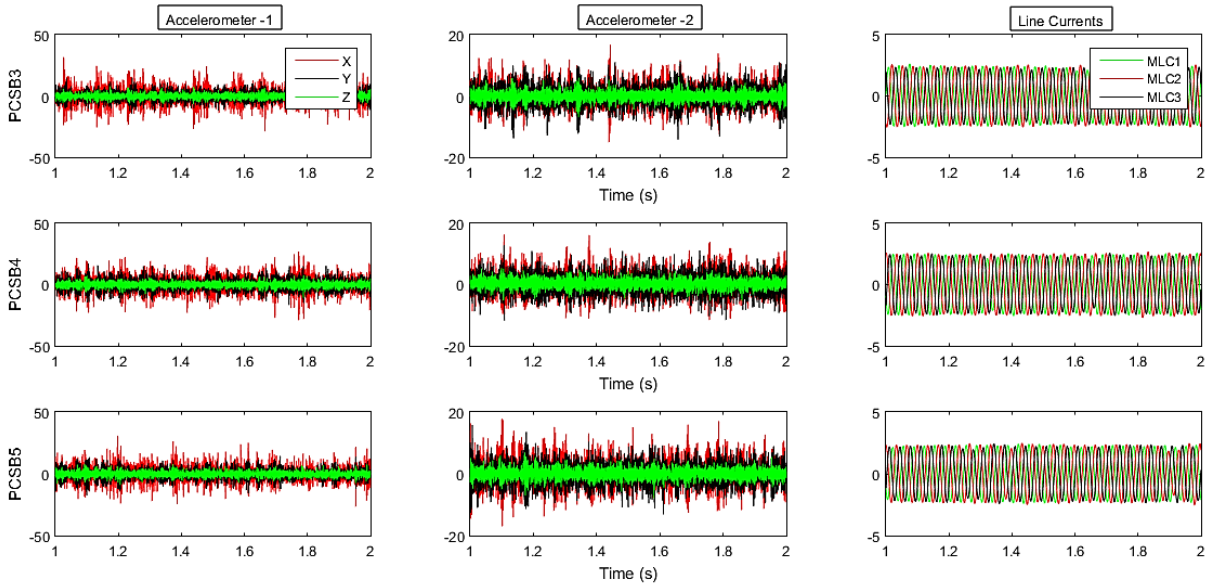


Figure C.10: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for PCSB3, PCSB4 and PCSB5 faults.

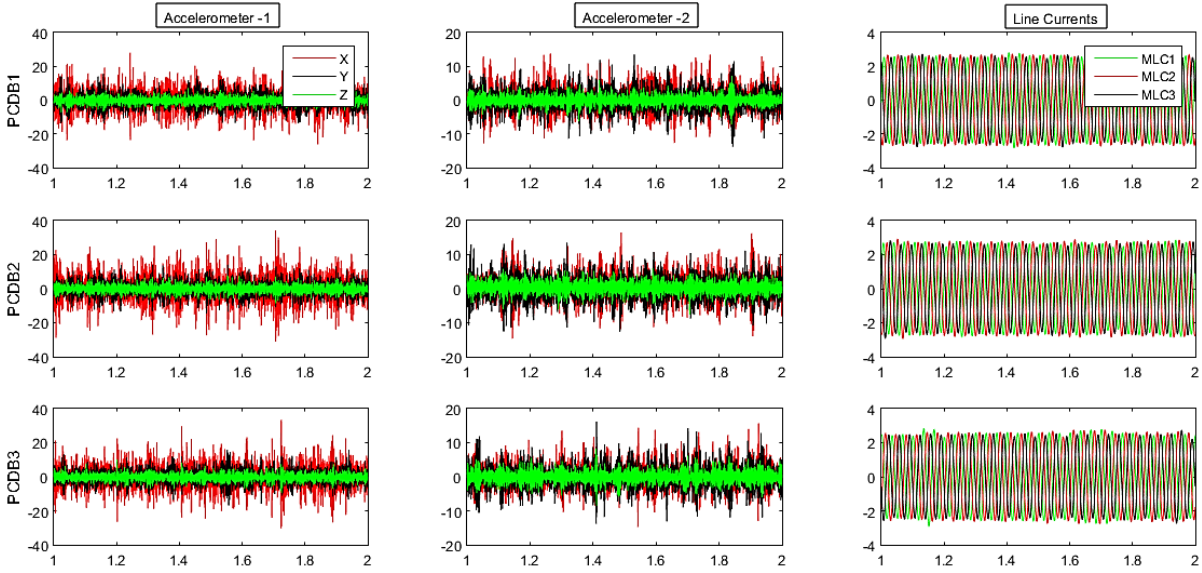


Figure C.11: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for PCDB1, PCDB2 faults.

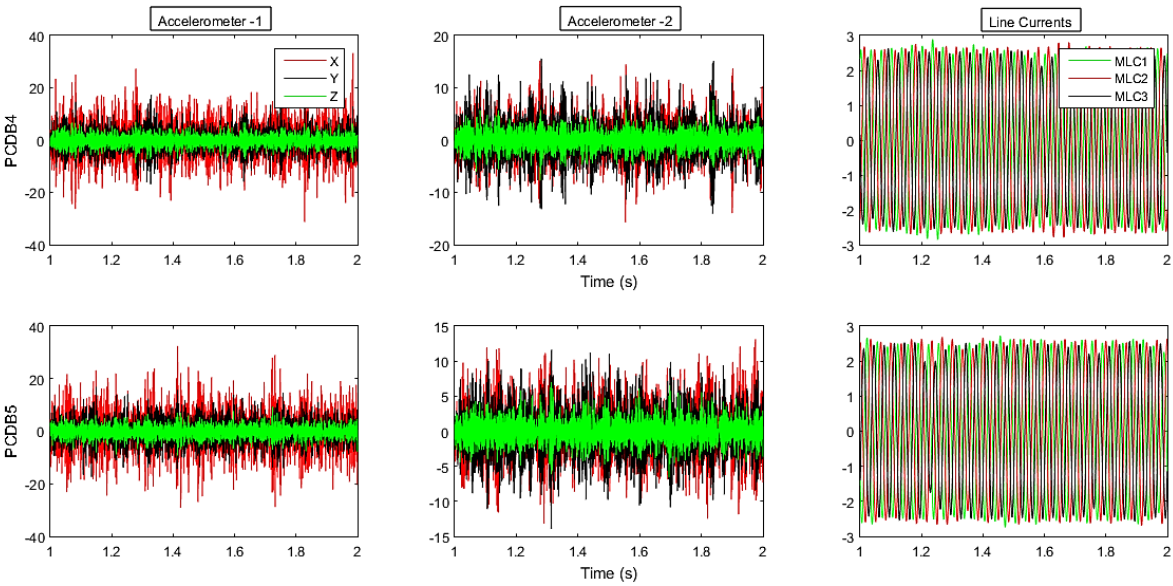


Figure C.12: Time domain amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for PCDB3, PCDB4, PCDB5 faults.

Appendix D: PSD plots of various CP faults high sampling rate

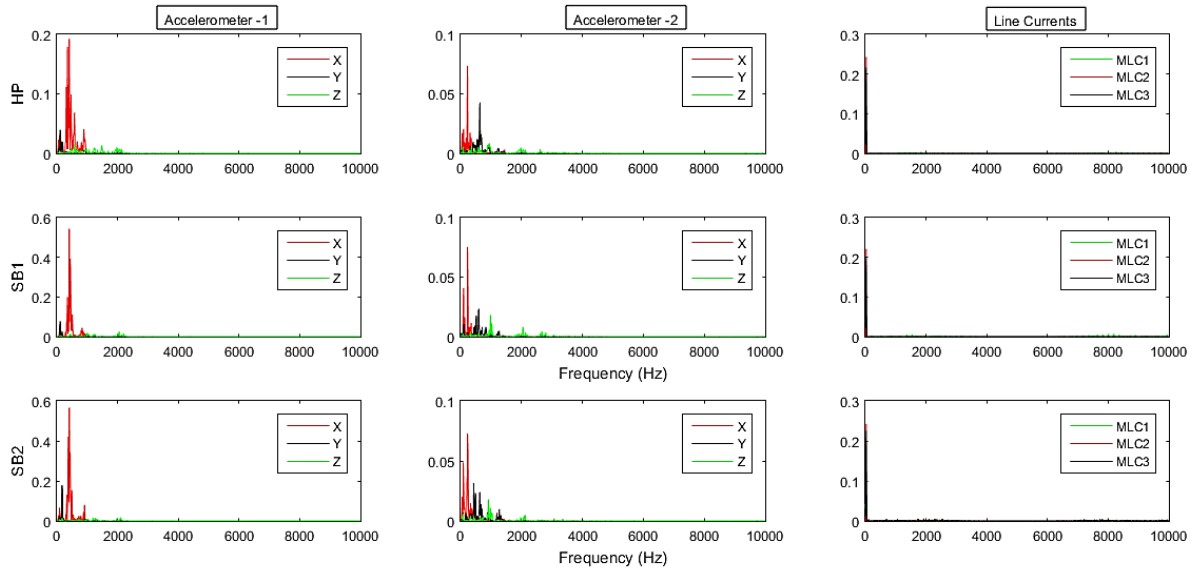


Figure D.1: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for HP, SB1 and SB2 faults.

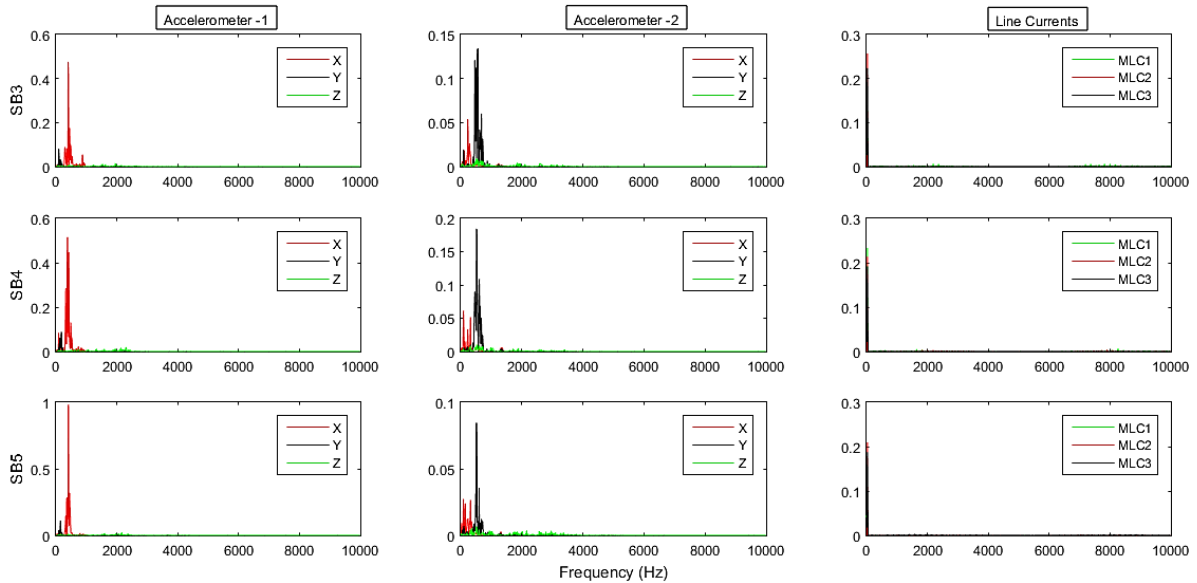


Figure D.2: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for SB3, SB4 and SB5 faults.

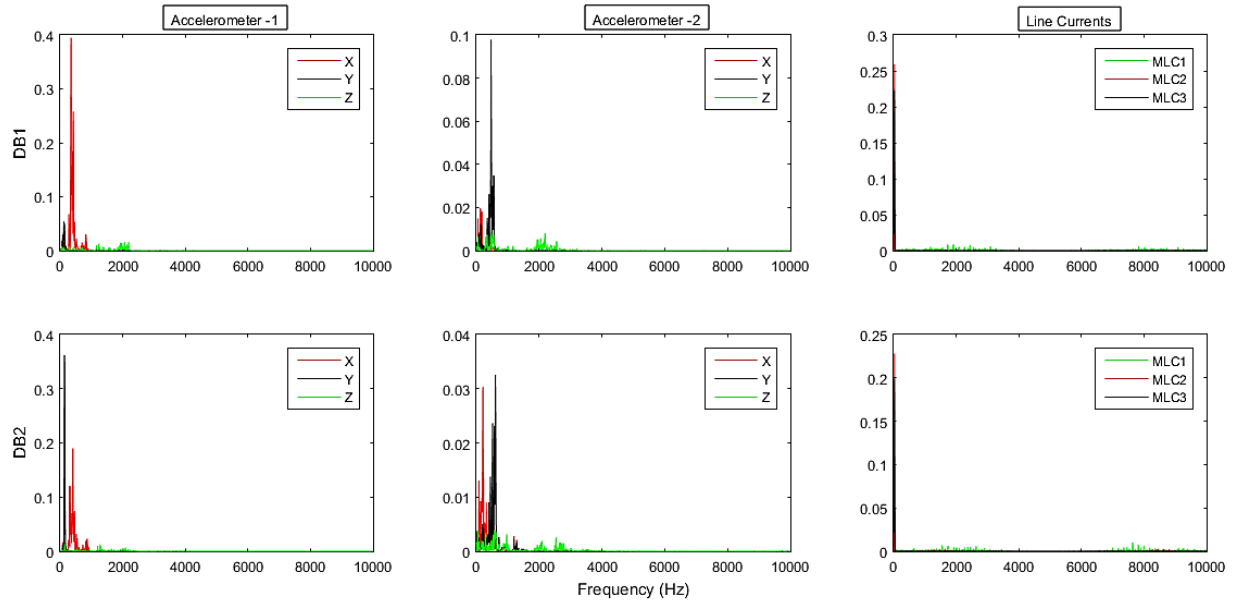


Figure D.3: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for DB1, DB2 faults.

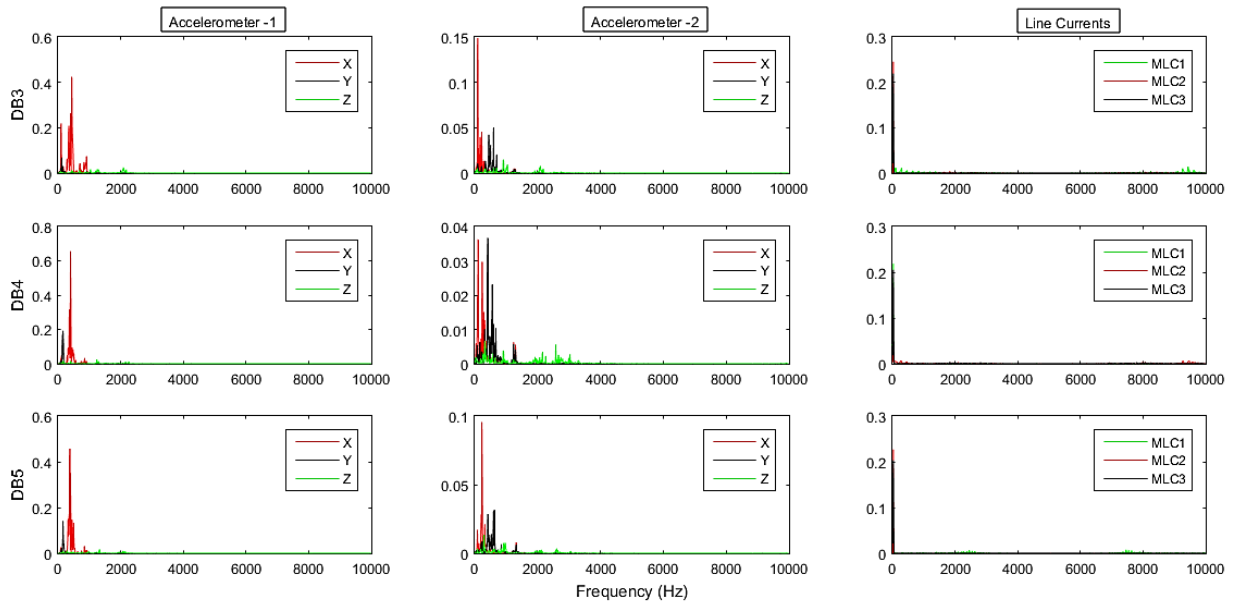


Figure D.4: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for DB3, DB4, DB5 faults.

