This is to certify that the
dissertation entitled

Fault Diagnosis and Failure Prognosis of Electrical Machines

presented by

Syed Sajjad Haider Zaidi

has been accepted towards fulfillment
of the requirements for the

Doctoral degree in Electrical Engineering

Major Professor's Signature

9 Aug 2010

Date

FAULT DIAGNOSIS AND FAILURE PROGNOSIS OF ELECTRICAL MACHINES

By

Syed Sajjad Haider Zaidi

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Electrical Engineering

2010

ABSTRACT

FAULT DIAGNOSIS AND FAILURE PROGNOSIS OF ELECTRICAL MACHINES

By

Syed Sajjad Haider Zaidi

Early detection, categorization and monitoring of faults can ensure safe and reliable operation, and increase the lifetime of a system. Fault is a condition corresponding to initial damage to a component or subsystem that, although does not affect the performance of it, can escalate to a failure. Diagnosis is the early detection of faults in the system and the assessment of its severity. On the other hand, failure prognosis is to identify the evolution of the fault condition and to predict the remaining useful life of the system.

The goal of this work is to develop a framework for fault diagnosis and failure prognosis which can detect and categorize the condition of an electromechanical system, and predicts its remaining useful life. In this work, methods are presented to identify transient faults using time-frequency analysis. The fault features are extracted from the motor current using Short Time Fourier Transform, Undecimated Wavelet Transform, Wigner Transform and Choi-Williams Transform. The presence of a fault is detected using spectrum energy density analysis and the categorization is performed by the pattern recognition classifiers, linear discriminant classifier and the nearest neighborhood classifier. The efficiency of each transform, to represent the underlying transient phenomenon, is compared by using Fisher discriminant ratio.

A prognosis algorithm is developed which predicts the remaining useful life of the system. Both the diagnosis and prognosis algorithms use the same time-frequency features extracted from the motor current. The prognosis algorithm is developed based on the statistical Hidden Markov Model. The model has three elements, state transition probabilities, state dependent observation densities and initial state prob-

ability distributions. Large data sets are required for the training of these elements, which are generally not available in the case of electromechanical systems. Methods are presented for the training of these elements from sparse data sets. For the computation of state transition probabilities, a method based on the Matching Pursuit decomposition is presented. The state dependent observation probability densities are defined as parametric densities and their statistics are computed from the experimental observations.

A survey of the state of the art diagnosis and prognosis methods is also presented in the dissertation. Possible faults in electromechanical system and their manifestation in the system parameters, and the experimental setup are also included. The proposed method is illustrated by examples using data collected from the experimental setup.

Dedication

Dedicated to the Twelfth Scion of Prophet Muhammad (PBUH)

Imam Mehdi (AS)

وَكُلَّ شَىْءٍ اَحْصَيْنٰهُ فِىْٓ اِمَامٍ مُّبِيْنٍ

"AND WE HAVE VESTED (THE KNOWLEDGE AND AUTHORITY) OF EVERYTHING IN THE MANIFEST IMAM."

The Holy Qur'an Verse 36:12

ACKNOWLEDGMENT

I would like to thanks to my advisor Professor Elias Strangas for his guidance, support, and his professional as well as personal example. He taught me how to ask honest questions about my own understanding, which was a great help. I really appreciate his efforts in transforming my attitude for research.

I would also like to thank the members of my PhD committee, Dr Selin Aviyente, Dr Hyder Radha and Dr Feeny Brian for their time. I am specially thankful to Dr Selin Aviyente for her guidance and support. She was a great help throughout the study period.

I would like to extend special thanks to those at National University for Science and technology (NUST) Pakistan and the Higher Education Commission (HEC) Pakistan for supporting my studies and to General Motor (GM) USA who helped make this research possible. I would like to thank Dr Mutasiem and Dr Kwang-kuen Shin for providing the funding for this work.

Special words of thanks go to my lab colleagues Carlo, Fan, Wes, Ramin and Shenalle. I would also like to thank my friends Attaullah, Khawar, Awais, Mabia, Ahmed Majeed, Sir Qaiser, Meena Bhabi, Sir Ajmal, Tariq, Jawad, Zabair, Waqar, Sir Hammad for always being there to share my pleasures and sorrows. I would also like to thank the members of the department whose help was much appreciated, including Brian, Gregg, Roxanne Peacock, and Pauline Van Dyke.

I owe special thanks to the love of my life my sister Auros and my brother Asad. I owe you fours years of life which I spent without you. I missed you each and every moment of my life. I am thankful to my in-Laws whose encouragement helped greatly.

I am obliged to my wife for all her efforts in making me a PhD. Her support was tremendous, she stood firm with me for all goods and bads of life, she encourage me great, and she was always well aware off and in conformation to all of my problems and commitments. I am thankful to her alot for being my partner.

I am thankful to motivating force of my life, my kids, Mehak, Kanwal, Hamza and Midhat. You guys are the greatest encouragement for me and you always cause me to think reasonably. Thanks for sharing PhD efforts with me and your mom.

All my efforts would have been in vain without the prayers of my father. He is the inspiration of my life and what all I am today, that is because of him. He my best friend, my teacher, my love and my inspiration.

Finally, I have no words to thank my late mother. All I can say is

"This was only because you wished for it,

I love you, miss you".

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Overview and Objectives of the Thesis

The present day world economics have ever increasing demand for cost-effective operation of critical equipment. The last few years have witnessed an increased interest in the reliable and safe operation of complex systems. In order to achieve this goal, fault diagnosis and failure prognosis have become key issues of interest for research. A fault is a condition corresponding to initial damage to a component or subsystem that, although does not affect the performance of it, can escalate to a failure. Fault diagnosis is the early detection of fault in the system and the assessment of its severity. On the other hand, failure prognosis is to identify the evolution of the fault condition and to predict the remaining useful life of the system. Diagnosis and prognosis share commonalities and generally prognosis is a succeeding activity of diagnosis. In this thesis, fault diagnosis and failure prognosis are collectively referred to as fault analysis

In many engineering and non engineering fields, fault diagnosis and failure prognosis find applications. Common applications of these methods can be found in the medical fields, which are the first to use diagnosis and prognosis tools for their subjects, the patients. Other examples are electro-mechanical systems, structural health

assessment systems, computer software fault detection and prediction methods, and manufacturing systems.

The task of fault diagnosis and failure prognosis becomes more challenging if the system is complex, nonlinear, noisy and contains different subsystems. Such system might not be modeled without extensive efforts and without significant approximations. The prognosis algorithms are difficult to realize in the absence of the system model. Moreover, most of the prognosis methods need huge amounts of historical data in order to implement them.

The principle objective of this thesis is to develop a generic method to deal with the problem of prognosis of complex systems, using hidden Markov model (HMM), a statistical tool for the probabilistic estimation of failure state, which works even with non linear systems in the presence of non Gaussian noise. The task was accomplished and specific objectives were successfully completed.

Diagnosis is not necessarily the preceding action of failure prognosis, but in most of the applications it is the preliminary phase. The first objective of this work was to develop an efficient diagnosis algorithm using non intrusive methods for complex nonlinear systems. The algorithm should be able to detect the presence of fault and should classify its severity. Supervised learning method is the overall approach, in which labeled data are available for algorithm training.

The second objective was to develop methods for the computation of HMM parameters. Generally, this needs large amounts of historic data to compute these parameters. In this work, methods for the parameters computation are developed using heuristic and experimental approaches with sparse data sets.

The third objective was to develop an algorithm which can estimate the probability of the next state and remaining useful life (RUL) of the system. Given the model and the sampled information, the algorithm computes RUL in terms of probability of the failure state.

In this work the starter of automobiles is the target system. A laboratory experiment was built for the system and data were collected for the validation of the proposed methods. Different signals were sampled from systems in healthy conditions and systems with seeded fault conditions. The developed prognosis algorithm was tested, results were obtained and conclusions are made.

1.2 Principal Contributions

This thesis presents non intrusive fault diagnosis and failure prognosis methods based on the supervised learning approach. The motor current signature analysis is performed and the fault features are extracted in the time frequency domain. The same health indicator is used for fault diagnosis and failure prognosis. In particular, the principal contributions of this work are:

1. **Selection of a suitable signal transform to represent transient repetitive faults in the electromechanical system:** In this work, four candidate transforms were compared for the extraction of fault features. A selection criterion based on the Fisher discriminant ratio is developed and the most suitable transform is identified.

2. **A general framework for fault diagnosis:** A complete framework for the diagnosis is presented for complex electromechanical systems having repetitive transient faults under varying load conditions. Different classifiers were evaluated for the classification of the transient faults. The computation efficiency and accuracy of classification is evaluated for each classifier.

3. **Methodologies to estimate HMM parameters are developed in presence of sparse data:** In this work, HMM is used for prediction of the failure state. Methodologies to compute the model parameters from sparse data are

developed. These methodologies use training data and classifier training output to compute the model parameters.

4. **A HMM based algorithm for failure prognosis capable of estimating the probability of failure at future time (RUL):** The remaining useful life (RUL) is estimated in terms of the failure state probability at the time of each sample. The algorithm uses the collected sample's features at the present state and the HMM parameters, initial state probabilities, state transition probabilities and state dependent observation densities, to compute the failure state probability.

5. **The validation of the proposed algorithm using experimental and simulated data.** The proposed algorithm is demonstrated using two different data sets, experimental and simulated. The simulated data are generated using the statistics obtained from the actual experimental data. The results obtained from the algorithm are presented and conclusions are made.

1.3 Organization of the thesis

There are four major parts to this thesis. The first part, contained in Chapter 2, provides a general overview of the literature related to the fault diagnosis and failure prognosis techniques. In addition, this chapter gives an account of the state-of-the-art techniques used for the development of fault diagnosis and failure prognosis algorithms. The second part is contained in Chapter 3 and Chapter 4. In Chapter 3, the fault diagnosis and failure prognosis problem is formulated and the selected approach is presented for fault analysis of the complex electromechanical systems. It explains the hierarchy of the algorithm, what selections are required and why they are instituted. The second chapter of this part, Chapter 4, illustrates the hardware

built for implementation of the diagnosis and prognosis algorithms. It explains the setting up of the electromechanical system in the laboratory, faults of the system, its operation and control, the sampled signals and sensors.

The third part of this thesis is comprised of Chapters 5 and 6, where a diagnosis and prognosis framework for electromechanical system is presented. Chapter 5, introduces the features extraction methods using time frequency distributions and diagnosis approaches using pattern recognition classifiers. Comparison of four distribution methods is performed and computed values of the Fisher discrimination ratio are presented. The fault diagnosis methods, with and without detection of fault event, are presented. The classification results using different classifiers are presented. In Chapter 6, a statistical modeling based prognosis method is presented. In this chapter, methodologies of the computation of model parameters from empirical data and algorithm for the estimation of RUL are presented. The chapter also contains examples illustrating the implementation of the developed methodologies on the sampled data. The last part of the thesis, Chapter 7, states the conclusions and suggested future work.

Chapter 2

Background

2.1 Scope and Objective of the Chapter

Fault diagnosis and failure prognosis have become an active field of research. Different diagnostics and prognostics methods, algorithms and techniques have been proposed in the literature. For the completeness of this thesis, it is considered essential to present a review of the related literature. Moreover, the developed diagnosis and prognosis methods involve different concepts related to time frequency analysis, pattern recognition, statistics and estimation. A comprehensive theoretical review of the related concepts will provide a better understanding of the problem. The objective of this chapter is to present a review of the literature and the theoretical concepts related with fault diagnosis and failure prognosis.