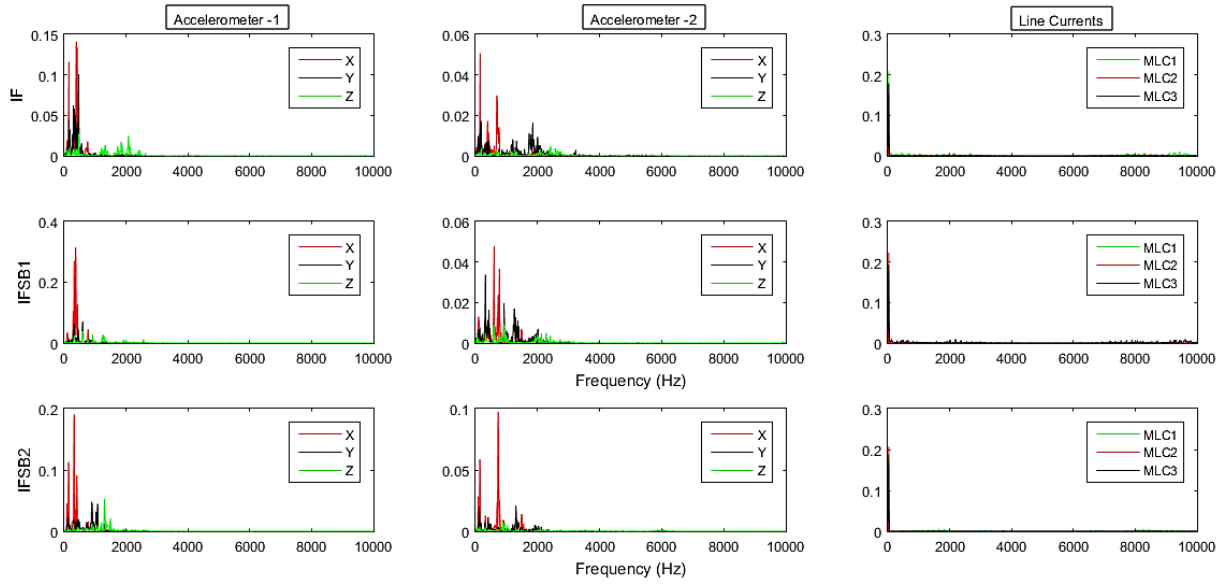


Figure D.5: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for IF, IFSB1, IFSB2 faults.

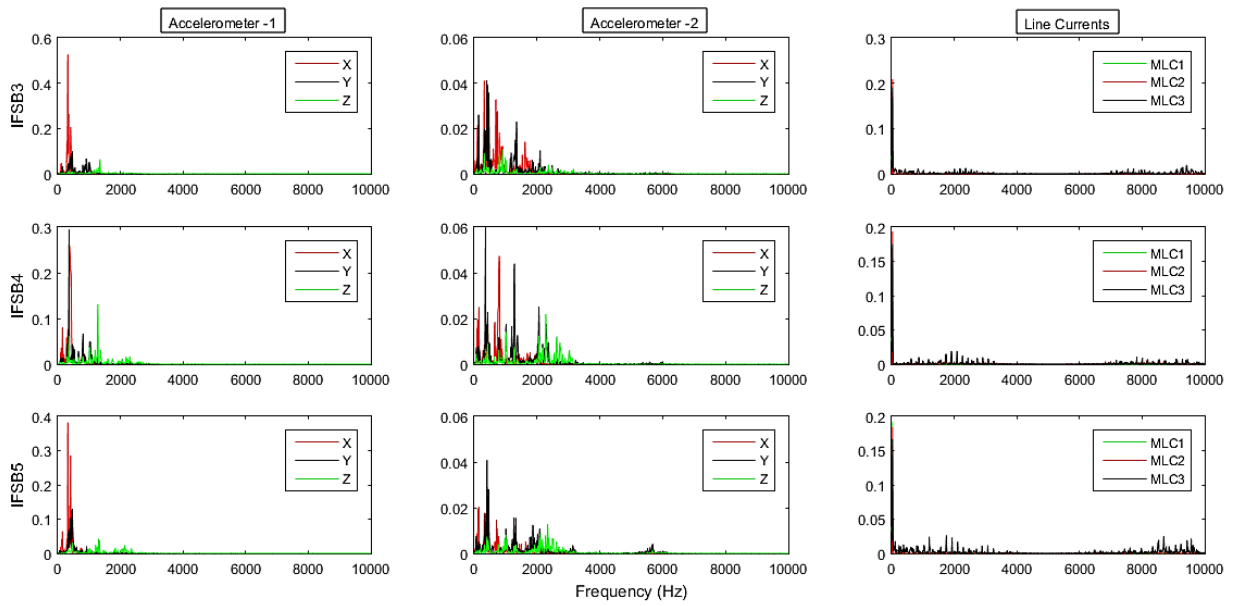


Figure D.6: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for IFSB3, IFSB4 and IFSB5 faults.

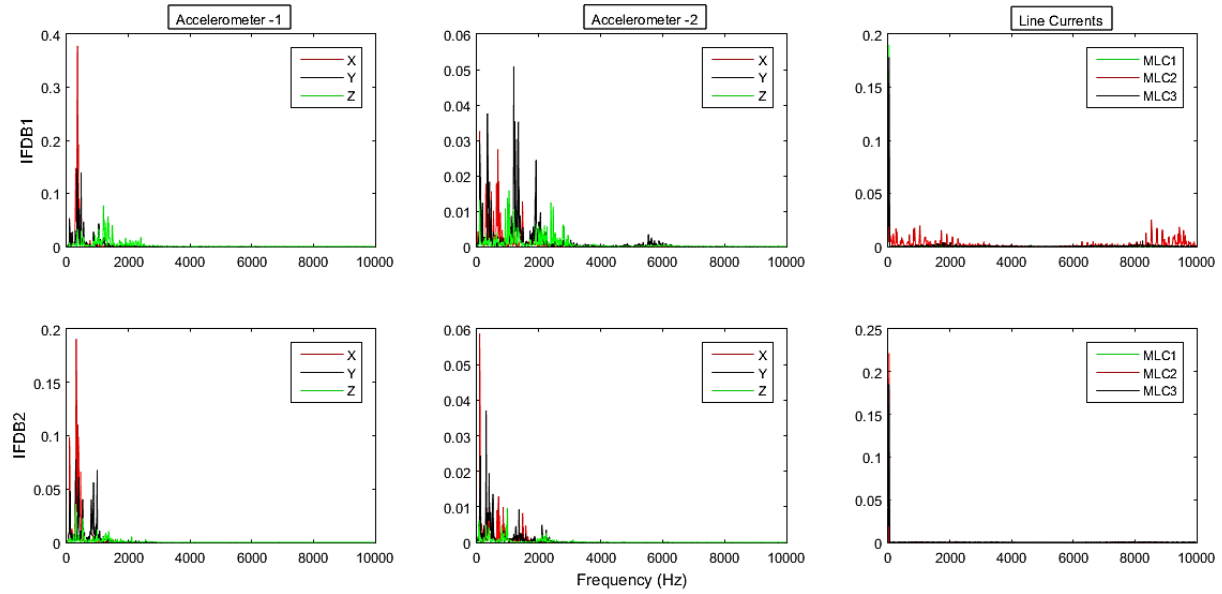


Figure D.7: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for IFDB1, IFDB2 faults.

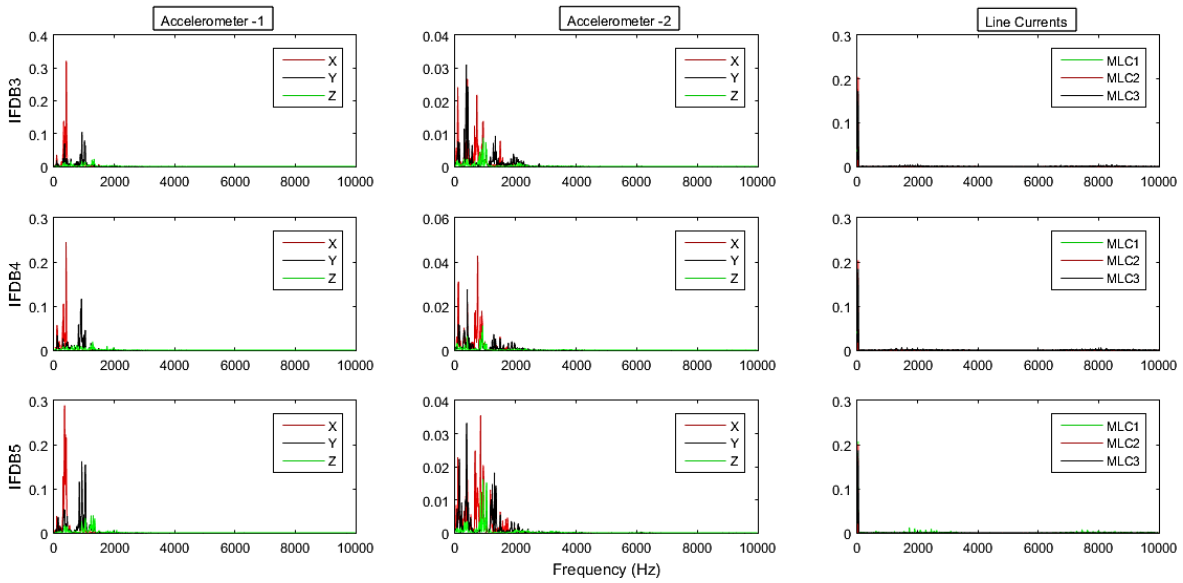


Figure D.8: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for IFDB3, IFDB4, IFDB5 faults.

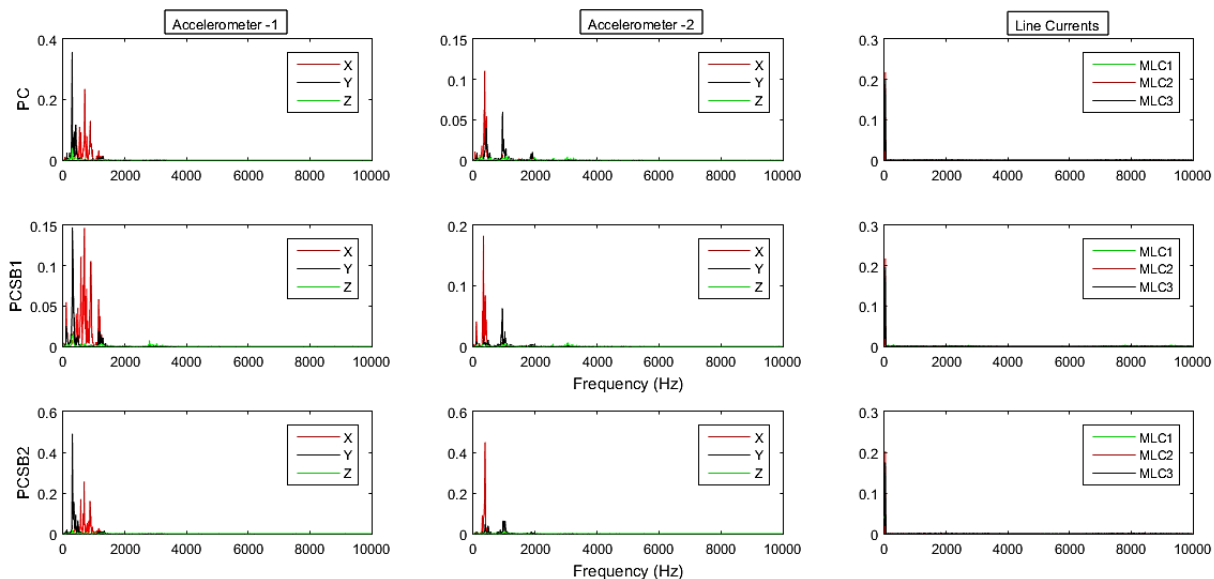


Figure D.9: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for PC, PCSB1, PCSB2 faults.

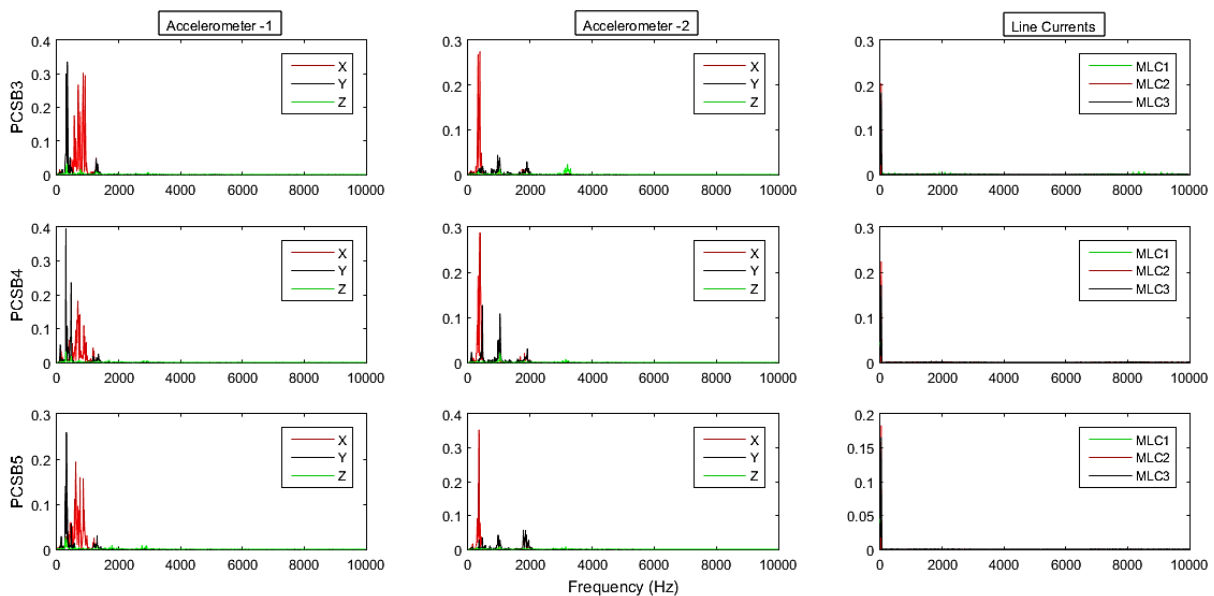


Figure D.10: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for PCSB3, PCSB4 and PCSB5 faults.

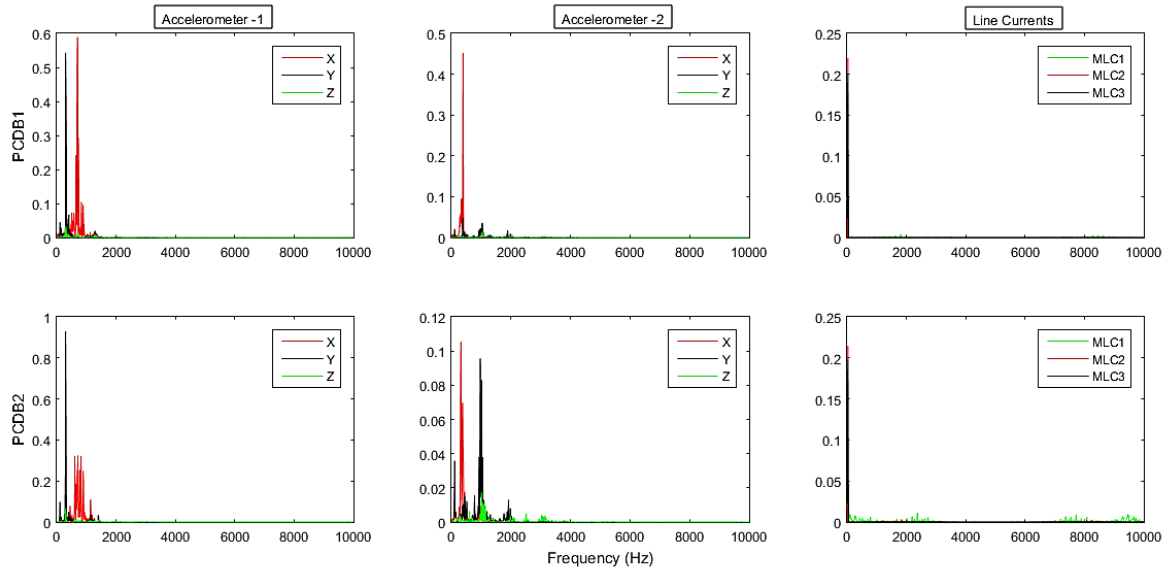


Figure D.11: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for PCDB1, PCDB2 faults.

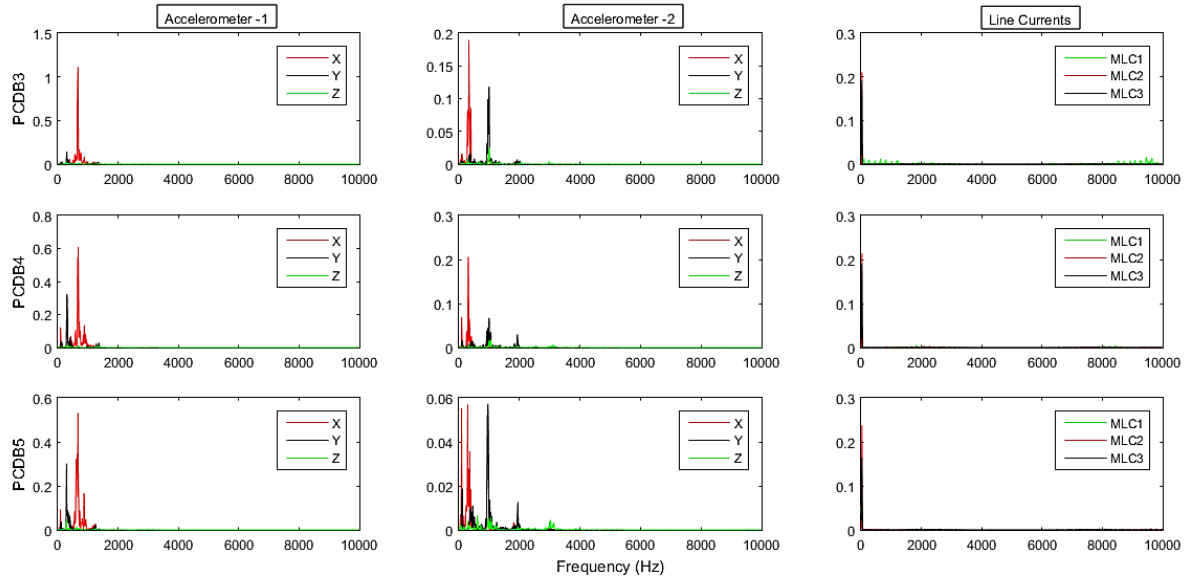


Figure D.12: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for PCDB3, PCDB4, PCDB5 faults.

Appendix E: PSD plots of various CP faults low sampling rate

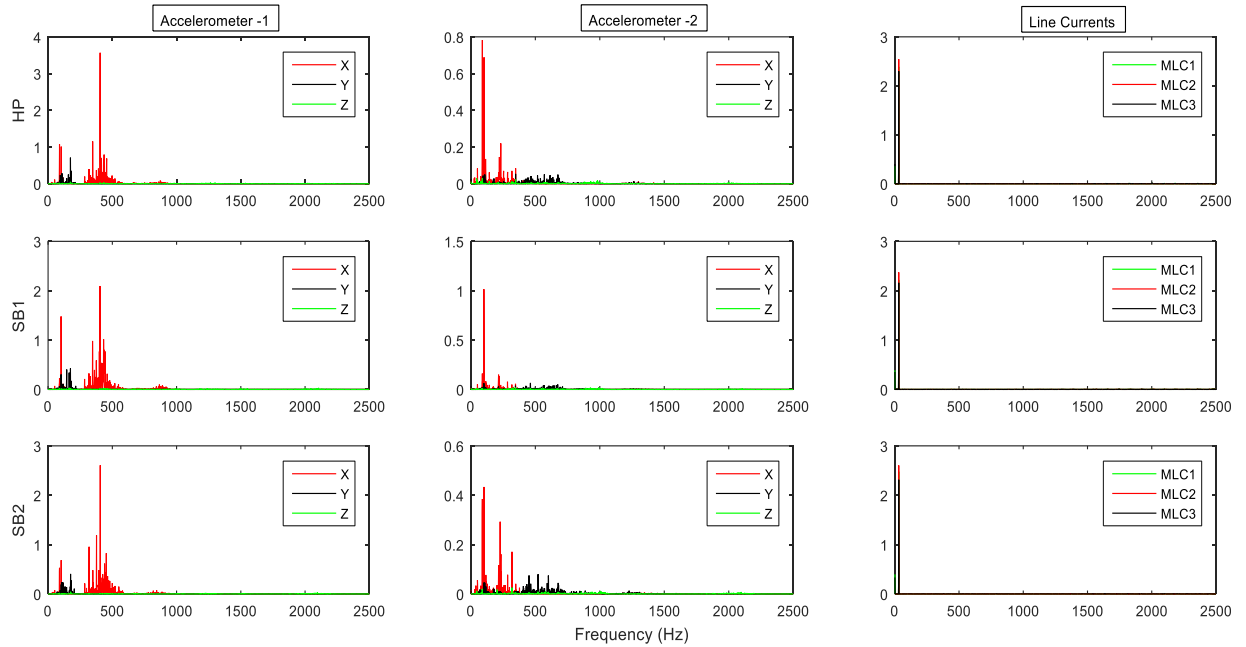


Figure E.1: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for HP, SB1 and SB2 faults.

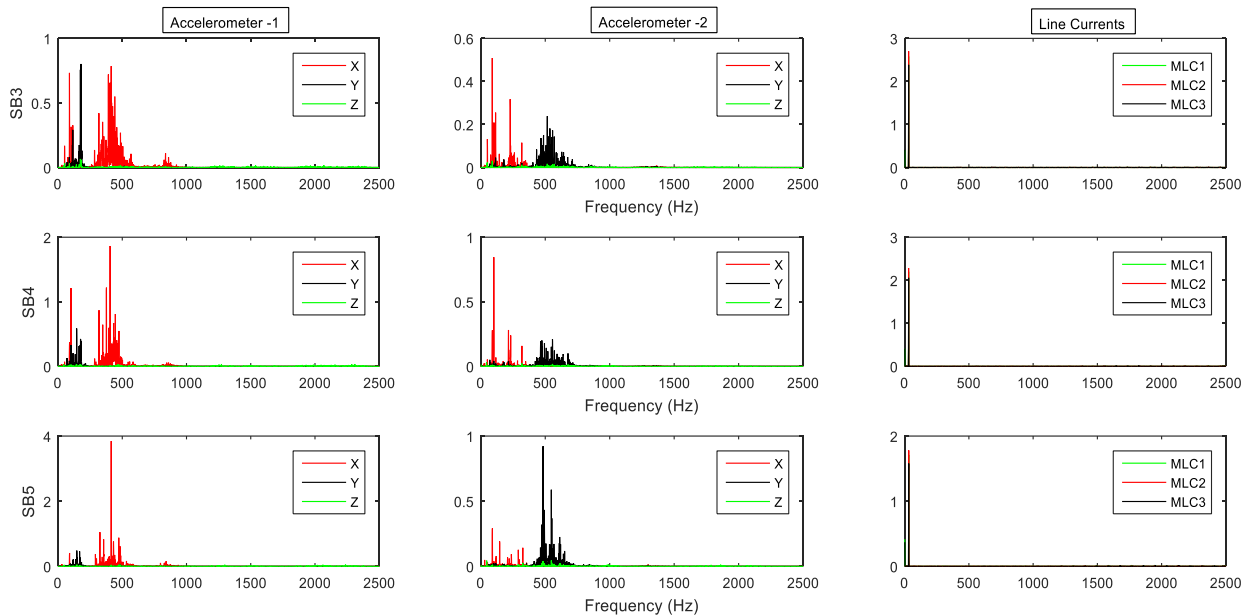


Figure E.2: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for SB3, SB4 and SB5 faults.

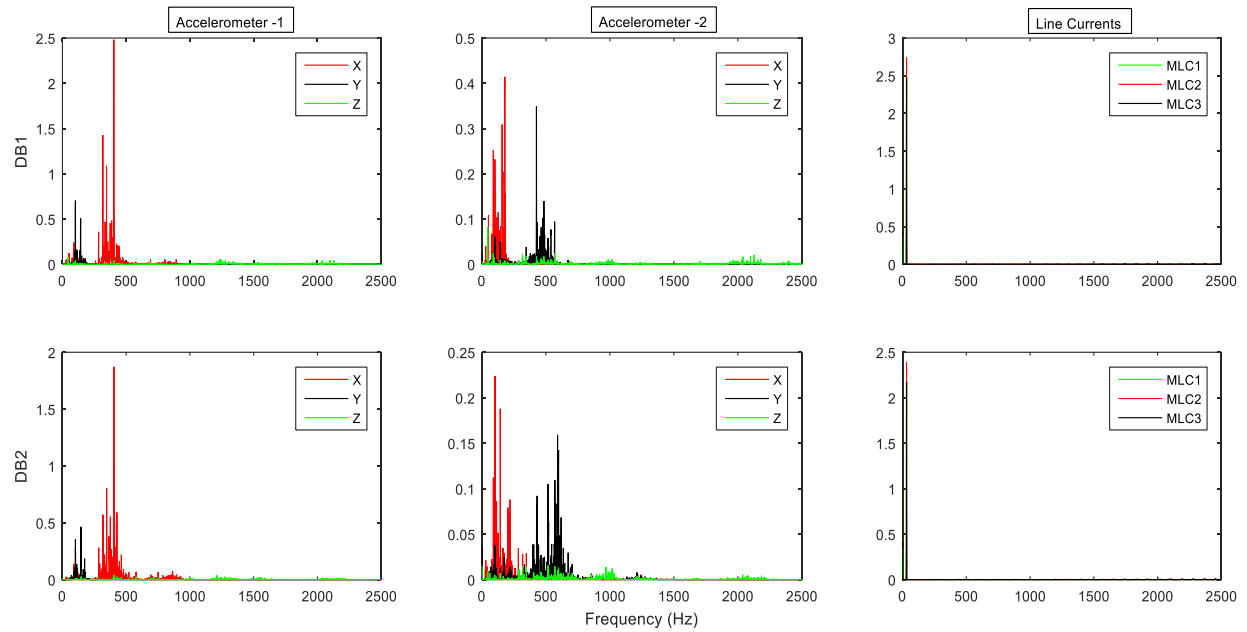


Figure E.3: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for DB1, DB2 faults.

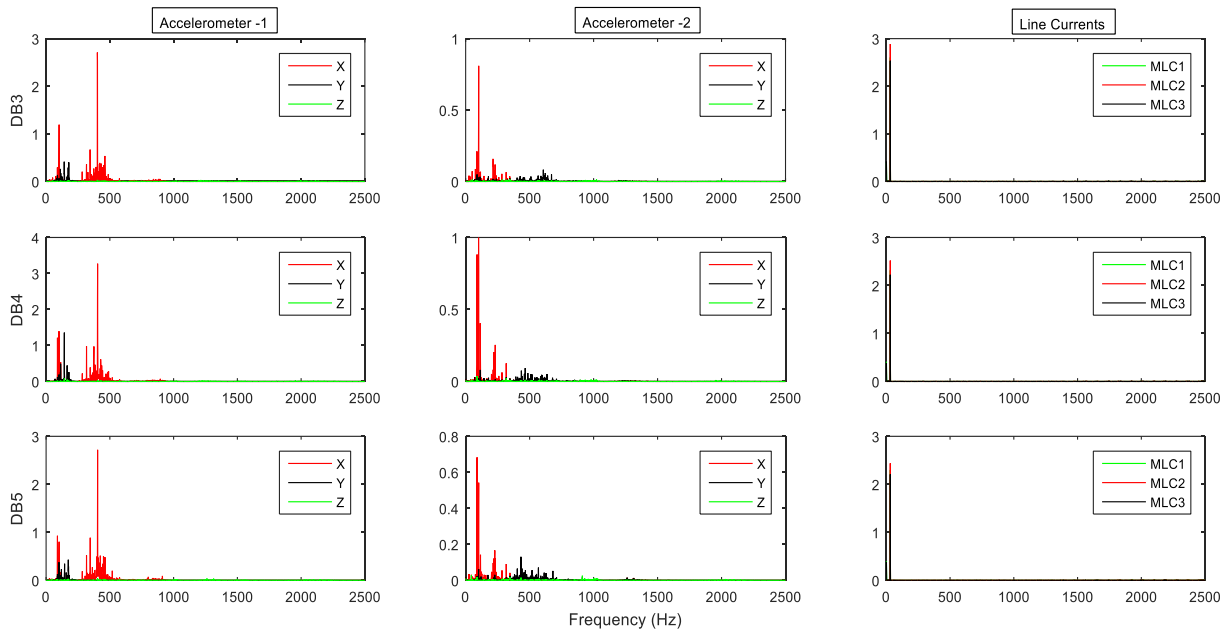


Figure E.4: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for DB3, DB4, DB5 faults.

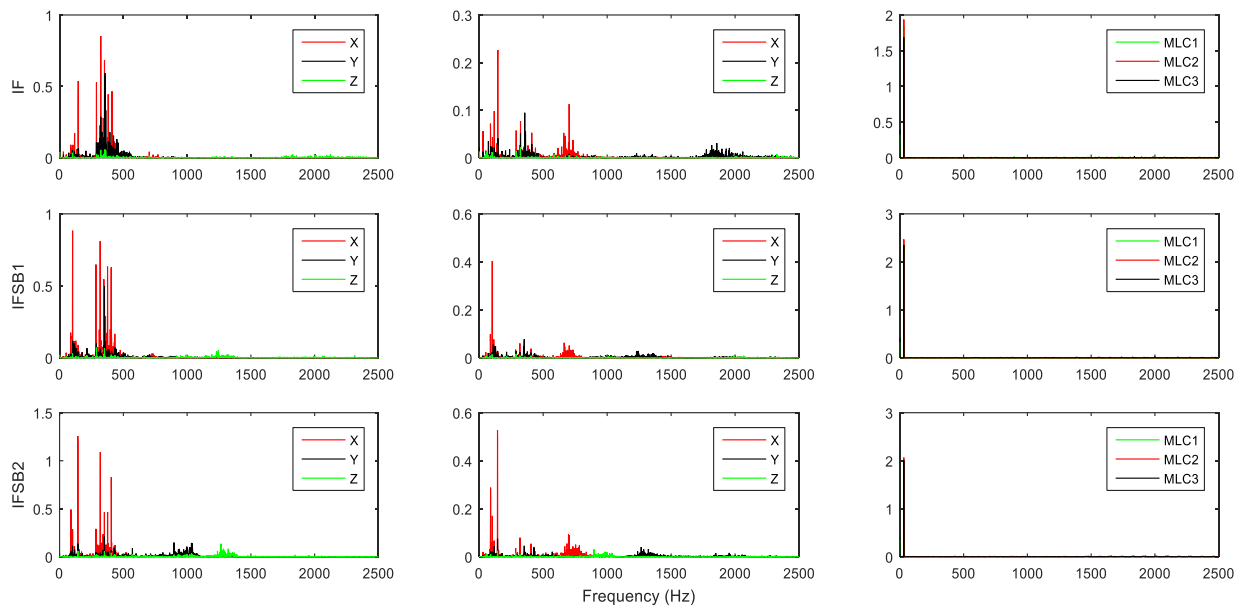


Figure E.5: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for IF, IFSB1, IFSB2 faults.

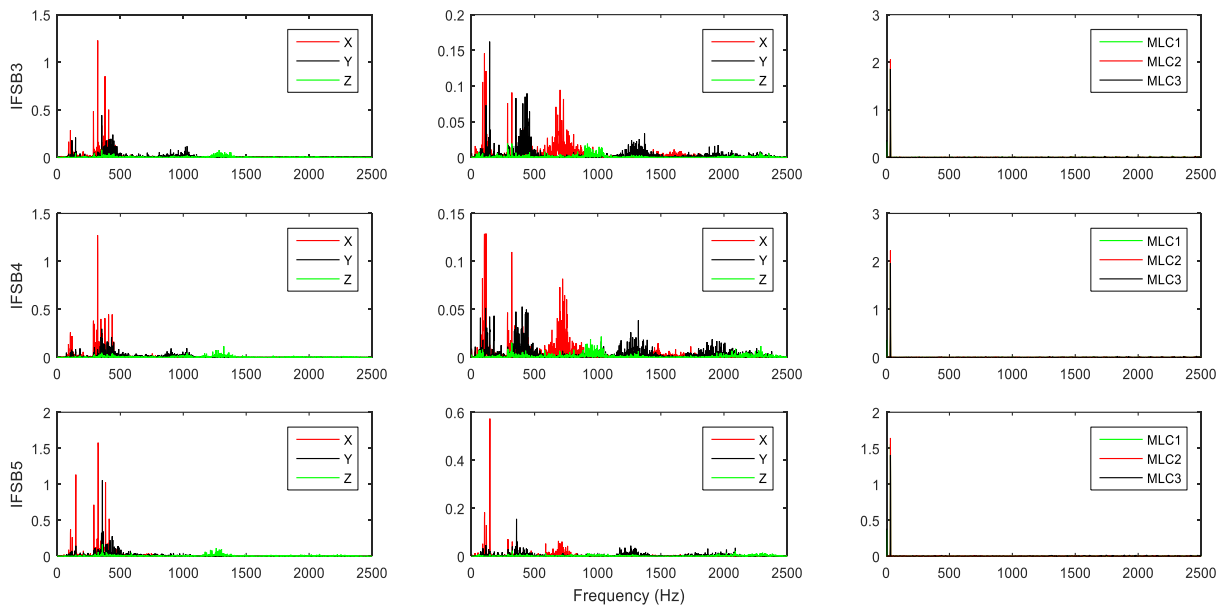


Figure E.6: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for IFSB3, IFSB4 and IFSB5 faults.