With the intention to achieve this objective, this chapter is arranged two sections. In Section 2.2 the literature review is presented, and in Section 2.3, theoretical concepts related to the feature extraction methods, pattern recognition classifiers and prognosticators are presented.

2.2 Literature Review

In this section, fault diagnosis and prognosis techniques are presented. There are two major categories of these techniques, intrusive and non-intrusive. The intrusive techniques are generally the classical methods for the fault diagnosis. Signatures in the vibration of a machine are often used to detect mechanical faults. These techniques require the installation of an accelerometer, which can be bulky and adds to cost. In [1], a three phase induction machine with a gearbox and its corresponding bearing assemblies are analyzed. The wavelet transform using the Daubechies 4 mother wavelet was applied to the fast Fourier transform (FFT) of the accelerometer output. The details coefficients at the first scale were the input to an artificial neural network (ANN) used for classification. The ANN was trained to detect faults including the presence of a small 'blip' of $2mm$ diameter welded onto a gear tooth, a triangle shaped area missing from a gear tooth, and a fractured inner race of the bearing housing. This technique was implemented offline.

For induction machines, parameter estimation is a typical method for condition monitoring [2]. For the parameter estimation the classical methods are locked-rotor test, no-load test or the DC test. These tests are intrusive, need special equipments and are to be conducted under off-line condition.

Other common intrusive methods for fault detection are temperature monitoring [3, 4], tagging compounds[5], high frequency injection[6, 7], axial leakage flux [8] and air gap flux signature analysis[9] and vibration signatures analysis[10, 11, 12]. Intrusive methods require installation of additional equipment, which is costly and may not be practical in many applications due to the nature of operation of the systems. Therefore, non-intrusive fault analysis is an attractive alternative.

2.2.1 Non-Intrusive Methods

Non intrusive analysis methods do not require additional sensors and installations. They only use voltage and current measurements from motor terminals and these signals are ready available. For a number of machines the stator current has been the monitored quantity, often without relating it to the underlying electromagnetic phenomena. In recent years, non intrusive fault detection methods have attracted interest.

Fault analysis methods can be divided in three groups, model based, signal based, and data based. Signal processing is an enabling technology for all three but with different impact and role. Moreover, with advances in digital technology over the last few years, adequate data processing capability is now available on cost-effective hardware platforms.

2.2.1.1 Model Based

Model-based diagnosis relies on a theoretical analysis of the machine whose model is used to predict fault signatures. The difference between measured and simulated signatures is used as a fault detector. Residual analysis and suitable signal processing are used to define a fault index.

The signal processing methods used in [13] were based on a Condition Monitoring Vector Database to find the presence of broken rotor bars in induction machines. First a set of Condition Monitoring Vectors (CMV) were determined through simulations using the time-stepping Finite Element (TSFE) technique, and a single vector was computed for each complete AC cycle, both in the presence and absence of a fault. The CMV is defined in (2.1),

$$CMV = \left[\frac{V_n}{V_p} \; \frac{I_n}{I_p} \; \frac{Z_n}{Z_p} \; A\left(f_{LSB}\right) \; \Delta\delta_{BB} \; \Delta\delta_{SC} \; \omega_m \; T_{dev} \right] \tag{2.1}$$

where V, I, and Z with the subscripts n and p are the negative and positive sequence components of the stator voltages, currents, and associated impedances; $A\left(f_{LSB}\right)$ is the amplitude of the low sideband frequency spectrum component of the stator current at the frequency $(1-2s)f_s$, where f_s is the power supply frequency; $\Delta\delta_{BB}$ and $\Delta\delta_{SC}$ represent the range of oscillation of the resultant mid air-gap magnetic field for broken rotor bars and stator winding inter-turn faults; and ω_m and T_{dev} are the motor speed and developed motor torque respectively. Finally, an artificial intelligence-based statistical machine learning approach, using Gaussian Mixture Models, was used to train a Bayesian maximum likelihood classifier. Experimental results showed that the algorithm could discern between various numbers of broken rotor bars.

In [14], brushless DC (BLDC) machines were analyzed using parameter estimation in a model-based technique. Based on the inverter supply voltage, the DC current, and the mechanical speed, a least-squares method was used to estimate parameters in a model of the machine. In the model for the electrical subsystem (2.2), estimates of R and k_E were obtained.

$$\bar{v}(t) = R\bar{i}(t) + k_E\omega_r(t) \tag{2.2}$$

In the model for the mechanical subsystem (2.3), estimates for J, c_C, and c_v were obtained.

$$J\dot{\omega}_r(t) = k_T\bar{i} - c_C\text{sign}(\omega_r(t)) - c_v\omega_r(t) \tag{2.3}$$

Here, c_C is the Coulomb friction coefficient, c_v is the viscose friction coefficient, and it is assumed that $k_T = k_E$. From the electrical model, the authors could determine whether the phase resistance of all coils had increased, indicating an increase in stator temperature, or a broken coil. From the mechanical model, the authors could detect increases in Coulomb and viscose friction.

Permanent magnet brush DC machines are analyzed for the presence of an open

phase/broken connector fault, shorting of adjacent commutator bars, and worn brush faults in [15]. This approach is model-based and uses block-pulse function series techniques to estimate parameters in a continuous-time system. This is advantageous as it eliminates the need to discretize the system so that an algorithm like the least-squares method can be used. Based on measured DC current, DC voltage, and mechanical angular velocity, estimates of the armature resistance and inductance, back-EMF coefficient, rotor inertia, and friction coefficient are obtained from the motor model. These parameters are passed to an ANN to determine the fault type. This system was implemented offline.

FEA (Flux2D) was used in [16] to determine the fault signatures in the DC bus voltage and current, stator currents, torque, and speed in a trapezoidal BLDC machine. The following faults were studied: Single-phase open circuit fault; phase-to-phase terminal short circuit; and internal turn-to-turn short circuit (across 6/26 turns of one coil). The machine was a 3-phase, 6 pole, 18 slot machine with a bifurcated stator tooth structure (helping to reduce torque ripple and doubling the frequency of cogging). The rotor of the machine had 6 Nd-Fe-B magnet poles magnetized radially. Six-step / voltage mode control was used. In this type of control, each switching pattern lasts 60 electrical degrees, and with a 6-pole machine, 20 mechanical degrees. An experimental setup was used to validate the FEA model.

A model-based approach based on wavelet analysis to detect faults in induction motors is presented in [17]. The faults analyzed are one broken rotor bar, two broken rotor bars, stator short circuit in single phase and stator short circuit in two phases. Application of the wavelet transform helped to remove variations due to changes in speed. The temporal coincidence of wavelet coefficients at different scales corresponding to an event was used for fault detection in this paper.

2.2.1.2 Data Based

Data-based diagnosis does not require any knowledge of machine parameters and model. It only relies on signal processing and on clustering techniques. Data sampled from the actual machine are processed to extract a set of features that are clustered in order to classify them. Eventually, decision process techniques are used to define a fault index. AI and pattern recognition techniques are widely used to achieve the above purposes

2.2.1.3 Signal Based

Signal-based diagnosis looks for known fault signatures in quantities sampled from the actual machine. Then, these signatures are monitored by suitable signal processing. Typically, frequency analysis is used, although advanced methods and/or decision-making techniques can be of interest. Here, signal processing plays a crucial role as it can be used to enhance signal-to-noise ratio and to normalize data in order to isolate the fault from other phenomena and to decrease the sensitivity to operating conditions.

The signal based techniques, which have received the most attention in the recent years, are implemented in the frequency domain, time domain and time-frequency domain. In [18] classical and modern diagnosis methods for machine faults based on signal processing techniques are presented.

2.2.1.3.1 Time Domain Analysis Time-domain analysis is a powerful tool for machine fault diagnosis as they offer a lower computational cost and, thus, require a reduced time acquisition period. Faults in induction machines are diagnosed by analyzing the starting current transient under a no-load conditions, where the only measurable and useful information exists in the large starting transient current of the motor. In [19], the oscillation of the electric power in the time domain becomes

mapped in a discrete waveform in an angular domain. Data-clustering techniques are used to extract an averaged pattern that serves as the mechanical imbalance indicator. The maximum covariance method is another technique that is based on the computation of the covariance between the signal and the reference tones in the time domain. However, in the case of nonstationary signals, these methods may not be very helpful.

In [20], a method to detect turn-turn insulation failures in induction machines was described. The line-neutral stator voltages were measured and filtered to remove the fundamental component of the machine excitation voltage. The RMS value of $v_{sum} = v_{an} + v_{bn} + v_{cn}$ of the filtered components is zero in a balanced machine, however in the case of turn-turn insulation failures, the number of shorted turns could be determined by the amplitude. This technique was implemented offline, and requires that the machine be star connected with the neutral accessible.

The Park's vector pattern of the currents is analyzed for stator voltage unbalance or an open phase in three-phase induction machines in [21]. The Park's vector pattern is plotted in the stationary frame of reference using a two-phase representation of the measured stator currents for one electrical cycle. The plot is analyzed using an ANN to check for a stator voltage unbalance or an open phase. The occurrence of either of these faults manifests itself in the deformation of the current Park's vector pattern corresponding to a healthy condition. This deformation leads to an elliptic pattern, whose major axis orientation is associated with the faulty phase. The severity of the deformation helps to distinguish between the two faults, with the open phase fault the most deformed. A mathematical model of the induction motor is not required. This system is implemented offline.

2.2.1.3.2 Frequency Domain Analysis Frequency domain analysis is a popular method in machine diagnosis. There are three main subclasses of this method: non-

parametric methods, parametric methods, and high-resolution methods. Fourier analysis and optimal bandpass filtering are the nonparametric methods. Autoregressive-moving-average model is a type of parametric methods which is employed for the estimation of linear time invariant systems from noise. Multiple signal classification (MUSIC) and eigenvector are called high-resolution methods[22].

Frequency domain analysis was used in [23] to analyze three-phase induction machines for the presence of broken rotor bars or end rings. An FFT is performed on the currents, and a diagnostic index is assigned equal to the sum of the amplitude of the two sideband current components at $(1 \pm 2s)f$, where s is the slip of the machine, and f is the fundamental component of the current. If the value of the diagnostic index exceeds a threshold, it is determined that either a broken rotor bar or end ring is present. Knowledge of the main nameplate data of the machine as well as the number of bars is required for this system. This system is implemented offline.

Stator line currents, circulating currents between parallel stator branches and forces between the stator and rotor were analyzed. These quantities were computed using finite element analysis (FEA) and noise was added to them (0 mean, 3% variance of the amplitude of the current). The power spectrum of each signal was used for classification. Faults analyzed included shorted turns, shorted coils, broken rotor bars, broken end rings, rotor eccentricities, and asymmetrical line voltages. Analysis was performed for both 35kW and 1600kW induction motors. Simulation results showed that classification of faults based on any of the above parameters was possible. Experimental results showed that classification of faults based on stator line currents was possible only when the measurement data was used for both training and testing of the classifier.

FEA was used in [24] to calculate electromagnetic variables and parameters, used to detect mechanical faults in PMAC machines. These parameters were flux linkages and inductances as functions of rotor position, and were stored in a look-up table

and then used in a transient simulation of the motor. This process was repeated for various rotor conditions. Faults were then directly mapped to the increase in harmonics. This technique detects the presence of static and dynamic eccentricities, and flux disturbances originating from defects to the permanent magnets. The authors determined that the current harmonic components can be analyzed in either the rotor or stationary reference frame. In the case of the static eccentricity, however, analysis in the stationary reference frame was required, since the fault effects in the rotor reference frame were obscured.