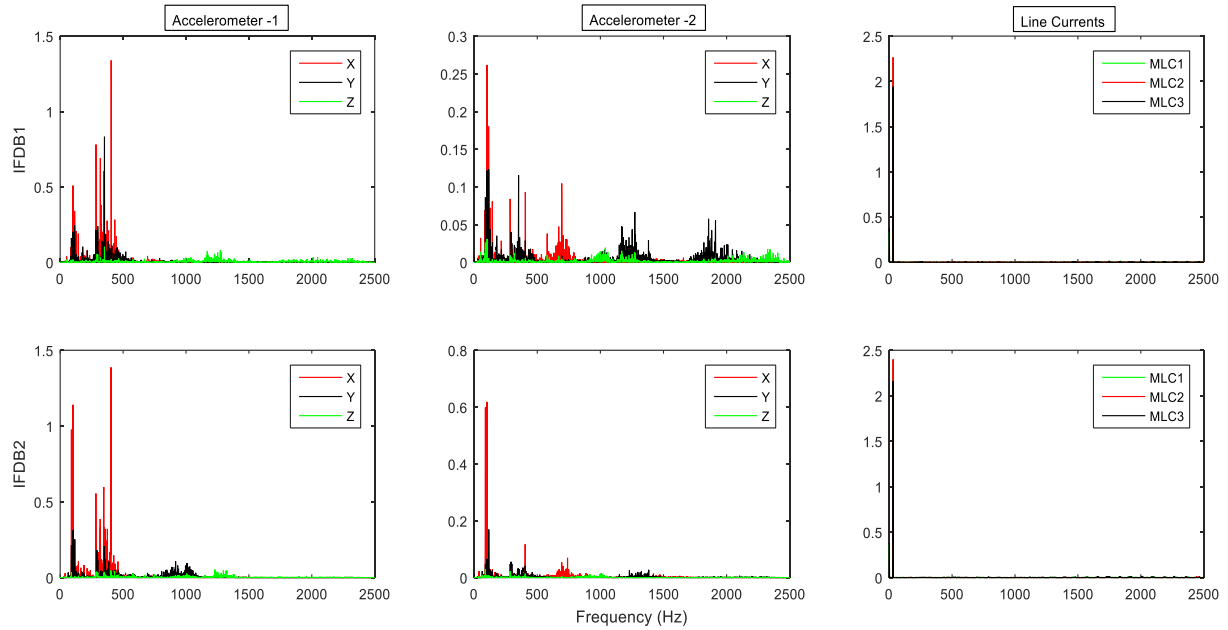


Figure E.7: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for IFDB1, IFDB2 faults.

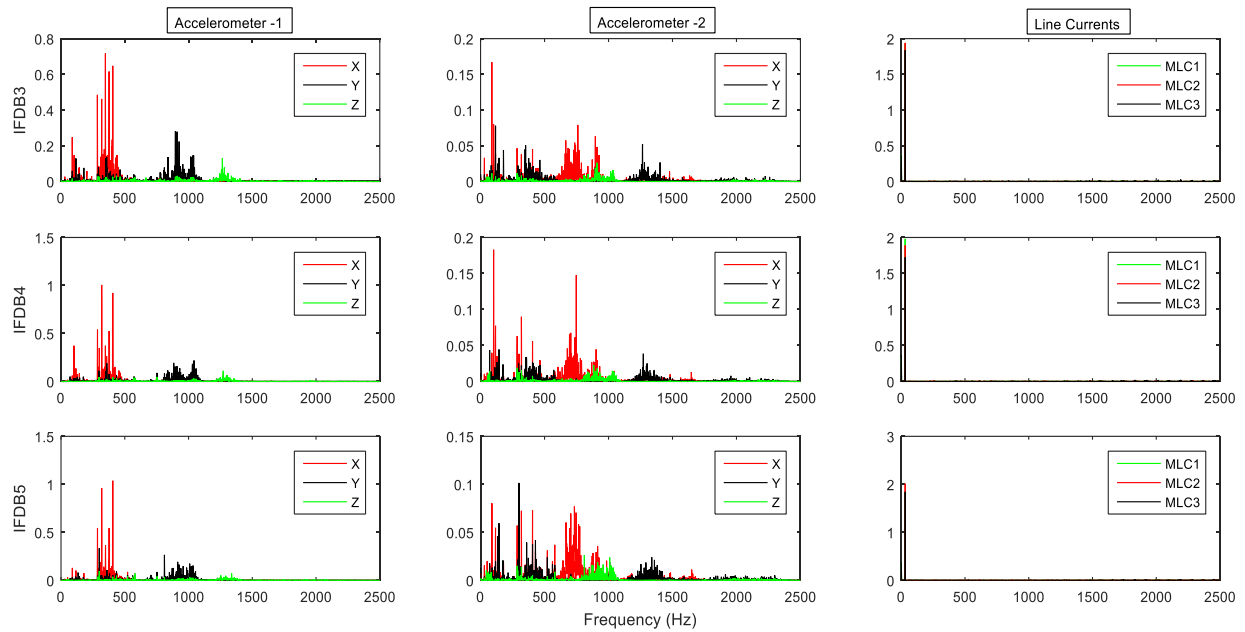


Figure E.8: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for IFDB3, IFDB4, IFDB5 faults.

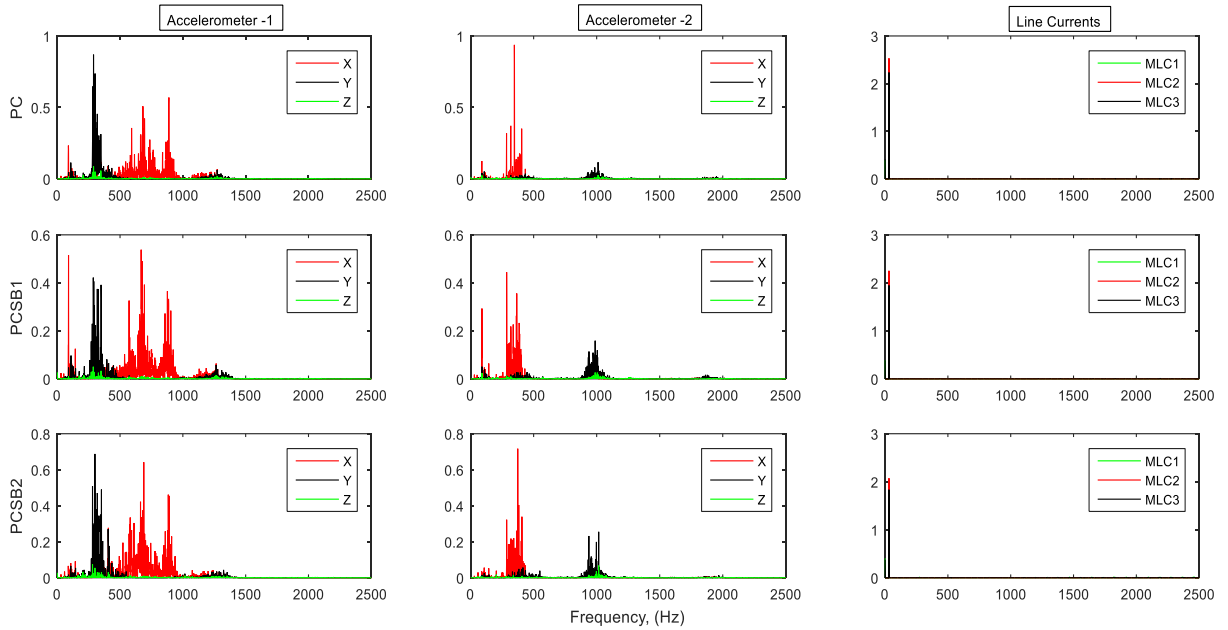


Figure E.9: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for PC, PCSB1, PCSB2 faults.

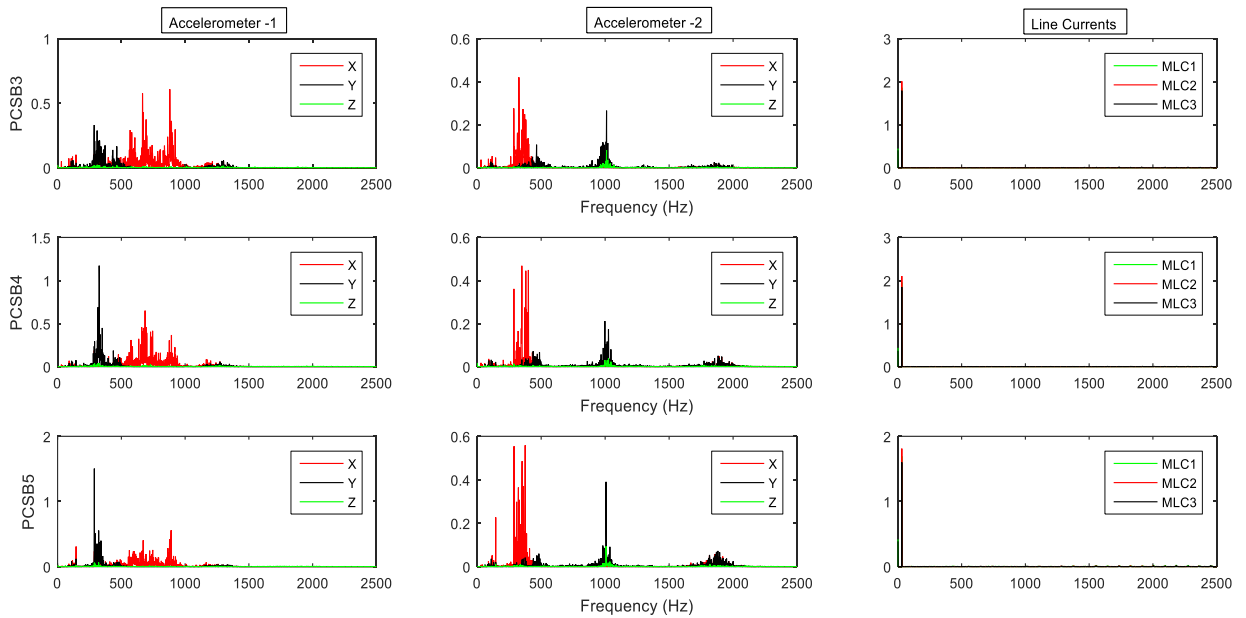


Figure E.10: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for PCSB3, PCSB4 and PCSB5 faults.

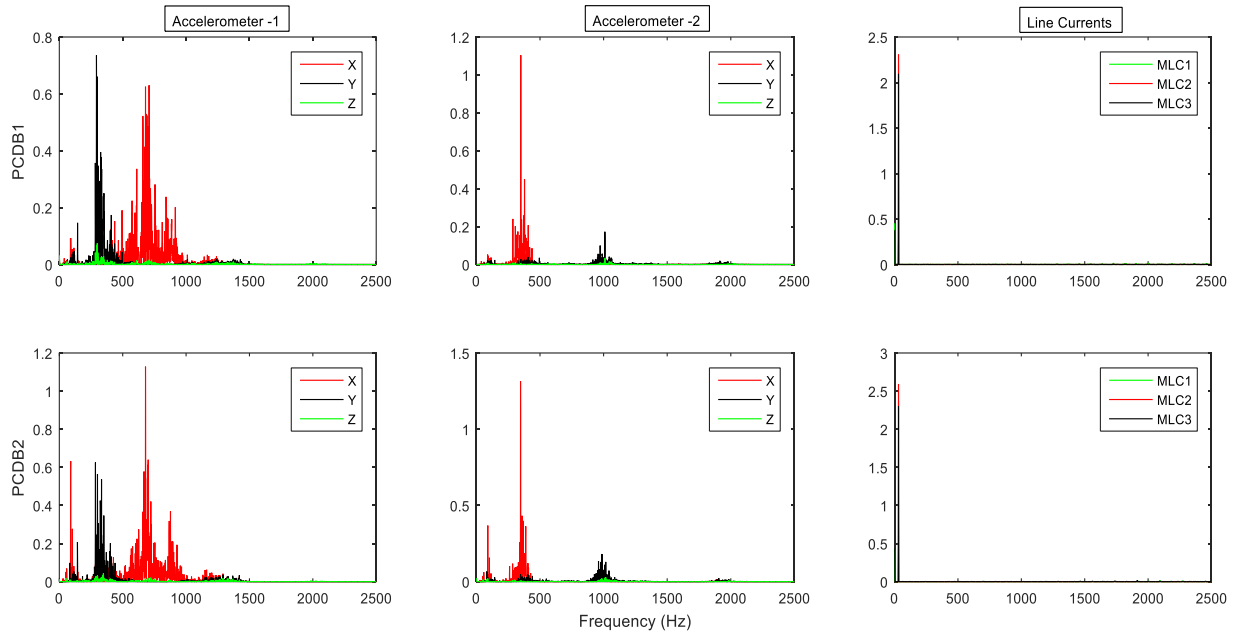


Figure E.11: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for PCDB1, PCDB2 faults.

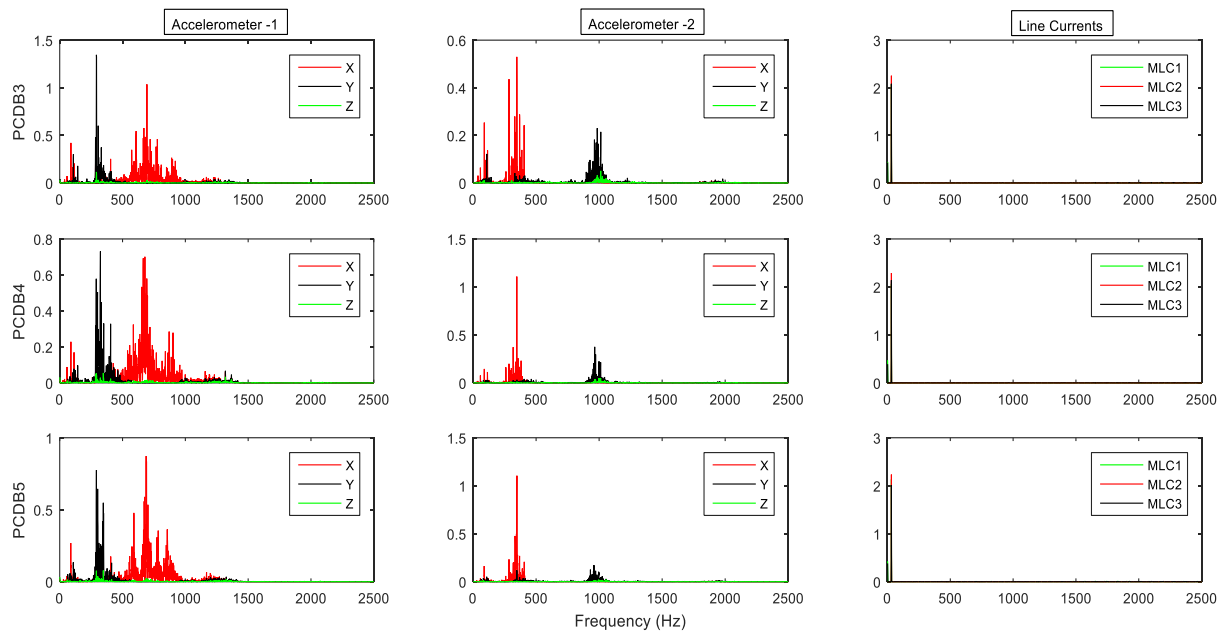


Figure E.12: PSD amplitude versus frequency plots using accelerometers 1 and 2 and line current probes for PCDB3, PCDB4, PCDB5 faults.

Appendix F: CWT plots of various CP faults low sampling rate

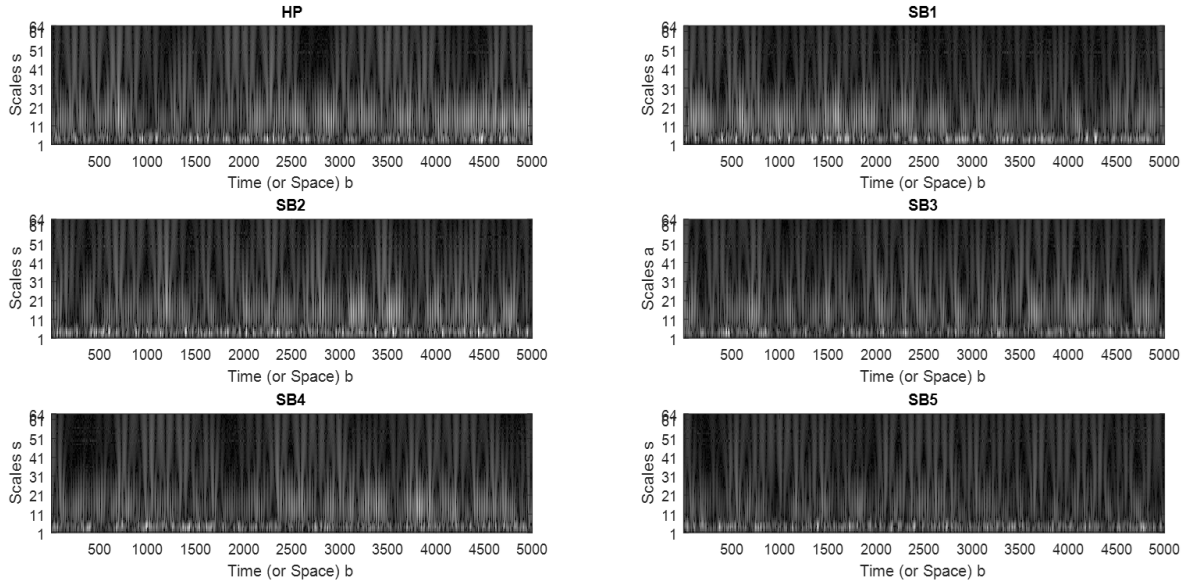


Figure F.1: CWT scale versus space plots using accelerometers 1 x-direction data for HP, SB1, SB2, SB3, SB4, SB5 faults using gaus1 wavelet.

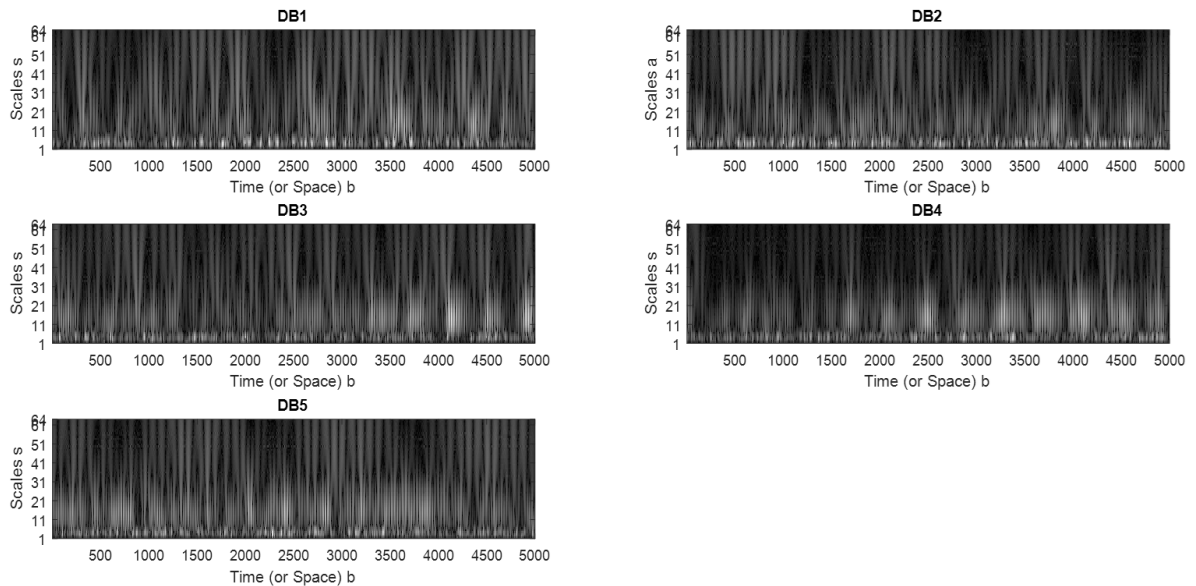


Figure F.2: CWT scale versus space plots using accelerometers 1 x-direction data for DB1, DB2, DB3, DB4, DB5 faults using gaus1 wavelet.

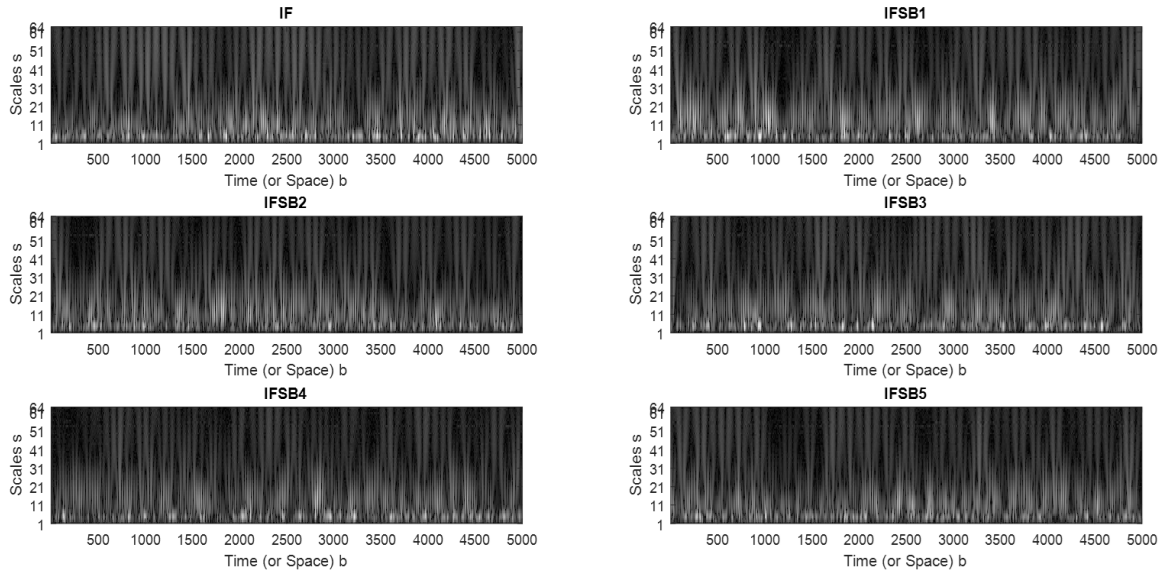


Figure F.3: CWT scale versus space plots using accelerometers 1 x-direction data for IF, IFSB1, IFSB2, IFSB3, IFSB4, IFSB5 faults using gaus1 wavelet.

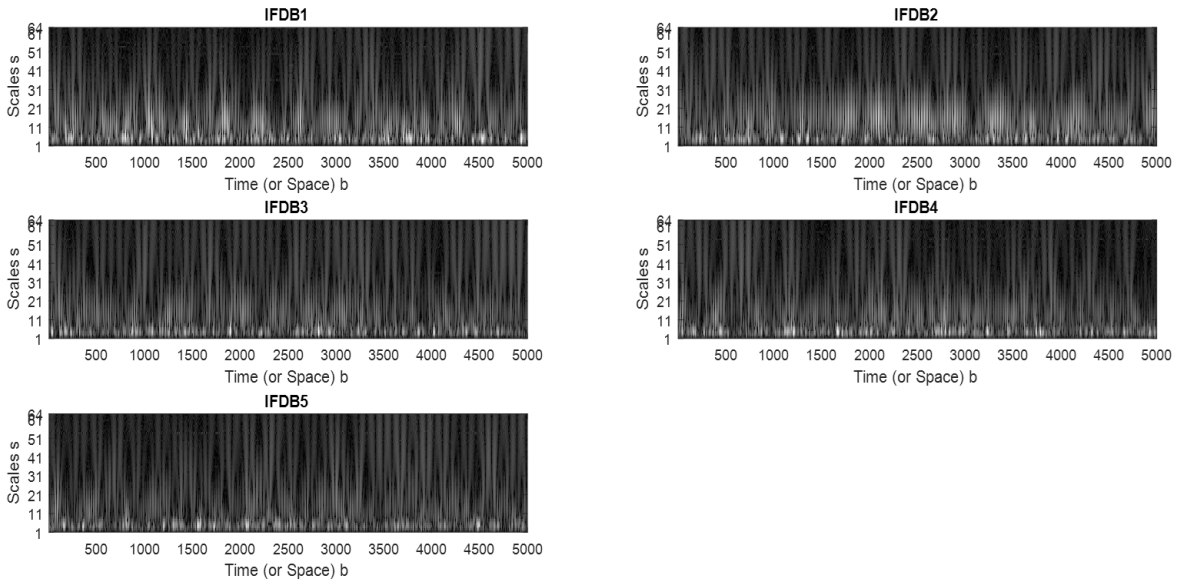


Figure F.4: CWT scale versus space plots using accelerometers 1 x-direction data for IFDB1, IFDB2, IFDB3, IFDB4, IFDB5 faults using gaus1 wavelet.

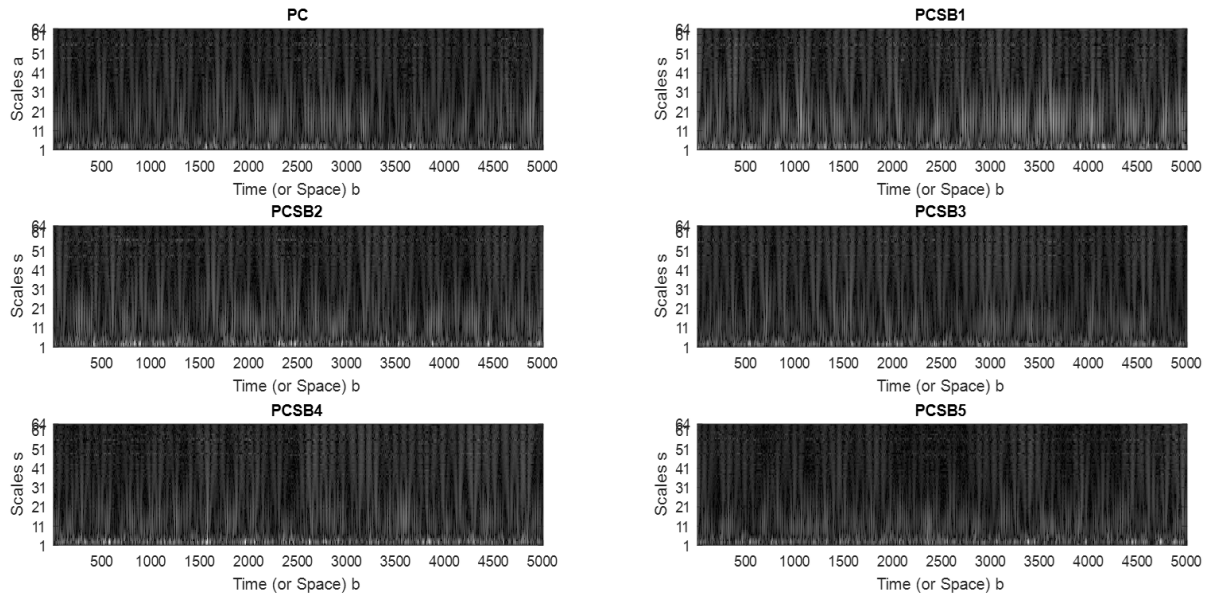


Figure F.5: CWT scale versus space plots using accelerometers -1 x-direction data for PC, PCSB1, PCSB2, PCSB3, PCSB4, PCSB5 faults using gaus1 wavelet.

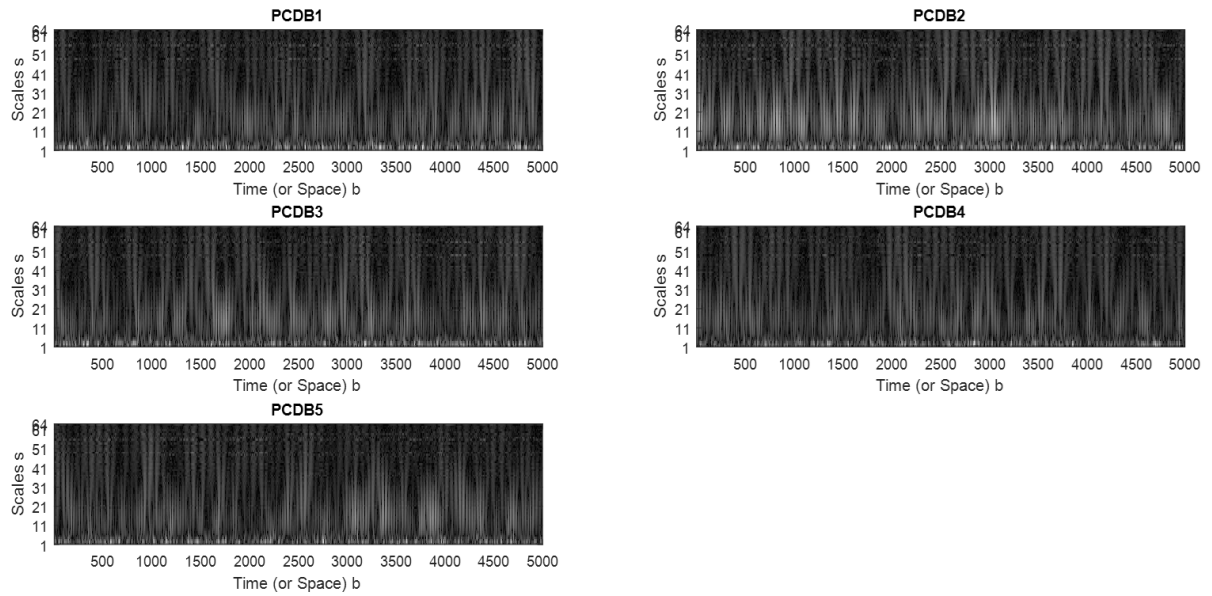


Figure F.6: CWT scale versus space plots using accelerometers 1 x-direction data for PCDB1, PCDB2, PCDB3, PCDB4, PCDB5 faults using gaus1 wavelet.



Appendix G: WPT plots of various CP faults low sampling rate

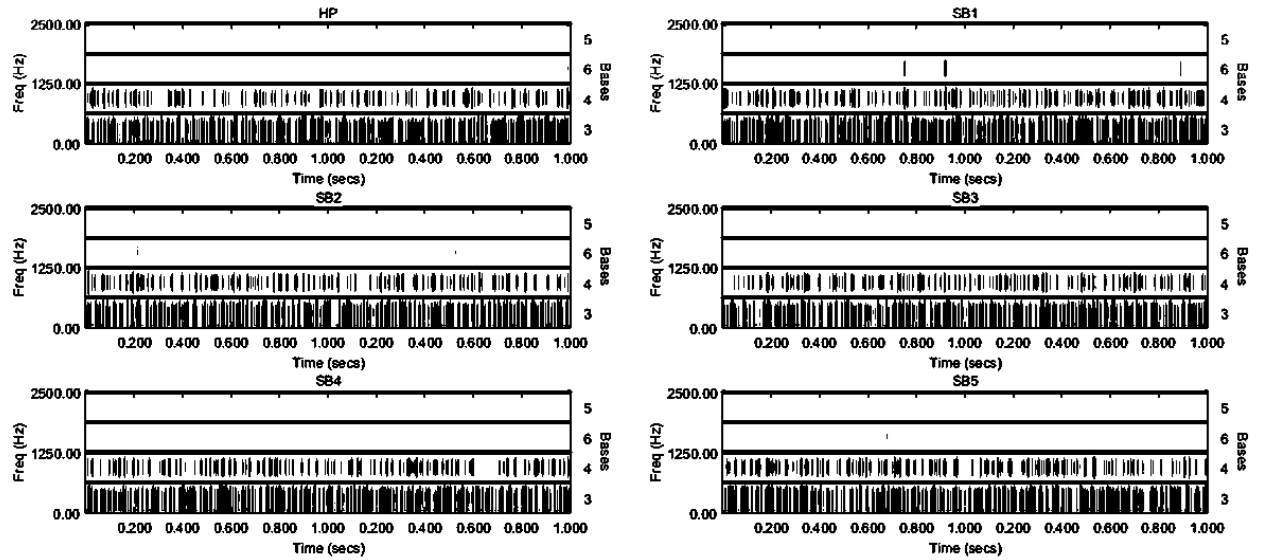


Figure G.1: WPT - Frequency versus time plots using accelerometers 1 x-direction data for HP, SB1, SB2, SB3, SB4, SB5 faults using db6 wavelet.