In [25], tool fracture in drilling operations, which is the chipping, breakage, or severe deformation of the cutting edges of a drill bit, is analyzed. When tool fracture occurs, a larger motor torque is generated due to the tool acting on the broken cutting edges. The authors show that tool fracture can be observed through the estimated motor torque. The torque is approximated as the square of one of the lowpass filtered, rectified, phase currents. The cutoff frequency of the lowpass filter is 166Hz. Increased amplitude of the estimated torque can indicate the presence of a tool fracture.

In [26, 27, 28], various diagnostic techniques are used to detect inner and outer race faults in bearings. It is of interest to determine the signatures in the spectrum of the vibration due to the bearing defects. In these papers, the expected frequencies associated with the occurrence of the faults are lower than $250Hz$. It can be observed in the spectrum, however, that excitation not only occurs at the expected frequencies, but also at higher frequencies, up to $12kHz$. The expected frequencies are based on the frequency of occurrence of the faults, i.e. the inverse of the period which successive balls contact one point in the race. The higher frequencies are due to the event that occurs each time a ball contacts the damaged part of the race. In the present research, fault identification is based on the higher frequencies due to the individual events.

In [29, 30, 31], various diagnostic techniques are used to detect faults in gears. The faults pertain to a single tooth that is missing, cracked, or has a notch cut in it.

Of particular interest are the signatures in the spectrum of the vibration due to the gear defects. In these papers, the tooth meshing frequency, i.e. the frequency which the defective tooth makes contact, ranges from $250 - 500Hz$. This can be considered to be the frequency of occurrence of the fault. It can be observed, however, that significant excitation occurs not only at the tooth meshing frequency, but at higher frequencies as well, up to approximately $5kHz$. The higher frequencies are due to the event that occurs each time the defective tooth makes contact.

Support Vector Machine (SVM) based classification, [32], was used in [33] to classify between faults in induction machines. SVMs are used to map a set of coefficients to a high-dimensional feature space where a set of 'best' separating hyperplanes are constructed. This mapping is based on a set of kernel functions, whose selection is critical. SVMs are binary classifiers, and hence, for multiclass problems, either one-against-all or one-against-one classifiers are developed. In this work, the one-against-one formulation was used requiring $n(n-1)/2$ classifiers, e.g. for 6 faults (and healthy), 21 classifiers would be required.

Several related techniques have been presented in [34, 35, 36] in which three-phase induction machines are analyzed. In [34], a single phase of stator current is monitored. The stator current spectrum is analyzed for the presence of harmonics indicating the possibility of two types of faults. In the case of air gap eccentricities, harmonics with frequencies predicted by (2.4)

$$f_{ecc} = f_e \left[1 \pm k \left(\frac{1-s}{\frac{p}{2}} \right) \right] \tag{2.4}$$

are observed, where f_e is the electrical supply frequency, $k = 1, 2, 3, \ldots$, s is the per unit slip, and p is the number of machine poles. In the case of rolling-element bearing

defects, harmonics with frequencies predicted by (2.5)

$$f_{bng} = |f_e \pm m \cdot f_v| \tag{2.5}$$

are observed where $m = 1, 2, 3, \ldots$ and f_v is one of the characteristic vibration frequencies. While this technique is performed offline, it is expanded upon in [35] to be an online diagnostic. Additional frequencies are monitored to detect broken rotor bars (2.6)

$$f_{brb} = f_e \left[k \left(\frac{1-s}{\frac{p}{2}} \right) \pm s \right] \tag{2.6}$$

where due to the normal winding configuration, $\frac{2k}{p} = 1, 5, 7, 11, 13, \ldots$. A selective frequency filter is added along with an ANN. The selective frequency filter provides the ability to discern between frequency bands possibly related to specific faults. The filter has a learning stage where an adaptive threshold is calculated from the FFT components. The frequency components which exceed the calculated threshold are placed in a table. It also has a reducing stage where it is decided which table entries will be provided as inputs to the ANN. The ANN has a training phase where is forms clusters which represent valid motor operating conditions. As the ANN is exposed to more and varied operating conditions, the number of acceptable classifications increases. After the training phase, the ANN switches to a fault-sensing mode. When a spectral signature falls outside the trained clusters, it is tagged as a potential motor fault. The ANN alarms the user only after multiple indications of a potential fault have occurred protecting the ANN from alarming on random signals which have been incorrectly identified. The techniques presented in [34, 35] do not require information on the motor or load characteristics. In [36], the authors combine the techniques applied in [34, 35], and expand upon them. In this case, all three stator currents as well as all line-line stator voltages are measured. A model-based thermal failure

protection algorithm is added. This system was implemented online. It is claimed that the system can operate with any motor design and load condition with only nameplate motor data.

2.2.1.3.3 Time Frequency Domain Analysis Recently, the application of signal processing techniques different from frequency analysis has been proposed to diagnose machine faults [37]. Time-frequency analysis consists of the 3-D time, frequency, and energy representation of a signal, which is inherently suited to indicate transient events in the signal. The Wigner distribution and its various permutations is an analysis technique that has been widely used in the detection of faults in mechanical systems [38, 39, 40, 41] together with wavelet transform and Hilbert–Huang transform[42]. Reasonable success was reported using wavelets to extract fault information from the stator current prior to classification. The problems of translation variance and the inability to closely approximate sinusoidal signals make their use difficult in rotor fault detection. In [40], the detection of rotor faults was investigated in electrical machines operating under continuously changing operating conditions. This allows an efficient diagnosis in every condition and not only during the motor start-up. In summary, the accurate signal processing of the electrical quantities acquired for monitoring purposes in the diagnostic process is a key issue.

A set of wavelets were introduced based on the shapes of widely encountered transient phenomena in the eddy currents on the surface of a steel mill in [43]. These new wavelets are high frequency oscillations enveloped by single and double-sided exponentials, a cosine-tapered rectangle, and Gaussian, Hanning and Hamming functions. A time-frequency scale distribution was developed and its power distribution was mapped onto a three dimensional image. Through analysis of these image patterns, detection and classification of faults is possible. Since the wavelets match specific transients, the modified wavelet transform can have a high sensitivity for certain

applications.

The techniques presented in [44, 45, 46, 47] were developed to detect the presence of several faults in automotive permanent magnet brush DC machines, mostly attributed to improper assembly. Both windshield wiper motors and fuel pump motors were used. The input coefficients to three fault detection algorithms were the modulus maxima of the first ten scales of the wavelet transform of the DC current. The classifier used were the decision tree, the nearest neighbor rule and the linear discriminant classifier

In [48], the authors used different time frequency transformations, such as Undecimated Wavelet Transform [49], Short Time Fourier Transform, Choi-Williams Transform [50] and Wigner-Ville Transform, in combination with the linear discriminant classifier, nearest neighborhood classifier and multiple discriminant based classifier [51] for the early identification and classification of the non-stationary electrical and mechanical faults in permanent magnet AC machines.

In [52], broken bar faults in induction machines are diagnosed using time frequency analysis of stator current and spectral power density as the health indicator. In [53], the authors proposed empirical mode decomposition of stator current to extract the fault frequencies. In [54] the authors proposed stator current decomposition using both the continuous and discrete wavelet transform for the detection of demagnetization in permanent magnetic synchronous machine, which a non-stationary fault. The motor current analysis is carried out for the detection of fault in the mechanical components attached to electrical machines. In [55] the authors proposed discrete wavelet transform based method for broken rotor bar detection in induction machines, which does not require slip information if only the low frequency bands are analyzed. The squared instantaneous magnitude of the stator current and the squared stator-current space-vector magnitude, are proposed as health indicators. In [56], an efficient method in terms of computational effort and memory requirements

is proposed for the diagnosis of machine faults using MCSA.

Although, diagnosis and prognosis are closely related and share commonalties, fault prognosis is a relatively new field of research. Extensive work has been done on diagnosis, i.e the early detection of faults in electrical machines, applying a variety of techniques as mentioned in preceding paragraphs. On the other hand, prognosis of failure has been further explored only recently.

2.2.2 Prognosis Methods

In practical terms prognosis, i.e. estimation of the remaining useful life (RUL) and prediction of the future state, is the next logical step to fault diagnosis and central to determining the need for timely maintenance, employment of auxiliary systems when such exist, or simply cease operation to avoid dangerous failure. Prognosis algorithms can also be classified as model based and data driven. In model based approach either mathematical models or models based on the equipment physics, or both, are used. The fault is considered continuously variable, whose evolution is defined by a deterministic or stochastic law.

One common model based prognosis technique uses particle filters [57, 58]. Orchard and Vachtsevanos in [59] used particle filters for the prediction of RUL of rotary systems having cracks. In [60], fuzzy logic classifier is used for the diagnosis of inverter faults and can be extended to prognosis. The misfiring in the Voltage Source Pulse Width Modulation inverter (PWM-VS) is detected by the fuzzy fault classifier, which compares the patterns generated from the output current from the healthy and faulty inverters. In [61] models of Time Dependent Dielectric Breakdown (TDDB) and Hot Carrier are used for the life estimation of the semiconductor devices. In [62], a Preisach model is used to estimate the state of charge of the batteries. In [63], the authors used a genetic algorithm for the monitoring and fault detection of an induction motor drive. In this diagnostic method, the motor current is analyzed

in the frequency domain and the decision is made by an expert system using fuzzy logic. In [64], data-driven similarity-based prognostic techniques were used for the RUL estimation. The system/equipment life patterns are obtained from the run-to-failure data, using the training data sets. Prognosis from a test sample is performed by pattern matching. Another data driven tool for prognosis is neural networks and neuro-fuzzy systems [65, 66, 67] and have been used for predicting the health and RUL of machines, bearings and batteries.

2.2.3 HMM as classifier and prognosticator

Hidden Markov Models (HMMs), which are used here as the prognosis tool, were introduced by Rabiner [68], as a powerful statistical modeling tool, having its main strength in its doubly statistical nature. HMMs are extensively used for speech recognition [69], hand gesture recognition [70] and text segmentation. They have been used for fault diagnosis, classification and prognosis, tool wear detection/prediction [71] and bearing faults monitoring [72]. In [73] and [74], HMMs are used for the structure damage classification where as HMMs were employed for monitoring machine tool wear in [75] and [76]. In [77] the authors use HMMs for the diagnosis of machine faults using features extracted from time frequency distributions. Easy model interpretation is a competitive edge of HMMs over the commonly used black-box modeling techniques such as neural network[78].

2.3 Theoretical Background

This section presents the theory related with the different tools used for the diagnosis and prognosis algorithm. Motor current signature analysis is performed for the machine health assessment. The fault features are extracted in time frequency domain and different candidate distribution and transforms are compared. Section 2.3.2,

presents the pattern recognition classifier and in Section 2.3.3, the prognostication method are discussed.

2.3.1 Features Extraction Methods

2.3.1.1 Short Time Fourier Transform

Fourier Transform provides a good representation of signals in the frequency domain, since it decomposes the signals in terms of complex exponentials at different frequencies. However, this representation does not contain any information about the occurrence of a particular frequency in time. This idea is illustrated in Figure 2.1, with the time-frequency tiling showing the joint time and frequency resolution provided by the Fourier transform. The faults studied in this work manifest themselves

Figure 2.1: Fourier Transform Tiling

as repeated short transients superimposed on the stator currents. Analysis of these

short transients, however, requires information in both frequency and time. The inability to provide time localization of a signal is a fundamental limitation of the FFT[79].