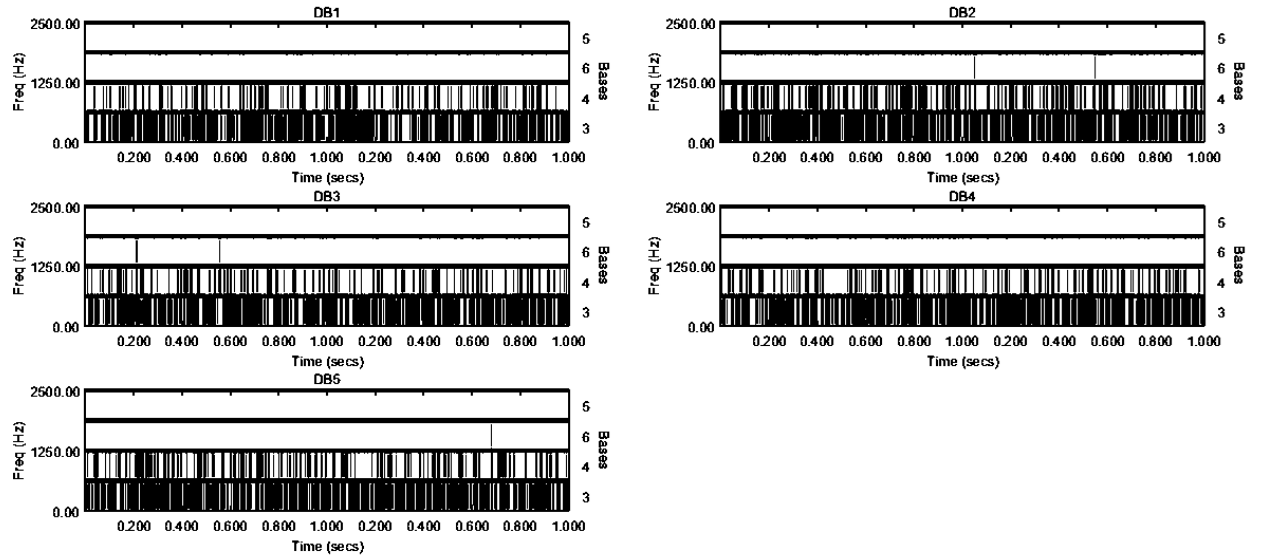


Figure G.2: WPT - Frequency versus time plots using accelerometers 1 x-direction data for DB1, DB2, DB3, DB4, DB5 faults using db6 wavelet.

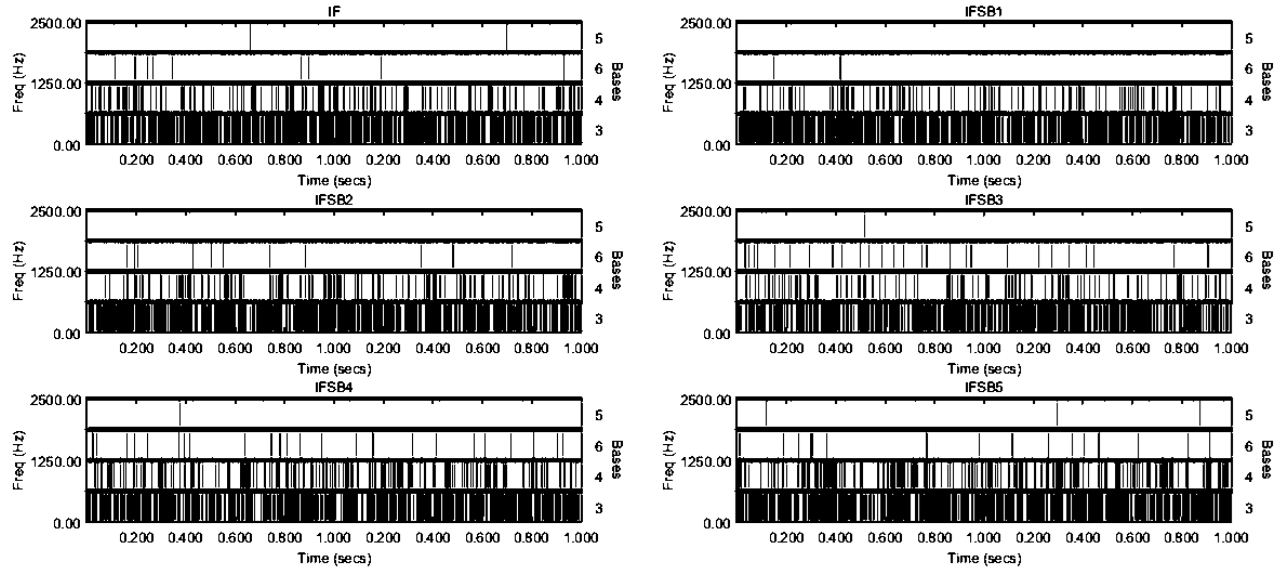


Figure G.3: WPT - Frequency versus time plots using accelerometers 1 x-direction data for IF, IFSB1, IFSB2, IFSB3, IFSB4, IFSB5 faults using db6 wavelet.

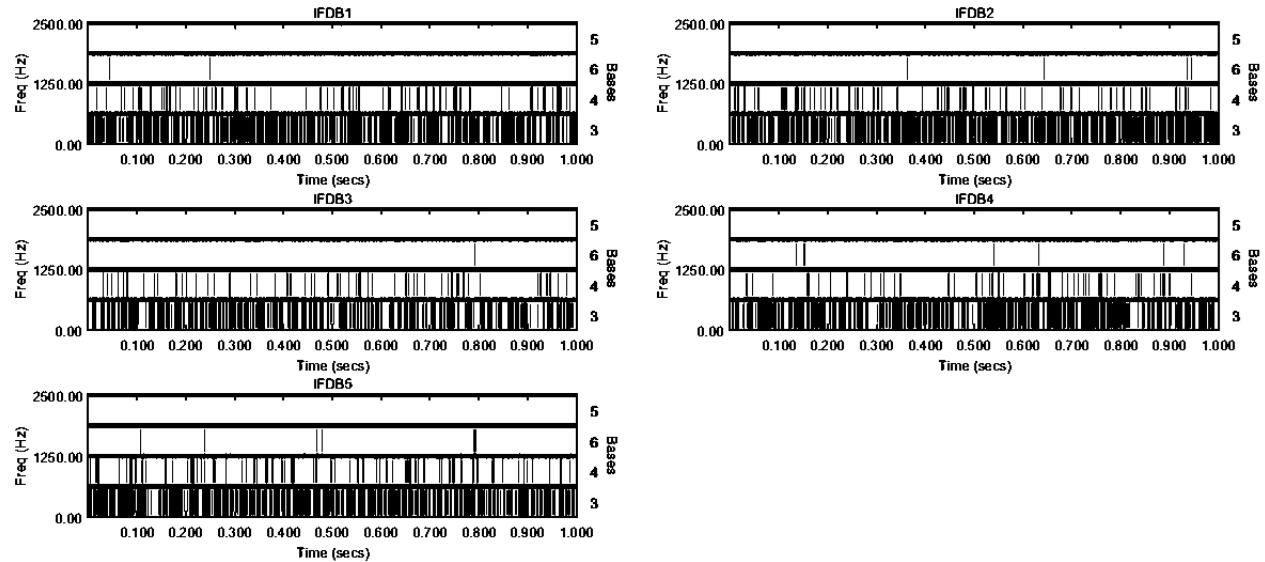


Figure G.4: WPT - Frequency versus time plots using accelerometers 1 x-direction data for IFDB1, IFDB2, IFDB3, IFDB4, IFDB5 faults using db6 wavelet.

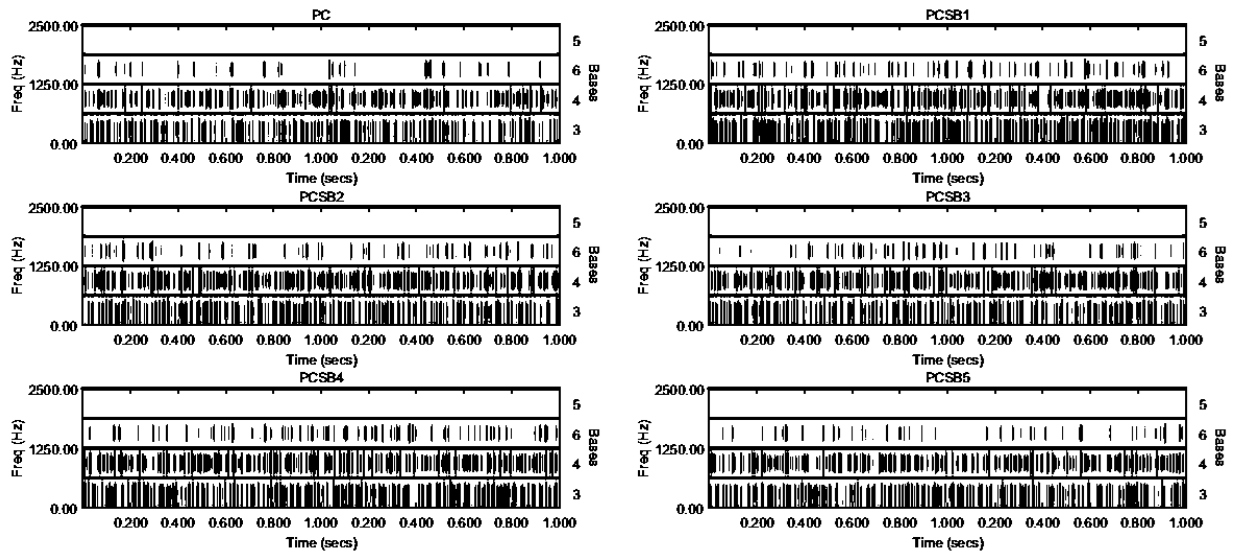


Figure G.5: WPT - Frequency versus time using accelerometers 1 x-direction data for PC, PCSB1, PCSB2, PCSB3, PCSB4, PCSB5 faults using db6 wavelet.

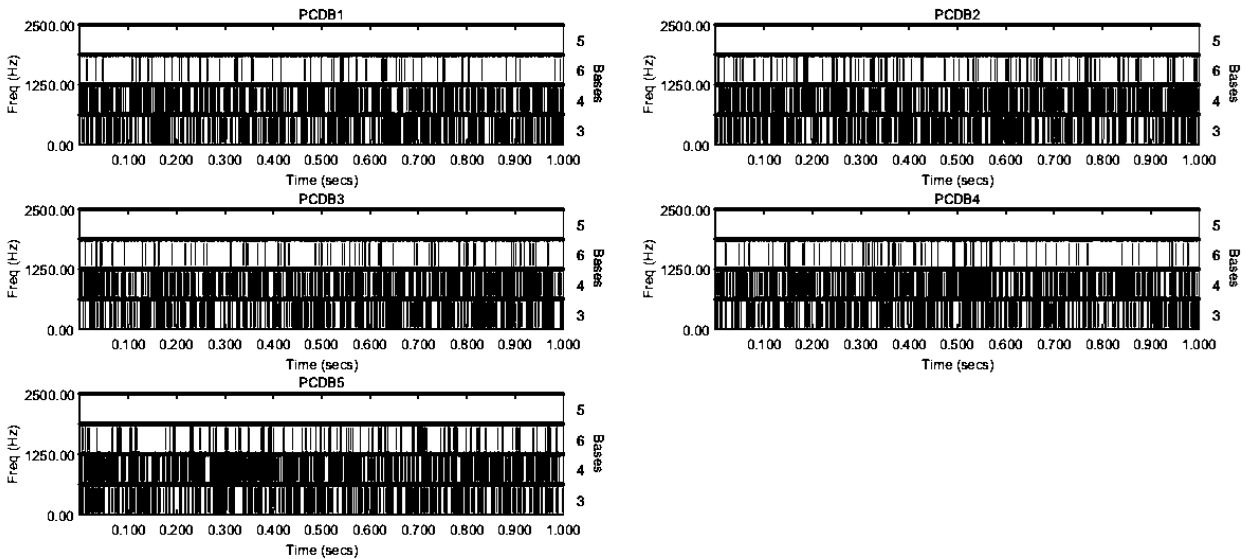


Figure G.6: WPT - Frequency versus time using accelerometers 1 x-direction data for PCDB1, PCDB2, PCDB3, PCDB4, PCDB5 faults using db6 wavelet.



Appendix H: Feature calculation for low sampling rate data

Table H.1: Sample calculation of various statistical features using time-domain data

S.No	x_i	$(x_i)^2$	$\mu = \frac{1}{5000} \sum_{i=1}^{5000} x_i$	$(x_i - \mu)$	$(x_i - \mu)^2$	$(x_i - \mu)^3$	$(x_i - \mu)^4$	$RMS = \sqrt{\frac{1}{5000} \sum_{i=1}^{5000} x_i^2}$	$\sigma^2 = \frac{1}{5000} \sum_{i=1}^{5000} (x_i - \mu)^2$	$\sigma = \sqrt{\sigma^2}$	$\chi = \frac{\frac{1}{N} \sum_{i=1}^{5000} (x_i - \mu)^3}{\sigma^3}$	$\kappa = \frac{\frac{1}{N} \sum_{i=1}^{5000} (x_i - \mu)^4}{\sigma^4}$
1	9.098877	82.78956	-0.08526	9.184136	84.34835	774.6666	7114.643	5.903941	34.84925	5.903326	-0.02541	3.026574
2	12.564515	157.867		12.64977	160.0168	2024.176	25605.37					
3	10.553911	111.385		10.63917	113.1919	1204.268	12812.41					
...	...	...		...	...	...	...					
...	...	...		...	...	...	...					
...	...	...		...	...	...	...					
4997	-4.901986	24.02947		-4.81673	23.20086	-111.752	538.2801					
4998	1.106755	1.224907		1.192014	1.420896	1.693727	2.018946					
4999	6.789185	46.09303		6.874444	47.25797	324.8723	2233.316					
5000	7.300931	53.30359		7.38619	54.5558	402.9594	2976.335					

Table H.2: Sample calculation of various statistical features using PSD data

S.No	x_i	$(x_i)^2$	$\mu = \frac{1}{2500} \sum_{i=1}^{2500} x_i$	$(x_i - \mu)$	$(x_i - \mu)^2$	$(x_i - \mu)^3$	$(x_i - \mu)^4$	$RMS = \sqrt{\frac{1}{2500} \sum_{i=1}^{2500} x_i^2}$	$\sigma^2 = \frac{1}{2500} \sum_{i=1}^{2500} (x_i - \mu)^2$	$\sigma = \sqrt{\sigma^2}$	$\chi = \frac{\frac{1}{N} \sum_{i=1}^{2500} (x_i - \mu)^3}{\sigma^3}$	$\kappa = \frac{\frac{1}{N} \sum_{i=1}^{2500} (x_i - \mu)^4}{\sigma^4}$
1	0.004805	2.3E-05	0.01361	-0.0088	7.8E-05	-6.8E-07	6.0E-09	0.066796	0.004276	0.065395	11.64689	181.1234
2	0.002431	5.9E-06		-0.01118	0.000124967	-1.4E-06	1.6E-08					
3	7.4E-07	5.5E-13		-0.01361	0.000185208	-2.5E-06	3.4E-08					
...	...	...		...	...	...	...					
...	...	...		...	...	...	...					
...	...	...		...	...	...	...					
...	...	...		...	...	...	...					
...	1.6E-06	2.6E-12		-0.01361	0.000185184	-2.5E-06	3.4E-08					
2498	1.3E-06	1.6E-12		-0.01361	0.000185194	-2.5E-06	3.4E-08					
2499	1.1E-05	1.1E-10		-0.0136	0.000184944	-2.5E-06	3.4E-08					
2500	1.1E-05	1.1E-10		-0.0136	0.000184938	-2.5E-06	3.4E-08					

References

- Abdi, H., and Williams, L.J. 2010. Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics **2**(4): 433-459. doi:doi:10.1002/wics.101.
- Abdulkarem, W., Amuthakkannan, R., and Al-Raheem, K.F. Centrifugal pump impeller crack detection using vibration analysis. *In* 2nd International Conference on Research in Science, Engineering and Technology (ICRSET). Dubai, UAE, 21 - 22 March 2014. pp. 206-211.
- Al-Hashmi, S., Gu, F., Li, Y., Ball, A.D., Fen, T., and Lui, K. Cavitation Detection of a Centrifugal Pump Using Instantaneous Angular Speed. *In* ASME 7th Biennial Conference on Engineering Systems Design and Analysis. Manchester, England, 19-22 July 2004. pp. 185-190.
- Alabied, S., Hamomd, O., Daraz, A., Gu, F., and Ball, A.D. Fault diagnosis of centrifugal pumps based on the intrinsic time-scale decomposition of motor current signals. *In* 23rd International Conference on Automation and Computing (ICAC). Huddersfield, UK, 7-8 September 2017. pp. 1-6.
- Albraik, A., Althobiani, F., Gu, F., and Ball, A. Diagnosis of centrifugal pump faults using vibration methods. *In* 25th International Congress on Condition Monitoring and Diagnostic Engineering Proceedings of the Journal of Physics: Conference Series, 2012. IOP Publishing. pp. 012139 (012131) - 012139 (012113).
- Alfayez, and Mba, D. 2005. Detection of incipient cavitation and determination of the best efficiency point for centrifugal pumps using acoustic emission. Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering **219**(4): 327-344. doi:10.1243/095440805X33225.