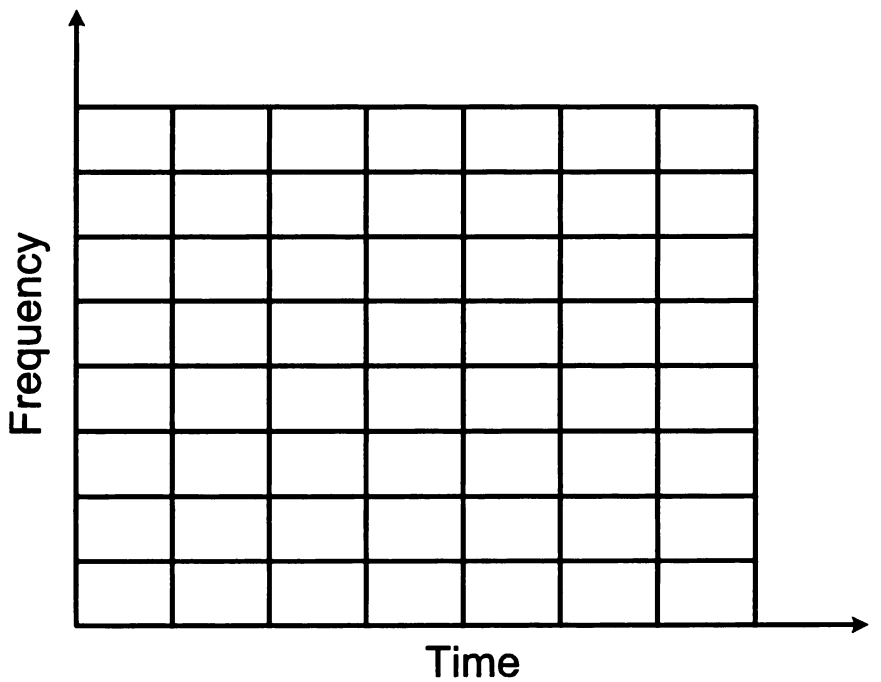

Figure 2.2: STFT Tiling

The short-time Fourier transform (STFT) is an extension of the Fourier transform, allowing for the analysis of non-stationary signals. Here, the signal is divided into small time windows, and each is analyzed using the FFT as follows:

$$STFT(t,f) = \int_{-\infty}^{\infty} h(t-\tau)x(\tau)e^{-j2\pi f\tau}d\tau \tag{2.7}$$

where h is the window function. This formulation provides the localization in time, while simultaneously capturing the frequency information. The resultant coefficients of the STFT are intuitive and easy to correlate with the original signal. Tiling for the STFT is given in Figure 2.2. The time-frequency tiling for the STFT is uniform

across time and frequency. In the implementation of the STFT, a design tradeoff must be made between time and frequency resolution. This is due to the principle of uncertainty, which limits the lower bound on the time-bandwidth product.

$$TB \geq \frac{1}{2} \tag{2.8}$$

The duration, T, and bandwidth, B represent the broadness of a signal in time and frequency. Good time localization, requires selection of narrow window in time domain, *h(t)*, whereas for good frequency resolution narrow window in frequency domain, *H(w)* is needed. But both *h(t)* and *H(w)* cannot be made arbitrarily narrow, causing an inherent trade-off between time and frequency localization[79].

2.3.1.2 Discrete Wavelet Transform

Wavelet analysis [49] is also suitable for non-stationary signals. The DWT has greater flexibility than the STFT. Different basis functions, or mother wavelets, may be used in wavelet analysis while the basis function for Fourier analysis is always the sinusoid. Unlike sinusoids, wavelets have finite energy concentrated around a point. One can choose, or design a wavelet to achieve the best results for a specific application.

Tiling for the DWT is shown in Fig. 2.3. Unlike Fourier methods, the tiling for the DWT is variable allowing for both good time resolution of high frequency components, and good frequency resolution of low frequency components in the same analysis.

The Hilbert space of measurable, square-integrable functions, $f(x) \in L^2(\mathbf{R})$, is defined in (2.9).

$$\int_{-\infty}^{+\infty} |f(x)|^2 dx < +\infty \tag{2.9}$$

A basis for a space $\mathcal{V}$ is defined as a set of linearly independent functions that span the space. That is, any function in $\mathcal{V}$ can be written as a linear combination of

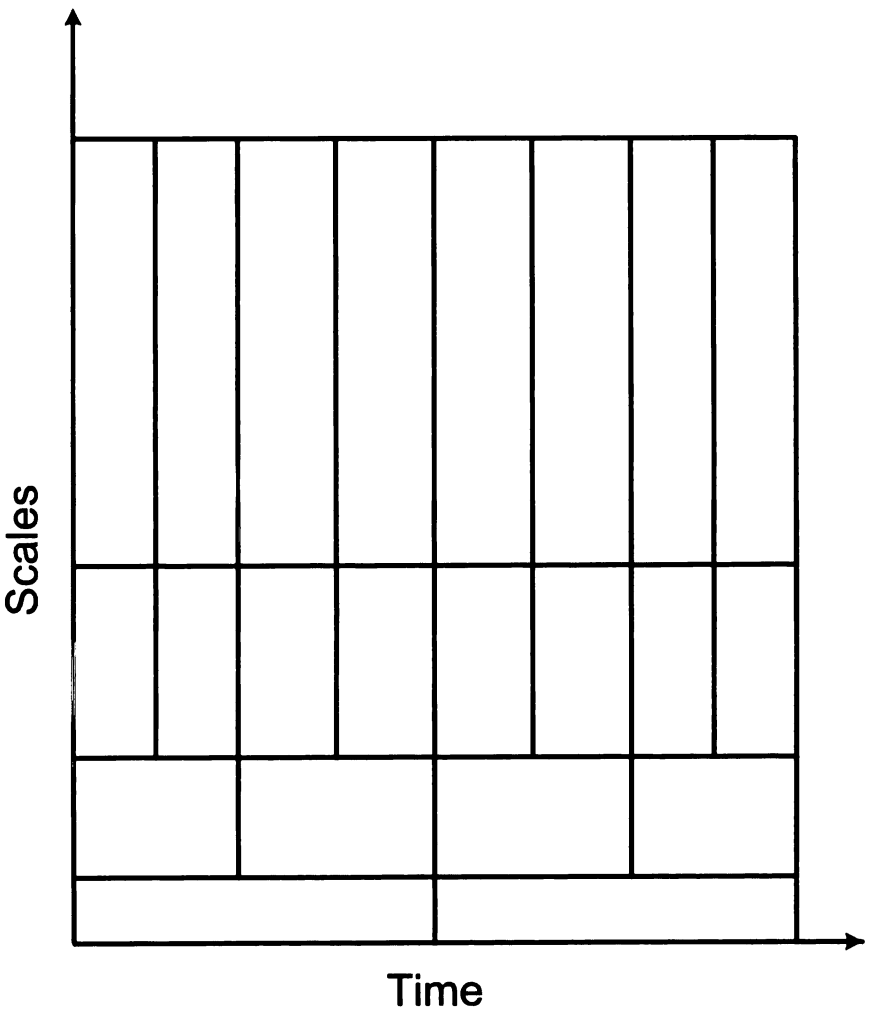

Figure 2.3: DWT Tiling

the basis functions. This can be illustrated by a linear decomposition (2.10), where $f(t)$ represents any function in the space $\mathcal{V}$, $\psi_\ell(t)$ are the basis functions, and a_ℓ are the scaling coefficients.

$$f(t) = \sum_\ell a_\ell \psi_\ell(t) \tag{2.10}$$

The discrete wavelet transform can be defined using the idea of multiresolution by starting with the scaling function and defining the wavelet function in terms of it. A basic one-dimensional scaling function can be designed to translate a function in time (2.11) where $\mathbf{Z}$ is the set of all integers.

$$\varphi_k(t) = \varphi(t-k) \quad k \in \mathbf{Z} \quad \varphi \in L^2 \tag{2.11}$$

Wavelet systems are two-dimensional, so a scaling function $\varphi_{j,k}(t)$ that both scales and translates a function $\varphi(t)$ is defined in (2.12),

$$\varphi_{j,k}(t) = 2^{j/2}\varphi\left(2^j\left(t - 2^{-j}k\right)\right) \quad j,k \in \mathbf{Z} \quad \varphi \in L^2, \tag{2.12}$$

where j is the $\log_2$ of the scale and $2^{-j}k$ represents the translation in time. A subspace of the $L^2(\mathbf{R})$ functions can be defined as the scaling function space $\mathcal{V}$. $\varphi_{j,k}(t)$ spans the space $\mathcal{V}_j$, meaning that any function in $\mathcal{V}_j$ can be represented by a linear combination of functions of the form $\varphi_{j,k}(t)$.

When discussing scaling functions in terms of multiresolution analysis the relationship between the span of scaling functions with different indices can be seen in (2.13-2.14).

$$\cdots \subset \mathcal{V}_{-2} \subset \mathcal{V}_{-1} \subset \mathcal{V}_0 \subset \mathcal{V}_1 \subset \mathcal{V}_2 \subset \cdots \subset L^2 \tag{2.13}$$

$$\mathcal{V}_{-\infty} = \{0\}, \quad \mathcal{V}_\infty = L^2 \tag{2.14}$$

Through this scaling, if a function $f(t) \in \mathcal{V}_j$, then $f(2t) \in \mathcal{V}_{j+1}$. In the case of the

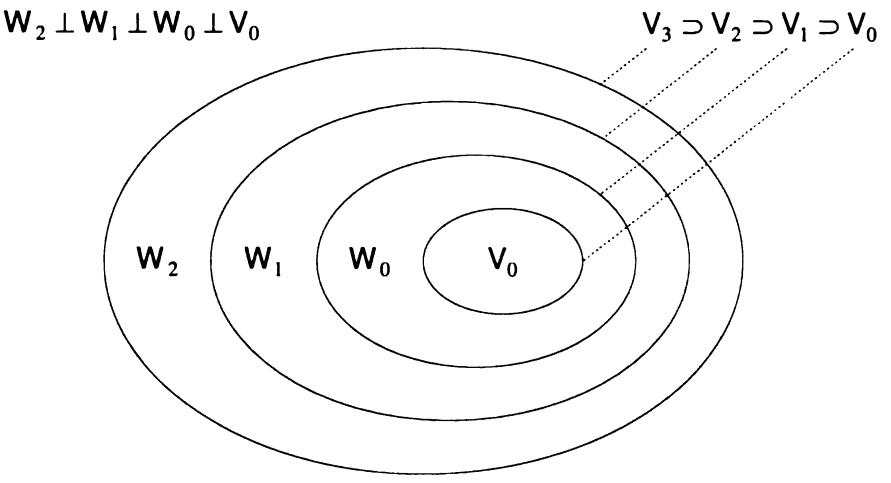

Figure 2.4: Scaling Function and Wavelet Vector Spaces

scaling function, $\varphi(t)$ is written as a function of $\varphi(2t)$ in (2.15),

$$\varphi(t) = \sum_n h_0(n)\sqrt{2}\varphi(2t-n), \quad n \in \mathbf{Z} \tag{2.15}$$

where h_0 is a set of coefficients discussed in Section 2.3.1.3.