- Alfayez, Mba, D., and Dyson, G. 2004. Detection of incipient cavitation and the best efficiency point of a 2.2 MW centrifugal pump using Acoustic Emission. *Journal of Acoustic Emission* **22**: 77-82.
- Alfayez, Mba, D., and Dyson, G. 2005. The application of acoustic emission for detecting incipient cavitation and the best efficiency point of a 60kW centrifugal pump: case study. *NDT & E International* **38**(5): 354-358.
doi:<https://doi.org/10.1016/j.ndteint.2004.10.002>.
- Alpaydın, E. 2008. *Maschinelles Lernen*. Oldenbourg Verlag.
- Ashokkumar, M. 2011. The characterization of acoustic cavitation bubbles—an overview. *Ultrasonics sonochemistry* **18**(4): 864-872.
- Azadeh, A., Ebrahimipour, V., and Bavar, P. 2010. A fuzzy inference system for pump failure diagnosis to improve maintenance process: The case of a petrochemical industry. *Expert Systems with Applications* **37**(1): 627-639.
- Azadeh, A., Saberi, M., Kazem, A., Ebrahimipour, V., Nourmohammadzadeh, A., and Saberi, Z. 2013. A flexible algorithm for fault diagnosis in a centrifugal pump with corrupted data and noise based on ANN and support vector machine with hyper-parameters optimization. *Applied Soft Computing* **13**(3): 1478-1485.
doi:<http://dx.doi.org/10.1016/j.asoc.2012.06.020>.
- Azizi, R., Attaran, B., Hajnaye, A., Ghanbarzadeh, A., and Changizian, M. 2017. Improving accuracy of cavitation severity detection in centrifugal pumps using a hybrid feature selection technique. *Measurement* **108**: 9-17.
doi:<https://doi.org/10.1016/j.measurement.2017.05.020>.

- Behzad, M., Bastami, A.R., and Maassoumian, M. Fault diagnosis of a centrifugal pump by vibration analysis. *In* ASME 7th biennial conference on engineering systems design and analysis. Manchester, England, 19-22 July 2004. ASME, Volume 3. pp. 221-226.
- Bell, D., Liu, W., Cheng, J., Greiner, R., and Kelly, J. 2002. Learning bayesian networks from data: An information-theory based approach. *Artificial Intelligence* **137**(1-2): 43-90.
- Bertsekas, D.P. 2014. *Constrained optimization and Lagrange multiplier methods*. Academic press.
- Bonarini, A. An introduction to learning fuzzy classifier systems. *In* International workshop on learning classifier systems. Berlin, Heidelberg 1999. Springer. pp. 83-104.
- Bordoloi, D.J., and Tiwari, R. 2017. Identification of suction flow blockages and casing cavitations in centrifugal pumps by optimal support vector machine techniques [journal article]. *Journal of the brazilian society of mechanical sciences and engineering* **39**(8): 2957-2968. doi:10.1007/s40430-017-0714-z.
- Brennen, C.E. 2013. *Cavitation and bubble dynamics*. Cambridge University Press.
- Camargo, L., and Yoneyama, T. 2001. Specification of training sets and the number of hidden neurons for multilayer perceptrons. *Neural Computation* **13**(12): 2673-2680.
- Campbell, C., and Ying, Y. 2011. Learning with support vector machines. *Synthesis lectures on artificial intelligence and machine learning* **5**(1): 1-95.
- Chang, C.-C., and Lin, C.-J. 2011. LIBSVM: A library for support vector machines. *ACM transactions of intelligent systems technology* **2**(3): 1-27. doi:10.1145/1961189.1961199.

- Chia-Chi, L., and Kuo-Wei, H. A rule-based classification algorithm: A rough set approach. *In* IEEE international conference on computational intelligence and cybernetics Bali, Indonesia, 12-14 July 2012. IEEE. pp. 1-5.
- Chih-Wei, H., and Chih-Jen, L. 2002. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks* **13**(2): 415-425. doi:10.1109/72.991427.
- Chini, S.F., Rahimzadeh, H., and Bahrami, M. Cavitation detection of a centrifugal pump using noise spectrum. *In* International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Proceedings of the 20th Biennial Conference on Mechanical Vibration and Noise, Parts A, B, and C, California, USA, 24-28 September 2005. ASME. pp. 13-19.
- Chudina, M. 2003. Noise as an indicator of cavitation in a centrifugal pump. *Acoustical Physics* **49**(4): 463-474.
- Cortes, C., and Vapnik, V. 1995. Support-vector networks [journal article]. *Machine Learning* **20**(3): 273-297. doi:10.1007/bf00994018.
- Čudina, M., and Prezelj, J. 2009. Detection of cavitation in situ operation of kinetic pumps: Effect of cavitation on the characteristic discrete frequency component. *Applied Acoustics* **70**(9): 1175-1182. doi:https://doi.org/10.1016/j.apacoust.2009.04.001.
- Dister, C.J. On-line health assessment of integrated pumps. *In* Aerospace conference proceedings Montana, USA, 8-15 March 2003. IEEE. pp. 3289-3294.
- Du, R., Elbestawi, M.A., and Wu, S.M. 1995. Automated monitoring of manufacturing processes, part 1: Monitoring methods. *Journal of engineering for industry* **117**(2): 121-132. doi:10.1115/1.2803286.

- Dular, M., Stoffel, B., and Širok, B. 2006. Development of a cavitation erosion model. *Wear* **261**(5-6): 642-655.
- Dutton, D.M., and Conroy, G.V. 1997. A review of machine learning. *The knowledge engineering review* **12**(4): 341-367.
- Ebrahimi, E., and Javidan, M. 2017. Vibration-based classification of centrifugal pumps using support vector machine and discrete wavelet transform. *Journal of vibroengineering* **19**(4): 2586-2597.
- Elger, D.F., and Roberson, J.A. 2016. *Engineering fluid mechanics*. Wiley Hoboken (NJ).
- Everett III, H. 1963. Generalized Lagrange multiplier method for solving problems of optimum allocation of resources. *Operations research* **11**(3): 399-417.
- Farokhzad, S., Ahmadi, H., Jaefari, A., Asadi Asad Abad, M., and Ranjbar Kohan, M. 2012. Artificial neural network based classification of faults in centrifugal water pump. *Vibroengineering* **14**(4): 1734-1744.
- Filippetti, F., Franceschini, G., Tassoni, C., and Vas, P. 2000. Recent developments of induction motor drives fault diagnosis using AI techniques. *IEEE Transactions on Industrial Electronics* **47**(5): 994-1004. doi:10.1109/41.873207.
- Frank, P.M. 1996. Analytical and qualitative model-based fault diagnosis – a survey and some new results. *European journal of control* **2**(1): 6-28. doi:[https://doi.org/10.1016/S0947-3580\(96\)70024-9](https://doi.org/10.1016/S0947-3580(96)70024-9).
- Fraser, H. Flow recirculation in centrifugal pumps. *In* *Turbomachinery and Pump Symposia*. Texas A&M University, USA 1981. pp. 95-100.

- Greene, R., Casada, D., and Ayers, C. 1995. Detection of pump degradation. Nuclear Regulatory Commission, Washington, DC (United States). Office of Nuclear Regulatory Research; Oak Ridge National Lab., TN (United States).
- Gulich, J.F. 2008. Centrifugal pumps. Springer.
- Halligan, G.R., and Jagannathan, S. 2011. PCA-based fault isolation and prognosis with application to pump [journal article]. The international journal of advanced manufacturing technology **55**(5): 699-707. doi:10.1007/s00170-010-3096-2.
- Hamomd, O., Alabied, S., Xu, Y., Daraz, A., Gu, F., and Ball, A. Vibration based centrifugal pump fault diagnosis based on modulation signal bispectrum analysis. *In* 23rd international conference on automation and computing (ICAC). Huddersfield, UK, 7-8 September 2017. IEEE. pp. 1-5.
- Harihara, P.P., and Parlos, A.G. Sensorless detection of cavitation in centrifugal pumps. *In* International Mechanical Engineering Congress and Exposition. Illinois, USA, 5-10 November 2006. ASME, Manufacturing Engineering and Textile Engineering. pp. 187-192.
- Harihara, P.P., and Parlos, A.G. 2008a. Sensorless Detection and Isolation of Faults in Motor-Pump Systems. (48661): 43-50. doi:10.1115/IMECE2008-66446.
- Harihara, P.P., and Parlos, A.G. Sensorless detection of impeller cracks in motor driven centrifugal pumps. *In* International mechanical engineering congress and exposition. Massachusetts, USA, 31 October - 6 November 2008b. ASME, Volume 5: Design, Analysis, Control and Diagnosis of Fluid Power Systems. pp. 17-23.

- Hart, R.J. Pumps and their systems-a changing industry. *In* Proceedings of the 19th international pump users symposium. Houston, Texas, U.A, February 2002. pp. 141-144.
- Hastie, T., and Tibshirani, R. Classification by pairwise coupling. *In* Advances in neural information processing systems. 1998. pp. 507-513.
- Hennecke, F. Reliability of pumps in chemical industry. *In* Proceedings of pump users international forum Karlsruhe, Germany 2000.
- Hsu, C.-W., Chang, C.-C., and Lin, C.-J. 2003. A practical guide to support vector classification.
- Isermann, R. 1984. Process fault detection based on modeling and estimation methods—A survey. *Automatica* **20**(4): 387-404. doi:[https://doi.org/10.1016/0005-1098\(84\)90098-0](https://doi.org/10.1016/0005-1098(84)90098-0).
- Isermann, R. 2005. Model-based fault-detection and diagnosis – status and applications. *Annual reviews in control* **29**(1): 71-85. doi:<https://doi.org/10.1016/j.arcontrol.2004.12.002>.
- Isermann, R. 2011. Fault-diagnosis applications: model-based condition monitoring: actuators, drives, machinery, plants, sensors, and fault-tolerant systems. Springer Science & Business Media.
- Isermann, R., and Ballé, P. 1997. Trends in the application of model-based fault detection and diagnosis of technical processes. *Control Engineering Practice* **5**(5): 709-719. doi:[https://doi.org/10.1016/S0967-0661\(97\)00053-1](https://doi.org/10.1016/S0967-0661(97)00053-1).
- Jackson, C. 1979. The practical vibration primer. Gulf Pub Co.

- Jain, A.K., Murty, M.N., and Flynn, P.J. 1999. Data clustering: a review. *ACM computing surveys (CSUR)* **31**(3): 264-323.
- Jardine, A.K.S., Lin, D., and Banjevic, D. 2006. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing* **20**(7): 1483-1510. doi:<https://doi.org/10.1016/j.ymssp.2005.09.012>.
- Jensen, F.V. 1996. *An introduction to Bayesian networks*. UCL press London.
- Kallesoe, C.S., Cocquempot, V., and Izadi-Zamanabadi, R. 2006. Model based fault detection in a centrifugal pump application. *IEEE transactions on control systems technology* **14**(2): 204-215. doi:10.1109/TCST.2005.860524.
- Kallesoe, C.S., Izaili-Zamanabadi, R., Rasmussen, H., and Cocquempot, V. Model based fault diagnosis in a centrifugal pump application using structural analysis. *In Proceedings of international conference on control applications*. Taipei, Taiwan, 2-4 September 2004. IEEE. pp. 1229-1235.
- Kamiel, B., Forbes, G., Entwistle, R., Mazhar, I., and Howard, I. 2005. Impeller fault detection for a centrifugal pump using principal component analysis of time domain vibration features. Department of Mechanical Engineering, Universitas Muhammadiyah Yogyakarta, Indonesia.
- Kotsiantis, S.B., Zaharakis, I., and Pintelas, P. 2007. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering* **160**: 3-24.
- Kotsiantis, S.B., Zaharakis, I.D., and Pintelas, P.E. 2006. Machine learning: a review of classification and combining techniques. *Artificial intelligence review* **26**(3): 159-190.

- Kumar, A., and Kumar, R. 2017. Time-frequency analysis and support vector machine in automatic detection of defect from vibration signal of centrifugal pump. *Measurement* **108**: 119-133. doi:<https://doi.org/10.1016/j.measurement.2017.04.041>.
- Lee, Y., Lin, Y., and Wahba, G. 2004. Multicategory support vector machines: Theory and application to the classification of microarray data and satellite radiance data. *Journal of the American Statistical Association* **99**(465): 67-81.
- Lima, C.A., Stopa, M.M., Filho, B.J.C., Miranda, L.S., Luiz, A.S.A., and Martinez, C.B. Cavitation detection in centrifugal pumps driven by frequency inverter fed motors. *In* Industry applications society annual meeting. Cincinnati, Ohio, USA, 1-5 October 2017. IEEE pp. 1-9.
- Lobanoff, V.S., and Ross, R.R. 2013. Centrifugal pumps: design and application. Elsevier.
- Lu, J., Yuan, S., Luo, Y., Yuan, J., Zhou, B., and Sun, H. 2016. Numerical and experimental investigation on the development of cavitation in a centrifugal pump. *Proceedings of the institution of mechanical engineers, part E: Journal of process mechanical engineering* **230**(3): 171-182. doi:[10.1177/0954408914557877](https://doi.org/10.1177/0954408914557877).
- Luo, Y., Zhixiang, X., Sun, H., Yuan, S., and Yuan, J. 2015. Research on the induction motor current signature for centrifugal pump at cavitation condition. *Advances in mechanical engineering* **7**(11): 1-13. doi:[10.1177/1687814015617134](https://doi.org/10.1177/1687814015617134).
- Makay, E. 1980. Centrifugal pump hydraulic instability. *In* Electric power research institute report EPRI CS-1445, Palo Alto, California, USA.
- McKee, Forbes, G.L., Mazhar, I., Entwistle, R., and Howard, I. A review of machinery diagnostics and prognostics implemented on a centrifugal pump. *In* Engineering asset

- management London, UK, 2014. *Edited by J. Lee and J. Ni and J. Sarangapani and J. Mathew.* Springer London. pp. 593-614.
- McKee, K., Forbes, G., Mazhar, M.I., Entwistle, R., and Howard, I. A review of major centrifugal pump failure modes with application to the water supply and sewerage industries. *In* ICOMS asset management conference Gold Coast, Australia, 16 May 2011. Asset Management Council.
- McKee, K.K., Forbes, G.L., Mazhar, I., Entwistle, R., Hodkiewicz, M., and Howard, I. 2015. A vibration cavitation sensitivity parameter based on spectral and statistical methods. *Expert systems with applications* **42**(1): 67-78.
- Mohanty, A.R., Pradhan, P.K., Mahalik, N.P., and Dastidar, S.G. 2012. Fault detection in a centrifugal pump using vibration and motor current signature analysis. *International journal of automation and control* **6**(3-4): 261-276.
- Muralidharan, V., and Sugumaran, V. 2012. A comparative study of naïve bayes classifier and bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis. *Applied soft computing* **12**(8): 2023-2029. doi:<http://dx.doi.org/10.1016/j.asoc.2012.03.021>.
- Muralidharan, V., and Sugumaran, V. 2013a. Feature extraction using wavelets and classification through decision tree algorithm for fault diagnosis of mono-block centrifugal pump. *Measurement* **46**(1): 353-359. doi:<https://doi.org/10.1016/j.measurement.2012.07.007>.
- Muralidharan, V., and Sugumaran, V. 2013b. Rough set based rule learning and fuzzy classification of wavelet features for fault diagnosis of monoblock centrifugal pump.

doi:<https://doi.org/10.1016/j.measurement.2013.06.002>.

- Muralidharan, V., and Sugumaran, V. 2013c. Selection of discrete wavelets for fault diagnosis of monoblock centrifugal pump using the J48 algorithm. *Applied artificial intelligence* **27**(1): 1-19. doi:10.1080/08839514.2012.721694.
- Muralidharan, V., Sugumaran, V., and Indira, V. 2014. Fault diagnosis of monoblock centrifugal pump using SVM. *Engineering science and technology, an international journal* **17**(3): 152-157.
- Murthy, S.K. 1998. Automatic construction of decision trees from data: A multi-disciplinary survey. *Data mining and knowledge discovery* **2**(4): 345-389.
- Nandi, S., Toliyat, H.A., and Li, X. 2005. Condition monitoring and fault diagnosis of electrical motors-a review. *IEEE transactions on energy conversion* **20**(4): 719-729. doi:10.1109/TEC.2005.847955.
- Nandi, S., and Toliyat, S.A. Fault diagnosis of electrical machines-a review. *In* Electric machines and drives, international conference IEMD Seattle, Washington, USA, 9-12 May 1999. pp. 219-221.
- Nasiri, M.R., Mahjoob, M.J., and Vahid-Alizadeh, H. Vibration signature analysis for detecting cavitation in centrifugal pumps using neural networks. *In* International conference on mechatronics. Istanbul, Turkey, 13-15 April 2011. IEEE pp. 632-635.
- Nelson, W.E. Pump vibrations. *In* Proceedings of the 9th international pump users symposium. 1992. Turbomachinery Laboratories, Department of Mechanical Engineering, Texas A&M University.

- Ning, B., Cheng, X.-x., and Wu, S. 2011. Research on centrifugal pump monitoring system based on virtualization technology. *Procedia Engineering* **15**: 1077-1081.
- Nordmann, R., and Aenis, M. 2004. Fault diagnosis in a centrifugal pump using active magnetic bearings. *International journal of rotating machinery* **10**(3). doi:10.1155/s1023621x04000193.
- Patel, T.H., and Darpe, A.K. 2009. Vibration response of misaligned rotors. *Journal of sound and vibration* **325**(3): 609-628. doi:<https://doi.org/10.1016/j.jsv.2009.03.024>.
- Pawlak, Z. Some issues on rough sets. *In Transactions on Rough Sets I Proceedings of the Lecture notes in computer science, Berlin, Heidelberg 2004. Springer Berlin Heidelberg*. pp. 1-58.
- Pawlak, Z., and Skowron, A. 2007. Rudiments of rough sets. *Information sciences* **177**(1): 3-27.
- Peng, and Chu. 2004. Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. *Mechanical systems and signal processing* **18**(2): 199-221. doi:[https://doi.org/10.1016/S0888-3270\(03\)00075-X](https://doi.org/10.1016/S0888-3270(03)00075-X).
- Peng, Y., Dong, M., and Zuo, M.J. 2010. Current status of machine prognostics in condition-based maintenance: a review. *The international journal of advanced manufacturing technology* **50**(1-4): 297-313.
- Perovic, S., Unsworth, P.J., and Higham, E.H. Fuzzy logic system to detect pump faults from motor current spectra. *In Industry applications conference. Chicago, Illinois, USA, 30 September - 4 October 2001. IEEE*. pp. 274-280

- Platt, J.C., Cristianini, N., and Shawe-Taylor, J. Large margin DAGs for multiclass classification. *In* Advances in neural information processing systems. 2000. pp. 547-553.
- Qi, X., and Neupauer, R.M. 2008. Wavelet analysis of dominant scales of heterogeneous porous media. *Water Resources Research* **44**(9).
- Quinlan, J.R. 2014. C4.5: programs for machine learning. Elsevier.
- Rao, B. 1996. Handbook of condition monitoring. Elsevier.
- Saberi, M., Azadeh, A., Nourmohammadzadeh, A., and Pazhoheshfar, P. Comparing performance and robustness of SVM and ANN for fault diagnosis in a centrifugal pump. *In* 19th International Congress on Modelling and Simulation Perth, Australia, 12–16 December 2011. pp. 433-439.
- Sakthivel, N.R., Nair, B.B., Elangovan, M., Sugumaran, V., and Saravanmurugan, S. 2014. Comparison of dimensionality reduction techniques for the fault diagnosis of monoblock centrifugal pump using vibration signals. *Engineering science and technology, an international journal* **17**(1): 30-38.
- Sakthivel, N.R., Nair, B.B., and Sugumaran, V. 2012a. Soft computing approach to fault diagnosis of centrifugal pump. *Applied soft computing* **12**(5): 1574-1581.
- Sakthivel, N.R., Sugumaran, V., and Babudevasenapati, S. 2010a. Vibration based fault diagnosis of monoblock centrifugal pump using decision tree. *Expert systems with applications* **37**(6): 4040-4049. doi:<https://doi.org/10.1016/j.eswa.2009.10.002>.
- Sakthivel, N.R., Sugumaran, V., and Nair, B.B. 2010b. Comparison of decision tree-fuzzy and rough set-fuzzy methods for fault categorization of mono-block centrifugal pump. *Mechanical systems and signal processing* **24**(6): 1887-1906.

- Sakthivel, N.R., Sugumaran, V., and Nair, B.B. 2012b. Automatic rule learning using roughset for fuzzy classifier in fault categorization of mono-block centrifugal pump. *Applied soft computing* **12**(1): 196-203.
- Samanipour, P., Poshtan, J., and Sadeghi, H. 2017. Cavitation detection in centrifugal pumps using pressure time-domain features. *Turkish journal of electrical engineering and computer sciences* **25**(5): 4287-4298.
- Shervani-Tabar, M.T., Ettefagh, M.M., Lotfan, S., and Safarzadeh, H. 2017. Cavitation intensity monitoring in an axial flow pump based on vibration signals using multi-class support vector machine. *Proceedings of the institution of mechanical engineers, Part C: Journal of mechanical engineering science* **232**(17): 3013-3026. doi:10.1177/0954406217729416.
- Sinha, J.K., and Rao, A.R. 2006. Vibration based diagnosis of a centrifugal pump. *Structural health monitoring* **5**(4): 325-332. doi:10.1177/1475921706067760.
- Sloteman, D.P. Cavitation in high energy pumps-Detection and assessment of damage potential. *In 23rd international pump users symposium*. Texas, USA 2007. Texas A&M University. pp. 29-38.
- Soylemezoglu, A., Jagannathan, S., and Saygin, C. 2011. Mahalanobis-taguchi system as a multi-sensor based decision making prognostics tool for centrifugal pump failures. *IEEE transactions on reliability* **60**(4): 864-878. doi:10.1109/TR.2011.2170255.
- Stopa, M.M., Cardoso Filho, B.J., and Martinez, C.B. 2014. Incipient detection of cavitation phenomenon in centrifugal pumps. *IEEE transactions on industry applications* **50**(1): 120-126. doi:10.1109/TIA.2013.2267709.

- Szenasi, F.R., Sparks, C.R., and Wachel, J. 1986. Pump noise. *In Pump handbook*. McGraw-Hill. pp. 8.101-108.118.
- Tabar, M.T.S., Hojjat Majidi, C., and Poursharifi, Z. 2011. Investigation of recirculation effects on the formation of vapor bubbles in centrifugal pump blades. *World academy of science* 5(1): 80-85.
- Tan, C.Z., and Leong, M.S. 2008. An experimental study of cavitation detection in a centrifugal pump using envelope analysis. *Journal of system design and dynamics* 2(1): 274-285.
- Tiwari, R. 2017. Rotor systems: Analysis and identification. CRC Press.
- Tse, P.W., Yang, W.-x., and Tam, H.Y. 2004. Machine fault diagnosis through an effective exact wavelet analysis. *Journal of sound and vibration* 277(4): 1005-1024. doi:<https://doi.org/10.1016/j.jsv.2003.09.031>.
- Vapnik, V. 1995. The nature of statistical learning theory. Springer science & business media.
- Vapnik, V.N. 1998. Statistical learning theory. Wiley New York.
- Vapnik, V.N. 1999. An overview of statistical learning theory. *IEEE transactions on neural networks* 10(5): 988-999.
- Verma, N.K., Agrawal, A.K., Sevakula, R.K., Prakash, D., and Salour, A. Improved signal preprocessing techniques for machine fault diagnosis. *In 8th international conference on industrial and information systems*. Peradeniya, Sri Lanka, 17-20 December 2013. IEEE. pp. 403-408.
- Walczak, B., and Massart, D. 1997. Noise suppression and signal compression using the wavelet packet transform. *Chemometrics and Intelligent Laboratory Systems* 36(2): 81-94.

- Wang, H., and Chen, P. Intelligent method for condition diagnosis of pump system using discrete wavelet transform, rough sets and neural network. *In* Second international conference on bio-inspired computing: Theories and applications. Zhengzhou, China, 14-17 September 2007a. IEEE. pp. 24-28.
- Wang, H., and Chen, P. 2007b. Sequential condition diagnosis for centrifugal pump system using fuzzy neural network. *Neural information processing-letters and reviews* **11**(3): 41-50.
- Wang, H., and Chen, P. 2009. Intelligent diagnosis method for a centrifugal pump using features of vibration signals [journal article]. *Neural computing and applications* **18**(4): 397-405. doi:10.1007/s00521-008-0192-4.
- Widodo, A., and Yang, B.-S. 2007. Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical systems and signal processing* **21**(6): 2560-2574. doi:<http://dx.doi.org/10.1016/j.ymssp.2006.12.007>.
- Wolfe, P. 1961. A duality theorem for non-linear programming. *Quarterly of applied mathematics* **19**(3): 239-244.
- Wolfram, Fussel, D., Brune, T., and Isermann, R. Component-based multi-model approach for fault detection and diagnosis of a centrifugal pump. *In* American control conference. Arlington, Virginia, USA, 25-27 June 2001a. pp. 4443-4448.
- Wolfram, A., Fussel, D., Brune, T., and Isermann, R. Component-based multi-model approach for fault detection and diagnosis of a centrifugal pump. *In* American Control Conference, 2001. Proceedings of the 2001. 2001b. IEEE. pp. 4443-4448.

- Worster, R.C. 1963. The flow in volutes and its effect on centrifugal pump performance. Proceedings of the institution of mechanical engineers **177**(1): 843-875. doi:10.1243/PIME_PROC_1963_177_061_02.
- Wu, J.-D., and Chen, J.-C. 2006. Continuous wavelet transform technique for fault signal diagnosis of internal combustion engines. NDT & E International **39**(4): 304-311. doi:https://doi.org/10.1016/j.ndteint.2005.09.002.
- Wu, J.-D., and Liu, C.-H. 2009. An expert system for fault diagnosis in internal combustion engines using wavelet packet transform and neural network. Expert Systems with Applications **36**(3, Part 1): 4278-4286. doi:https://doi.org/10.1016/j.eswa.2008.03.008.
- Xue, H., Li, Z., Wang, H., and Chen, P. 2014. Intelligent diagnosis method for centrifugal pump system using vibration signal and support vector machine. Shock and vibration **2014**: 1-14. doi:10.1155/2014/407570.
- Yan, R., Gao, R.X., and Chen, X. 2014. Wavelets for fault diagnosis of rotary machines: A review with applications. Signal processing **96**: 1-15. doi:https://doi.org/10.1016/j.sigpro.2013.04.015.
- Yang, J., and Lövsund, P. Development and validation of a human-body mathematical model for simulation of car-pedestrian collisions. In IRCOBI Conference. Hannover, Germany 1997. pp. 133-149.
- Yedidiah, S. 2012. Centrifugal pump user's guidebook: problems and solutions. Springer Science & Business Media.
- Young, F.R. 1999. Cavitation. World Scientific.

- Yu, L., and Liu, H. 2004. Efficient feature selection via analysis of relevance and redundancy. *Journal of machine learning research* **5**(Oct): 1205-1224.
- Yuan, S.-F., and Chu, F.-L. 2006. Support vector machines-based fault diagnosis for turbo-pump rotor. *Mechanical systems and signal processing* **20**(4): 939-952.
- Yunlong, Z., and Peng, Z. 2012. Vibration fault diagnosis method of centrifugal pump based on emd complexity feature and least square support vector machine. *Energy procedia* **17**: 939-945.
- Zhang, G.P. Neural networks for classification: a survey. *In* Systems, man, and cybernetics, part c (applications and reviews). November 2000. IEEE. pp. 451-462.
- Zhang, N., Yang, M., Gao, B., Li, Z., and Ni, D. 2014. Unsteady pressure pulsation and rotating stall characteristics in a centrifugal pump with slope volute. *Advances in mechanical engineering* **6**: 1-11. doi:10.1155/2014/710791.
- Zhao, Hu, Lei, Y.G., and Zuo, M.J. 2010. Vibration-based fault diagnosis of slurry pump impellers using neighbourhood rough set models. *Proceedings of the institution of mechanical engineers, Part C: Journal of mechanical engineering science* **224**(4): 995-1006. doi:10.1243/09544062jmes1777.
- Zhao, Wang, Z., Lu, C., Ma, J., and Li, L. Fault diagnosis for centrifugal pumps using deep learning and softmax regression. *In* 12th world congress on intelligent control and automation (WCICA). Guilin, China, 12-15 June 2016. IEEE. pp. 165-169.
- Zouari, R., Sieg-Zieba, S., and Sidahmed, M. 2004. Fault detection system for centrifugal pumps using neural networks and neuro-fuzzy techniques. *Surveillance* **5**: 11-13.

Publications from the Present Research

Journal Papers

1. **Rapur, J.S.**, and Tiwari, R. 2017. Experimental Time-domain Vibration Based Fault Diagnosis of Centrifugal Pumps using SVM. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering 3(4): 044501-044507. doi:10.1115/1.4035440.
2. **Rapur, J.S.**, and Tiwari, R. 2018. Automation of multi-fault diagnosing of centrifugal pumps using multi-class support vector machine with vibration and motor current signals in frequency domain. Journal of the Brazilian Society of Mechanical Sciences and Engineering 40(6): 278. doi:10.1007/s40430-018-1202-9.
3. Panda, A.K., **Rapur, J.S.**, and Tiwari, R. 2018. Prediction of Flow Blockages and Impending Cavitation in Centrifugal Pumps using Support Vector Machine (SVM) Algorithms Based on Vibration Measurements. Measurement doi:https://doi.org/10.1016/j.measurement.2018.07.092.
4. **Rapur, J.S.**, and Tiwari, R. 2018. On-line Time Domain Vibration and Current Signals Based Multi-Fault Diagnosis of Centrifugal Pumps using Support Vector Machines, Journal of Nondestructive Evaluation, Springer.
5. **Rapur, J.S.**, and Tiwari, R. 2018. Experimental Diagnosis and Severity Estimation of Centrifugal Pump Faults using Multiclass Support Vector Machine and Wavelet Packet Transform Analysis. (paper needs to be resubmitted).
6. **Rapur, J.S.**, and Tiwari, R. 2018. Multi-fault diagnosis of combined hydraulic and mechanical centrifugal pump faults using continuous wavelet transform and support

vector machines. *Journal of Dynamic Systems, Measurement and Control*. ASME (provisionally accepted).

7. **Rapur, J.S.**, and Tiwari, R.. A review on condition based monitoring and fault identification in centrifugal pumps (paper under preparation).

Conference Papers

1. **Rapur, J.S.**, and Tiwari, R. 2016. Severity assessment and classification of blockage of centrifugal pumps in frequency domain of vibration data using support vector machine algorithms. *In Vibrations in Rotating Machinery - VIRM 11* Manchester, UK.
2. **Rapur, J.S.**, and Tiwari, R. Multi-Class Experimental Fault Severity Assessment and Classification in Centrifugal Pumps using Support Vector Machine Algorithm. *In International Conference on Condition Monitoring*, GITAM University, Vishakapatnam, IN, 26-27 October 2016.
3. **Rapur, J.S.**, and Tiwari, R. An intelligent and robust fault diagnosis system for identification of centrifugal pump defects in frequency domain using polluted vibration and current signatures. *In International Conference on Vibration Problems*. IIT Guwahati, Guwahati, IN. 29-November to 2-December 2017.
4. **Rapur, J.S.**, and Tiwari, R. 2017. A Compliant Algorithm to Diagnose Multiple Centrifugal Pump Faults With Corrupted Vibration and Current Signatures in Time-Domain. *In ASME Gas Turbine India*, ASME. Bengaluru, IN. 7 to 8 December 2017 (58516): V002T005A007. doi:10.1115/GTINDIA2017-4615.
5. **Rapur, J.S.**, and Tiwari, R. 2017. Wavelet Packets Based Adaptable Algorithm to Diagnose Multiple Independent and Coexisting Centrifugal Pump Faults with

- Corrupted Vibration and Current Signatures. *In* National Symposium on Rotor Dynamics. IIT Patna, Patna, IN. 12 to 13 December 2017.
6. **Rapur, J.S.**, and Tiwari, R. 2018. A Robust Algorithm to Detect Multiple Centrifugal Pump Faults with Corrupted Vibration and Current Signatures using Continuous Wavelet Transform. *In* 47th Turbomachinery and Pump Symposium. Texas A&M University, Houston, USA. 18 to 20 September 2018.
7. **Rapur, J.S.**, and Tiwari, R. 2019. Multi Fault Diagnosis of Centrifugal Pumps with Time, Frequency and Wavelet-Based Features Using Support Vector Machines. *In* Proceedings of the 10th International Conference on Rotor Dynamics Windsor Barra Convention Center, Rio de Janeiro, Brazil, 23 to 27 September 2018.– IFToMM. Springer International Publishing. pp. 43-57.