Another subspace of the $L^2(\mathbf{R})$ functions is the wavelet vector space $\mathcal{W}$. A wavelet spans the space $\mathcal{W}_j$, which represents the difference between two scaling function spaces, $\mathcal{V}_j$ and $\mathcal{V}_{j+1}$. It can be seen that (2.16) extends to (2.17).

$$\mathcal{V}_1 = \mathcal{V}_0 \oplus \mathcal{W}_0 \tag{2.16}$$

$$L^2 = \mathcal{V}_0 \oplus \mathcal{W}_0 \oplus \mathcal{W}_1 \oplus \cdots \tag{2.17}$$

The relationship between the scaling function and wavelet vector spaces is illustrated in Figure 2.4.

The scale of the initial space $\mathcal{V}_j$ can be chosen arbitrarily, but is usually chosen to be the coarsest detail of interest in a signal. It can even be chosen as $j = -\infty$ where

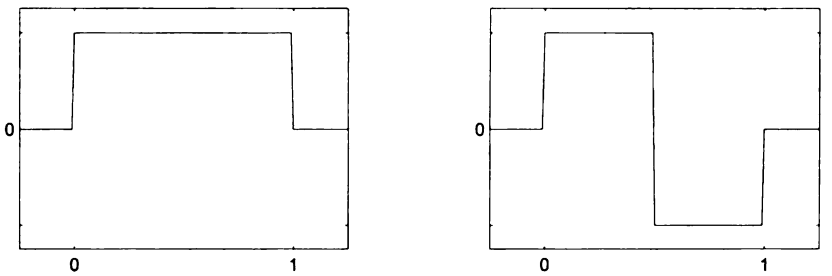

Figure 2.5: Haar Scaling and Wavelet Functions

L^2 can be reconstructed in terms of only wavelet functions (2.18).

$$L^2 = \cdots \oplus W_{-2} \oplus W_{-1} \oplus W_0 \oplus W_1 \oplus W_2 \oplus \cdots \tag{2.18}$$

In the case of the wavelet function, $\psi(t)$ is written as a function of $\varphi(2t)$ in (2.19),

$$\psi(t) = \sum_n h_1(n)\sqrt{2}\varphi(2t-n), \quad n \in \mathbf{Z} \tag{2.19}$$

where h_1 is a set of coefficients discussed in Section 2.3.1.3.

A very basic wavelet system with a scaling function and a wavelet function to make up the detail between one level of decomposition and the next is the Haar system shown in Figure 2.5.

Any function in $L^2(\mathbf{R})$ can be written as an expansion of a scaling function and wavelets (2.20), where $c_{j_0}(k)$ are the scaling function coefficients, $\varphi_{j_0,k}(t)$ is the scaling function at the initial scale j_0, $d_j(k)$ are the wavelet function coefficients and $\psi_{j,k}(t)$ are the wavelet functions spanning the space between V_{j_0} and L^2.

$$f(t) = \sum_{k=-\infty}^{\infty} c_{j_0}(k)\varphi_{j_0,k}(t) + \sum_{k=-\infty}^{\infty}\sum_{j=j_0}^{\infty} d_j(k)\psi_{j,k}(t) \tag{2.20}$$

2.3.1.3 Filter Banks

In order to perform the Discrete Wavelet Transform on a computer, computational methods must be developed. The DWT can be performed without using calculus, but rather additions and multiplications in the form of convolutions [49].

If the decomposition is considered linear (2.10), and if the basis functions are orthogonal (2.21),

$$\langle \psi_k(t), \psi_\ell(t) \rangle = \int \psi_k(t)\psi_\ell(t)dt = 0, \quad k \neq \ell \tag{2.21}$$

the coefficients of the decomposition, a_k, can be determined by calculating the inner product (2.22).

$$a_k = \langle f(t), \psi_k(t) \rangle = \int f(t)\psi_k(t)dt \tag{2.22}$$

In the two-dimensional case of the wavelet transform, the same techniques can be used to calculate the scaling coefficients (2.23) and the wavelet coefficients (2.24).

$$c_j(k) = \langle f(t), \varphi_{j,k}(t) \rangle = \int f(t)\varphi_{j,k}(t)dt \tag{2.23}$$

$$d_j(k) = \langle f(t), \psi_{j,k}(t) \rangle = \int f(t)\psi_{j,k}(t)dt \tag{2.24}$$

Finally the scaling function coefficients for a coarse scale can be determined from the scaling function coefficients at the next finer scale by convolving the coefficients at the finer scale with the recursion coefficients $h_0(n)$ and then down-sampling (2.25).

$$c_j(k) = \sum_m h_0(m - 2k)c_{j+1}(m) \tag{2.25}$$

The coefficients $h_0(n)$ are referred to as the decomposition lowpass filter coefficients. The same can be done in the case of the wavelet coefficients using the recursion

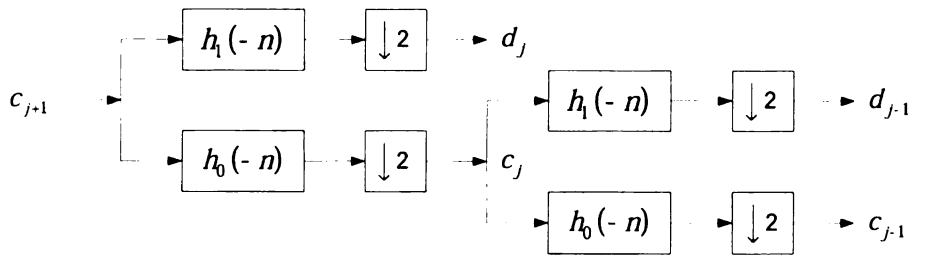

Figure 2.6: Two-Stage Filter Bank Analysis Tree

coefficients $h_1(n)$ (2.26) where $h_1(n) = (-1)^n h_0(1-n)$.

$$d_j(k) = \sum_m h_1(m-2k)c_{j+1}(m) \tag{2.26}$$

The coefficients $h_1(n)$ are referred to as the decomposition highpass filter coefficients. An example of a filter bank analysis tree is illustrated in Figure 2.6.

The down-sampling operation does not result in the loss of signal information. In the filter bank structure shown in Figure 2.6, there is enough information to reconstruct c_{j+1} in either the combination of c_j and d_j, or the combination of c_{j-1}, d_{j-1} and d_j. Despite down-sampling, either of these combinations of coefficients will have approximately the same number of values as c_{j+1}. Signal reconstruction from DWT coefficients is not used in this work, however it is discussed in detail in [49].

2.3.1.4 Undecimated Discrete Wavelet Transform

One of the drawbacks of the DWT is that it is not a shift-invariant transformation. This makes pattern recognition problems based on the DWT more difficult since the DWT coefficients resulting from decomposition of a signal and a shifted version of the signal can be very different. Only in special cases, where the signal is shifted by specific powers of two will the outputs be shifted versions of one another.

The UDWT adds the property of shift-invariance to the DWT. Here, the down-sampling step is omitted from the DWT algorithm, and zeros are inserted between

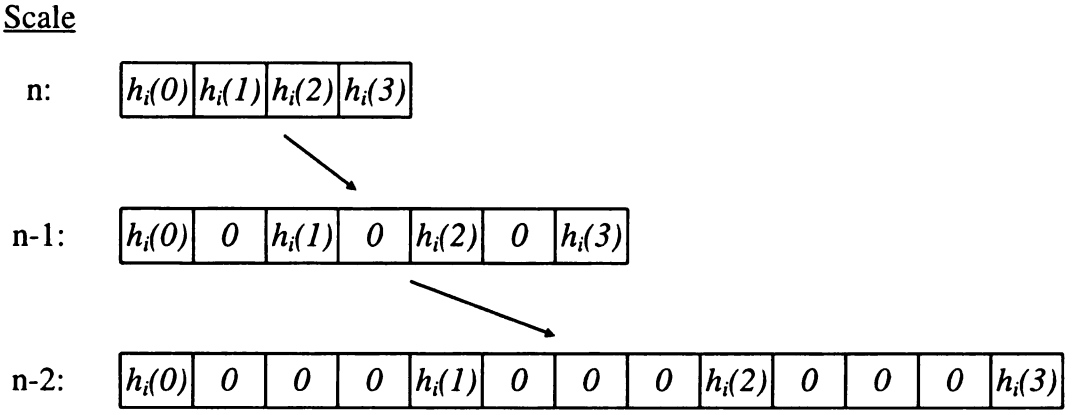

Figure 2.7: UDWT Filters Modified by the "Algorithme à Trous"

the filter coefficients at each successive scale, as shown in Fig. 2.7. This is known as the "Algorithme à Trous" [80, 81]. While circular convolution is used in [80], linear convolution is used here, since the signals are not periodic. The coefficients at each end of the convolution where the filter and the signal are not completely overlapping, known as end effects, are replaced by zeros. As these zeros propagate through each scale of the UDWT, the number of end effect coefficients increases. The fact that the length of the wavelet filters increase at each scale also contributes to additional end effects.

2.3.1.5 Wigner Ville Distribution

Wigner distribution(WVD)[79] of a signal $s(t)$ is defined as:

$$W(t,w) = \frac{1}{2\pi}\int s^*(t-\frac{\tau}{2})s(t+\frac{\tau}{2})e^{-j\tau\omega}d\tau \tag{2.27}$$

WVD gives the energy distribution of the signal as a function of time and frequency. It has high time-frequency resolution and does not suffer from the temporal vs. frequency resolution tradeoff encountered in STFT. The distribution preserves the energy of the underlying signal. Following are the basic properties of this distribution:

1. Reality The Wigner distribution is always real, even if the signal is complex.

2. Symmetry For symmetric spectra the Wigner distribution is symmetric in the frequency domain and for real spectra the time waveform i symmetrical and the Wigner distribution is symmetrical in time

$$W(t,\omega) = W(t,-\omega)\texttt{for real signals} \equiv \texttt{symmetrical spectra}, S(\omega) = S(-\omega) \tag{2.28}$$

$$W(t,\omega) = W(-t,\omega)\texttt{for real spectra} \equiv \texttt{symmetrical signals}, S(t) = S(-t) \tag{2.29}$$

3. Marginals The distribution satisfies the time frequency marginals

$$\int W(t,\omega)d\omega = |s(t)|^2 \tag{2.30}$$

$$\int W(t,\omega)dt = |S(\omega)|^2 \tag{2.31}$$

4. Time Shift If the signal is shifted in time, the distribution is shifted accordingly if $s(t) \rightarrow s(t-t_o)$ then $W(t,\omega) \rightarrow W(t-t_o,\omega)$

5. Frequency Shift If the signal is shifted in spectrum, the distribution is shifted accordingly if $s(t) \rightarrow e^{j\omega_o t}s(t)$ then $W(t,\omega) \rightarrow W(t,\omega-\omega_o)$

The major shortcoming of WVD occurs for multicomponent signals in terms of the cross-terms. When applied to multicomponent signals, it produces interference terms known as crossterms, which can hinder the correct identification of signal components. It is possible to attenuate the crossterms significantly using ambiguity domain kernel functions.

2.3.1.6 Choi Williams Distribution

Choi-Williams distribution(CWD)[79] of a signal *s(t)* is defined as:

$$C(t,f) = \int\int\int \phi(\theta,\tau)s(u-\frac{\tau}{2})s^*(u-\frac{\tau}{2})e^{j(\theta u-\theta t-\tau\omega)}dud\theta d\tau \tag{2.32}$$

where $\phi(\theta,\tau) = e(-\frac{(\theta,\tau)^2}{\sigma})$ is the kernel function that acts as a filter on the signals autocorrelation function. This distribution can be thought of as a filtered/smoothed version of the WVD and the amount of smoothing is controlled by σ. This smoothing removes the cross-terms seen in the WVD at the expense of reduced resolution.

1. Reality The CWD distribution real if the kernal satisfies the identical conditions $\phi(\theta,\tau) = \phi^*(-\theta,-\tau)$

2. Marginals The distribution satisfies the time and frequency marginals

$$\int C(t,\omega)d\omega = |s(t)|^2 \tag{2.33}$$

$$\int C(t,\omega)dt = |S(\omega)|^2 \tag{2.34}$$

   if $\phi(\theta,0) = 1$ and $\phi(0,\tau) = 1$ respectively. As the kernel function satisfies $\phi(0,0) = 1$, total energy is preserved.

3. Time Shift As the kernel is independent of time, the distribution is time shift invariant. If the signal is shifted in time, the distribution is shifted accordingly if $s(t) \rightarrow s(t-t_o)$ then $C(t,\omega) \rightarrow C(t-t_o,\omega)$

4. Frequency Shift As the kernel is independent of frequency, the distribution is frequency shift invariant. If the signal is shifted in spectrum, the distribution is shifted accordingly if $s(t) \rightarrow e^{j\omega_o t}s(t)$ then $C(t,\omega) \rightarrow C(t,\omega-\omega_o)$

The features from the motor current in time frequency domain were used as input for the classifier and prognosticator. The selection of the suitable distribution is based on the fisher discriminant ratio.

2.3.1.7 Fisher Discriminat Ratio

The Fisher discriminant ratio is a measure that quantifies the discrimination capacity of features regardless of the classifier. The classification results obtained by the different classifiers give information about which class each fault belongs to. However, these results are not necessarily informative of the separation of the different fault classes and how the time-frequency coefficients cluster with the different methods. The goodness of the features can be quantified through the classification accuracy or through Fisher discriminant ratio, which is independent of the classifier.

Fisher's discriminant ratio [51] can be used both as a classification method and as a class separability criterion. In this work, it is as an indicator of the class separability for the four different time-frequency analysis methods. For multi-class data, Fisher's discriminant ratio is defined as:

$$\begin{aligned} F(\mathbf{X}) &= \frac{S_B}{S_W} = \frac{\| \sum_{i=1}^{C} K_i(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T \|_2^2}{\sum_{i=1}^{C} s_i^2}, \\ \mathbf{m_i} &= \frac{1}{K_i} \sum_{j \in C_i} \mathbf{x}_j, \\ s_i^2 &= \frac{1}{K_i} \sum_{j \in C_i} \|\mathbf{x}_j - \mathbf{m}_i\|_2^2, \\ \mathbf{m} &= \frac{1}{K} \sum_{i=1}^{K} \mathbf{x}_i, \end{aligned} \tag{2.35}$$

where $\mathbf{X}$ is the matrix of the extracted features, K_i is the number of samples in each class, $\mathbf{m}_i$ is the mean of each class, s_i^2 is the variance within each class, and K is the total number of samples. Fisher discriminant ratio can be interpreted as the

ratio of the inter-class distance to inner-class scatter. Fisher's criterion is motivated by the intuitive idea that the discrimination power is maximized when the spatial distribution of different classes are as far away as possible from each other and the spatial distribution of samples from the same class are as close as possible to each other.

In the application proposed, this criterion will be applied to the time-frequency features extracted by the four methods to compare the discrimination power of the different transforms and $\mathbf{X}$ is the matrix of the time-frequency coefficients.

2.3.1.8 Energy Calculations

The time-marginal is used as a measure of energy for the coefficients of the STFT. This is defined as the sum of the squares of the coefficients at each time step.

In the case of the UDWT, the energy is computed by giving increasing weight to coefficients in higher scales. This gives more emphasis to the high frequency information in the signal, but does not discard low frequency information entirely.

$$E(t_2) = {d_{2,1}}^2 \cdot 2^2 + {d_{1,0}}^2 \cdot 2^1 + {d_{0,0}}^2 \cdot 2^0 \tag{2.36}$$

For the WVD and CWD, the energy is computed based only on the coefficients at the highest scale. For example, the energy at time instance t_2 is computed using (2.37).

$$E(t_2) = {d_{2,1}}^2 \tag{2.37}$$

2.3.2 Pattern Recognition Classifier

Once the features of the signal are available, i.e. the coefficients resulting from the STFT, UDWT, WVD and CWD, the type and the severity of the fault is determined by a classifier. The classifier considered for this work are the following:

2.3.2.1 Linear Discriminant Classifier

The linear discriminant classifiers (LDCs)[51] are trained on the input feature vectors of a set of known faults. The feature space is divided in C sub-regions, where C is the number of fault classes, corresponding to different fault severities. Weighting coefficients are computed for each class that maximize the linear-discriminant function for the corresponding input feature vectors. The linear-discriminant function is defined as:

$$D_c(x) = x_1\alpha_{1c} + x_2\alpha_{2c} + \ldots + x_k\alpha_{kc} + \alpha_{k+1,c}$$
$$c = 1, 2, \ldots, C \tag{2.38}$$

where x is the k-dimensional feature vector and α are the normalized weighting coefficients for the C-th class. A sample vector belongs to a particular class if its discriminant function for that class is greater than for the any other class, i.e., x belongs to class j if:

$$D_j(x) > D_k(x) \tag{2.39}$$

for every $k \neq j$. The weighting coefficients are adjusted from their initial guess through a training procedure. The algorithm for this procedure makes adjustments to the weighting coefficients until each of the training sample vector is correctly classified.

Young and Calvert [82] show that this training algorithm will converge in a finite number of steps. When a sample vector is correctly classified, no adjustment to the weighting coefficients is made. When a sample vector is incorrectly classified, or

$$D_j(\mathbf{x}) \leq D_l(\mathbf{x}),$$

where

$$D_l(\mathbf{x}) = \max_{l \neq j} \left[D_1(\mathbf{x}), \ldots, D_K(\mathbf{x}) \right],$$

adjustments are made to $\boldsymbol{\alpha}_j$ (2.40) and $\boldsymbol{\alpha}_l$ (2.41) only,

$$\boldsymbol{\alpha}_j(i+1) = \boldsymbol{\alpha}_j(i) + a\mathbf{x}_i \tag{2.40}$$

$$\boldsymbol{\alpha}_l(i+1) = \boldsymbol{\alpha}_l(i) - a\mathbf{x}_i, \tag{2.41}$$

where a is a gain constant.

Discriminant functions have minimal storage requirements after the training phase since, for each class, only a single vector of weighting coefficients need to be stored. Storage of training samples is no longer required during the classification phase. For multiclass problems ($K > 2$), it can be said that the classes are linearly separable if linear discriminant functions $D_1(\mathbf{x}), \ldots, D_K(\mathbf{x})$ exist, such that (2.42) is true.

$$D_j(\mathbf{x}) > D_k(\mathbf{x}) \quad \text{for every } \mathbf{x} \text{ in } C_j \text{ and all } k \neq j \tag{2.42}$$

2.3.2.2 k-means Classification

k-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed *a priori*. The main idea is to define k centroids, one for each cluster. These centroids can be chosen in different ways, such as choosing k random points or using the means from training data as centroids. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. In pattern recognition, k-means is a method for classifying objects based on closest training examples in the feature space. The distance is usually measured using the Euclidean distance (Euclidean

distance classifier EDC). When all of the data points have been assigned to clusters, this procedure is repeated by measuring the means of each cluster and using these as the centroids for clustering. This procedure is repeated until there are no changes in the k centroid.

The alternate method for the distance measurement is the Mahalanobias distance[51] (Mahalanabois distance classifier, MDC). This distance not only takes into account the means of the cluster but also adjusts itself for the clusters' variances. This the given as

$$D_m(x) = \sqrt{(x-\mu)^T \Sigma^{-1} (x-\mu)} \tag{2.43}$$

where μ and Σ are the means and variances of the clusters respectively.

2.3.2.3 Multiple Discriminant Analysis Classifier

Multiple discriminant analysis (MDA) classifier is a pattern recognition technique [51], which is used to analyze the fault signals. Analysis of the faults with different time-frequency transforms indicates that the time-frequency feature vectors contain redundant information and that not all regions of the time-frequency plane contain discriminative information [83]. Suitable data reduction methods are required to choose the 'optimal' lower dimensional transformation of the time-frequency feature vector.

MDA transforms from high dimensional feature space to a low dimensional space and maximizes the discrimination between different classes. It is an extension of Fisher Discriminant Ratio, which uses the ratio of intra-class scatter to inter-class scatter. Given the time-frequency feature vectors x and the class labels, the intra-class and inter-class scatter matrices are computed. Intra-class scatter matrix is

defined as:

$$\sum_w = \sum_{i=1}^{C} \sum_{x \in C_i} (x - m_i)(x - m_i)^T \tag{2.44}$$

and the inter-class scatter matrix is defined as

$$\sum_b = \sum_{i=1}^{C} K_i (m_i - M)(m_i - M)^T \tag{2.45}$$

where m_i is the class average, M is the total average, and K_i is the number of samples in class i. The optimal linear transformation Φ is defined such that following is maximized:

$$J(\Phi) = \frac{\Phi^T \sum_b \Phi}{\Phi^T \sum_w \Phi} \tag{2.46}$$

The optimal Φ is found by solving the generalized eigenvalue problem, $\sum_b \Phi = \lambda \sum_w \Phi$. For C classes, the linear transform projects to a $C - 1$ dimensional space moreover, lower dimensional space is the one which give the maximum separation between different classes.

The MDA classifier calculates the distance of each new incidence from the centroid of all the clusters (generated during the training phase) and assigns it to the class for the distance is the least.

2.3.2.4 Support Vector Machine Classifier

An support vector machine (SVM) aims to fit an optimal separating hyperplane (OSH) between classes by focusing on the training samples that lie at the edge of the class distributions, the support vectors. The OSH is oriented such that it is placed at the maximum distance between the sets of support vectors. It is because of this orientation that SVM is expected to generalize more accurately on unseen cases relative to classifiers that aim to minimize the training error such as neural

networks. Thus, with SVM classification only some of the training samples that lie at the edge of the class distributions in feature space (support vectors) are needed in the establishment of the decision surface.

The training data is projected into arbitrary higher dimensional feature space, using non-linear mapping. This mapping is implicitly defined by the algorithm. The selection of OSH is carried out in higher dimensional space.

$$x = (x_1, x_2, ..., x_n) \mapsto \phi(x) = (\phi_1(x), \phi_2(x), ..., \phi_n(x)) \tag{2.47}$$

The simplest form of SVM classifier is maximal margin classifier and it will explained. It is convex optimization problem, which tries to minimize a quadratic function under linear inequality constraints. Suppose we have linearly separable training data as shown in 2.8 and we to find a optimal hyperplane for the separation of it.

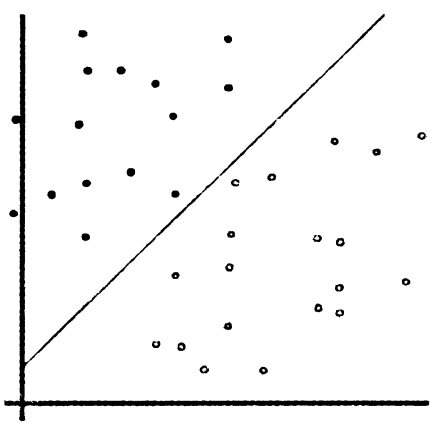

Figure 2.8: Linearly Separable Data

Suppose the set of labeled training data is

$$(y_1, x_1), ..., (y_l, x_l)\ y_i \in \{-1, 1\} \tag{2.48}$$

This data set is linearly separable if there exist a vector w and a scalar b such that

the inequalities

$$wx_i + b \leq 1 \;\; if \;\; y_i = 1 \tag{2.49}$$

$$wx_i + b \geq -1 \;\; if \;\; y_i = -1 \tag{2.50}$$

are valid for all the elements of the training set. Eq. 2.49 and 2.50 can be combined in the following form:

$$y_i(wx_i + b) \geq 1 \quad i = 1, ..., l \tag{2.51}$$

The optimal hyperplane

$$w_0 x + b = 0 \tag{2.52}$$

is the unique one which separates the training data with maximal margin. It determines the direction where the distance between the projections of training vectors of two different classes is maximal. *Support Vectors* are those data point against which this margin pushes up.

2.3.2.4.1 Optimal Hyperplane Selection Selection the optimal hyperplane is a constrained optimization problem given by

$$\min_{w,b} \frac{1}{2} w^t w \;\; subject\ to \;\; y_i(w^t \phi(x_i) + b) \geq 1 \;\; i = 1, ..., l \tag{2.53}$$

The optimization problem mentioned in Eq. 2.53 is for the case where the training data is completely separable. It is a very strong assumption and may not be true in most of the practical and generalized applications[84]. For non separable training data, a non-negative variable ζ is introduced in the equation 2.53 and it is expressed as

$$\min_{w,b,\zeta} \frac{1}{2} w^t w + c \sum_{i=1}^{n} \zeta_i \;\; subject\ to \;\; y_i(w^t \phi(x_i) + b) \geq 1 - \zeta_i \;\; i = 1, ..., l \tag{2.54}$$

2.3.3 Prognosticators

There are many popular techniques for prognosis, some of which are reviewed in this section.

2.3.3.1 Kalman Filter

It is one of the most popular and common methods[85], which works with linear systems in presence of Gaussian noise. It provides an efficient recursive computational solution by minimizing the least-squared error. It can estimate the past, present and future states of the system under observation.

Kalman filter solves the general problem of estimating the state of a discrete-time process that is governed by the linear stochastic difference equation given as Eq 2.3.3.1

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \tag{2.55}$$

and 2.3.3.1 as output equation

$$z_k = Hx_k + v_k \tag{2.56}$$

where x_k is the state vector at time k, z_k is the output, A, B and H are the systems matrices, v_k and w_k are the process and measurement noise respectively assumed to be Gaussian white noise with zero mean and Q and R are process noise covariance and output noise covariance matrices respectively.

$$p(w) \sim N(0, Q) \tag{2.57}$$

$$p(v) \sim N(0, R) \tag{2.58}$$

If $\hat{x}_k^- \in \Re n$ is a priori state estimate at step k, given knowledge of the process prior to step k, and $\hat{x}_k \in \Re n$ is a posteriori state estimate at step k given measurement z_k. A priori and a posteriori estimate errors are defined as

$$\bar{e}_k = x_k - \hat{x}_k^- \tag{2.59}$$

and

$$e_k = x_k - \hat{x}_k \tag{2.60}$$

The a priori estimate error covariance is then

$$P_k^- = E[e_k^- e_k^{-T}] \tag{2.61}$$

and the a posteriori estimate error covariance is

$$P_k = E[\bar{e}_k e_k^T] \tag{2.62}$$

There are two stages of Kalman filter. First is the prediction, in which the states are projected ahead in time. Suppose we have a initial guess about the state $\hat{x}_{k-1}$ and error covariance P_{k-1} at time k-1. The prediction stage will project the state and covariance from time step k-1 to time step k. With zero input the system predict states will be

$$\begin{aligned} \hat{x}_k^- &= A\hat{x}_{k-1} \\ P_k^- &= AP_{k-1}A^T + Q \end{aligned} \tag{2.63}$$

The next stage is update or correct. In this face the Kalman gain is computed.

$$K_k = P_k^- H^T (H P_k^- H_T + R)^{-1} \tag{2.64}$$

The Kalman gain is used to update the state estimate and error covariance estimate.

$$\begin{aligned} \hat{x}_k &= \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-) \\ P_k &= (I - K_k H) + P_k^- \end{aligned} \tag{2.65}$$

The measurement noise covariance R is measured prior to the operation of the filter, which is generally possible, however, the determination of the process noise covariance, Q, is more difficult therefore reasonable values are selected.

After each predict and update, the process is repeated with the previous a posteriori estimates used to project or predict the new a priori estimates and filter eventually converges to the correct state of the system. The recursive nature is one of the very appealing features of the Kalman filter.

Due to the stringent conditions of linearity, the extended Kalman filter is the more suitable alternate for real world systems. The extended Kalman filter linearizes the systems around the point of interest and solution is obtained recursively as in case of Kalman filter.

2.3.3.2 Particle Filters

Particle filter is a Bayesian method in which the solution of filtering problem is computed by recursively estimation[59]. The filtering problem is to estimate the first two moments of the state vector which is governed by the dynamic state space model having noisy observation. A discrete time controlled process can be expressed in state space form by the stochastic difference equation of the form

$$x_k = f_k(x_{k-1}, w_{k-1}) \tag{2.66}$$

and a measurement equation $y \in \Re^k$ given by

$$y_k = h_k(x_k, v_k) \tag{2.67}$$

The equation Eq.(2.66) is called the state transition(dynamic) equation whereas the Eq.(2.67) is called the correction, update or output equation. At time t_k, x_k is the state vector, w_k is the dynamic noise, y_k is the real observation vector and v_k is the observation noise vector. The function f_k gives the relationship between the previous state and the current state and the function h_k links the current state to the output[57].

In Bayesian form, instead of the future state vector, the probability density (pdf) of the future state vector is estimated. Following the pattern of update and measurement equations, the prior pdf is calculated using the update equation and the posterior pdf using the measurement equation. Eq.(2.66) gives the predictive conditional transition density, $p(x_k|x_{k-1}, y_{k-1})$, of the current state given the *previous* states and *previous* observations [86]. The observation or measurement equation Eq.(2.67) gives the likelihood function of the current measurement given the current state, $p(y_k|x_k)$. If $p(x_{k-1}|y_{k-1})$ is defined as the previous posterior density then the prior pdf $p(x_k|y_{1:k-1})$ is defined using the Baye's rule as:

$$p(x_k|y_{1:k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|y_{k-1})dx_{k-1} \tag{2.68}$$

The correction step generates the posterior probability density function from

$$p(x_k|y_{1:k}) = c * p(y_k|x_k)p(x_k|y_{1:k-1}) \tag{2.69}$$

The filtering problem is to recursively estimate the first two moments of x_k given y_k. For a general distribution, p_x, this consists of recursive estimation of the expected value of any function of x, say $\langle g(x) \rangle_{p(x)}$, using Eqns. 2.68 and 2.69. This requires the calculation of the integral of the form[87].

$$\langle g(x) \rangle_{p(x)} = \int g(x)p(x)dx \tag{2.70}$$

2.3.3.2.1 Monte Carlo Integration Monte Carlo (MC) methods are stochastic techniques, meaning they are based on the use of random numbers and probability statistics to investigate problems[88]. In this technique, the solution of a numerical problem is estimated by means of an artificial sampling experiment. The estimate is usually given as the average value, in a sample, of some statistic whose mathematical expectation is equal to x. The main idea behind this method is centered around the availability of very large computational power of present day processors. MC estimate advocates that when a large number of samples are drawn from the required posterior distribution, it is possible to approximation the intractable integral.Suppose we want to numerically evaluate a multidimensional integral:

$$I = \int g(x)dx \tag{2.71}$$

where x$\in \Re_x^n$. In MC methods, numerical integration is factorized as $g(x) = f(x)\pi(x)$ in such a way that $\pi(x)$ is a probability density satisfying $\pi(x) \geq 0$ and $\int \pi(x)dx = 1$. The assumption is that it is possible to draw $N \gg 1$ samples $\{x^i; i = 1, ..., N\}$ distributed according to $\pi(x)$. MC estimation of the integral

$$I = \int f(x)\pi(x)dx \tag{2.72}$$

is the mean sample

$$I_N = \frac{1}{N}\sum_{i=1}^{N} f(x^i) \tag{2.73}$$

If the samples x^i are independent then I_N is an unbiased estimate and according to law of large numbers I_N will almost surely converge to I. However, $\pi(x)$ a posterior density in Bayesian estimation, is not possible to sample efficiently, the reason being its multivariate, nonstandard nature and limited possible assess. The possible solution is use of "Importance Sampling Method".

2.3.3.2.2 Importance Sampling Ideally we want to generate samples directly from $\pi(x)$ and estimate I using Eq.(2.72). But suppose we can only generate samples from a density *q(x)*, which is similar to $\pi(x)$. Then we need to introduce correcting weights for the sampled set in order to make MC estimation possible. The pdf *q(x)* is referred to as the *importance* density provided that the following condition is fulfilled:

$$\pi(x) > 0 \Rightarrow q(x) > 0 \tag{2.74}$$

for all $x \in \Re_x^n$, which means that both π and $q(x)$ have same support. It is a necessary condition for importance sampling to hold. If this condition is met, any integral of the form (2.72) can be written as:

$$I = \int f(x)\pi(x)dx = \int f(x)\frac{\pi(x)}{q(x)}q(x)dx \tag{2.75}$$

$q(x) > 0$. A MC estimate of I is computed by generating $N \gg 1$ independent samples $\{^i, i = 1...N\}$, distributed according to *q(x)* and forming the weighted sum:

$$I_N = \frac{1}{N}\sum_{i=1}^{N} f(x^i)\tilde{\omega}(x^i) \tag{2.76}$$

where

$$\tilde{\omega}(x^i) = \frac{\pi(x^i)}{q(x^i)} \tag{2.77}$$

are the importance weights. If the normalizing factor of the desired density is unknown we need to perform the normalization of the importance weights. Then the estimate I_N is as follow:

$$I_N = \frac{\frac{1}{N}\sum_{i=1}^{N} f(x^i)\tilde{\omega}(x^i)}{\frac{1}{N}\sum_{j=1}^{N} \tilde{\omega}(x^i)} \tag{2.78}$$

Particle filter estimation is a type of MC integration, which is also called sequential importance sampling. In this method recursive Bayesian estimation is performed by implementing MC simulations. The required posterior density function 2.69 is represented by a set of random samples with associated weights. As the number of samples becomes very large, the MC representation converges to the actual functional description of the posterior pdf.

2.3.3.3 Hidden Markov Model

A HMM is a stochastic technique for modeling signals that evolve through a finite number of states[68, 73]. The states are assumed hidden and responsible for producing observations. An HMM assumes that the system is Markovian, this means the behavior depends only on the current state. The objective is to characterize the states given the observations. S_k is the hidden state at time k and O_k as the observation sequence, assuming that there are C possible states. The main objective is to determine hidden parameters (states) from the observable parameters. There are three basic problems to solve by HMM.

1. *Problem 1:* Given the observation sequence $y = \{y_1 y_2 ... y_k\}$ and set of model parameters $\theta = \{\pi, A, B\}$ how to efficiently compute *p(y)*, that is the probability of the observation sequence, given the model.

2. *Problem 2:* Given the observation sequence $y = \{y_1 y_2 ... y_k\}$ and set of model parameters $\theta = \{\pi, A, B\}$ how to choose corresponding state sequence $x = \{x_1, x_2 ..., x_k\}$, which is optimal to generate the observation sequence. The optimal measure can be the maximum likelihood.

3. *Problem 3:* How do we adjust the parameters $\theta = \{\pi, A, B\}$ to maximize the likelihood of all observation sequences.

In this work, HMM is used to get the solution of the Problem 2, which tries to maximize the likelihood of the states given the observable variable. An algorithm is developed which uses HMM for the prognosis. The model has three elements:

- $\boldsymbol{\pi}$: C x 1 initial state distribution vector where the i_{th} element is the probability of being in state i at time $k = 0$. $p(S_0 = i)$.

- **A**: C x C state-transition matrix where the $(i, j)^{th}$ element is the probability of being in state j at time $k + 1$, given that it is in state i at time k, $p(S_{k+1} = j | S_k = i)$.

- **B**: State-dependent observation density B. Its j^{th} element is the probability of observing O_k at time k given the system is in state j, $b_j(O_k) = p(O_k | S_k = j)$.

The model parameters are collectively denoted by $\lambda = \{\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B}\}$. In order to implement the HMM based prognosis algorithm, the model parameters need to be trained. The state transition probabilities($\boldsymbol{A}$) and state dependent observation densities ($\boldsymbol{B}$) are generally obtained from the historical data collected from large number of observations and the initial state probability distributions ($\boldsymbol{\pi}$) depend on the implementation area and the nature of operation of the system being studied. In this work, computational methods for the HMM elements ($\lambda = \{\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B}\}$) and algorithm for prognosis based on these elements are developed and presented in Chapter 6.

Chapter 3

Problem Formulation and Selected Approach

3.1 Scope and Objective of the Chapter

The first objective of the present chapter is to formulate the problem addressed in the thesis. It describes the available inputs, major constraints and limitations in the case of complex electromechanical systems. Moreover, the required intermediate data and information needed to compute the solution are also identified. The second objective of the chapter is to postulate the approach to address the problem. Based on the selected approach, an overview of the different phases of the algorithm development are also presented.

3.2 Problem

In this thesis, fault analysis of the complex electromechanical systems is addressed. Generally, electromechanical systems are non-linear, noisy and complex. Early fault detection, categorization and prognosis require continuous monitoring of the systems.

Mostly, the deployment, role and location of the electromechanical systems confine the analysis methods only to the ones which are non-intrusive in nature. As the faults in the system manifest themselves in the machine current[89], motor current signature analysis is an attractive non-intrusive fault analysis tool. By examining the features extracted from the machine current, faults can be analyzed. This method has been successfully implemented for the diagnosis of major machine faults, for example turn-to-turn short circuit, cracked /broken rotor bars, bearing deterioration etc. [89, 90, 91].

The objective of the thesis is to develop algorithms of fault diagnosis and failure prognosis of the transient faults in complex electromechanical systems, using non-intrusive methods. For the diagnosis, it is assumed that there is no apriori information available. For prognosis, the limitations are the non linearity of the system, presence of non Gaussian noise, non availability of the historic data from which fault progression trends can be extracted, and non availability of system model. The objective is to develop a methodology while remaining within these constraints.

3.3 The Selected Approach

Diagnosis and prognosis techniques are generally divided in three groups, model based, data based and signal based. The modeling of electromechanical systems is not practical due to their complex construction and the requirement of extensive approximations, which makes model based analysis methods an inappropriate choice. For the data based methods, huge data is required to define system trends and fault patterns, which is mostly not available in electromechanical systems. The signal based techniques are considered suitable options for this analysis and they are applied on the lines of the supervised learning methods. In the supervised learning, the data is collected from the system in known health conditions and based on the decision rules

developed, health conditions of unknown systems are categorized and prognosticated. The adopted approach has three major parts, analysis of the feature extraction methods, classifiers and the prognosticator. Algorithms are to be developed for feature extraction, diagnosis and prognosis. The overview of the methodology adopted is shown as Fig. 3.1

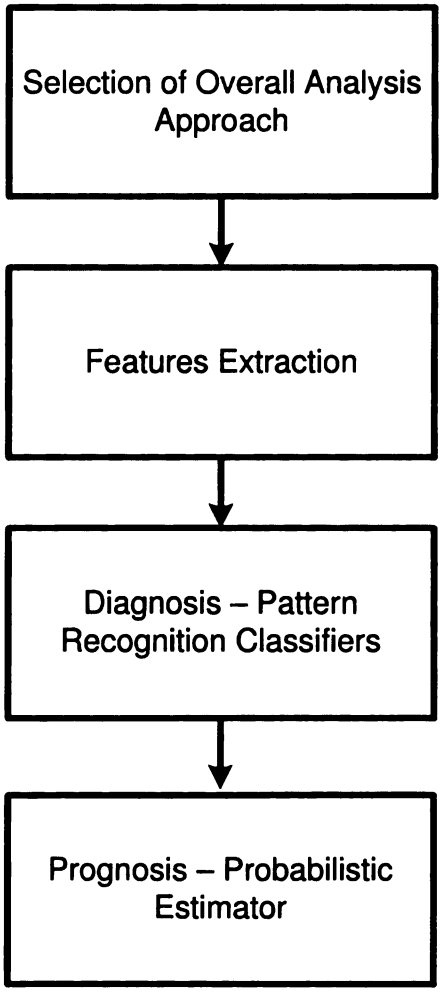

Figure 3.1: Methodology

3.3.1 Analysis of the Feature Extraction Methods

The fault features can be extracted in time, frequency and time-frequency domains. The time domain features can be the peak values or covariance comparisons. They offer lesser computational cost. However, they are not very efficient in the case of transient signal analysis. The frequency domain features or the spectrum analysis is extensively used for the fault diagnosis, however, in case of transient signals the frequency domain features are not very discriminative. In the time-frequency analysis the signal is represented in three dimensions, time, frequency, and amplitude. It is inherently suited to indicate transients in the signal. Different transforms have been proposed for efficient representation of transient events. In this work, the fault features are extracted in the time-frequency domain using four different transforms, short time Fourier transform, undecimated wavelet transform, Wigner transform and Choi-Williams transform. The Fisher discrimination ratio is used as the figure of merit for the selection of most suitable transform for the transient repetitive faults.

3.3.2 Analysis of the Diagnostic Methods

The features extracted using time-frequency distribution are the input of the classifiers. For the fault classification different candidate methods are analyzed, which are linear discriminant classifier, nearest neighborhood classifier, support vector machine classifier and Mahalanobias distance classifier . The comparative analysis of the classifiers is performed in terms of classification accuracy and the computational time.

3.3.3 Analysis of the Prognosticators

Prognosis succeeding the activity of diagnosis and it gives the estimate of the RUL of the equipment. Different prognosis methods have been presented in the literature,

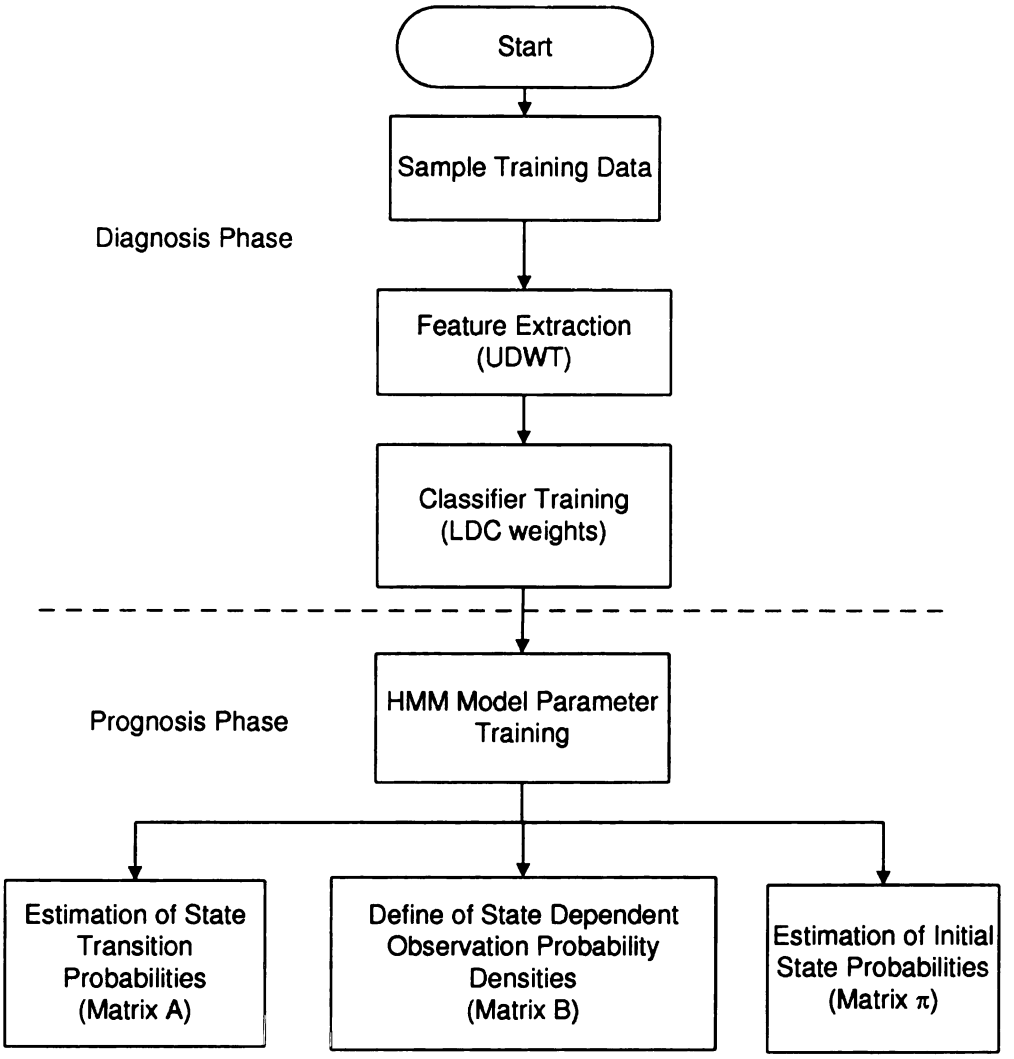

Figure 3.2: Training block diagram

namely Kalman filter, fuzzy logic predictor and particle filters. These prediction methods are model based techniques. In this problem, these methods cannot be employed as the model of the system is not available. A statistical technique HMM-based prognosis method is presented for the estimation of RUL.

HMM has three parameters, (matrices A,B *and* π) as mentioned in Chapter 2. These parameters needs to be trained before engaging the model for the prediction. The training of parameters requires a large amount of data, which is not always

available. In the presence of limited data, this task becomes more challenging. In this work methods are developed for the computation of these parameters from the experimental data. For the computation of state transition probabilities(matrix A) a heuristic method is developed, which is based on the matching pursuit decomposition. The state dependent observation densities are defined as parametric densities. The outcome of the training of the classifier and the experimental observations are used to obtain the parameters of the densities. A prognosis algorithm is developed, which computes the RUL in terms of the probability of the failure state, given the trained model.

3.4 Algorithm Execution Phases

Supervised learning is the overall approach for the development of the algorithm. In supervised learning, the execution is performed in two phases. The first phase is called the training phase. During the training phase, it is assumed that all the required information about the system is available. The decision rules are developed, provided the known set of information. Following is the testing phase, during which the decision rules are applied to the data sampled from the systems of unknown health conditions. Both diagnosis and prognosis algorithms need to be trained, and both use the same set of time frequency features extracted from the machine current. The prognosis phase utilizes the information acquired from the diagnosis. Figure 3.2 and Figure 3.3 show the block diagrams of the training and testing phases respectively.

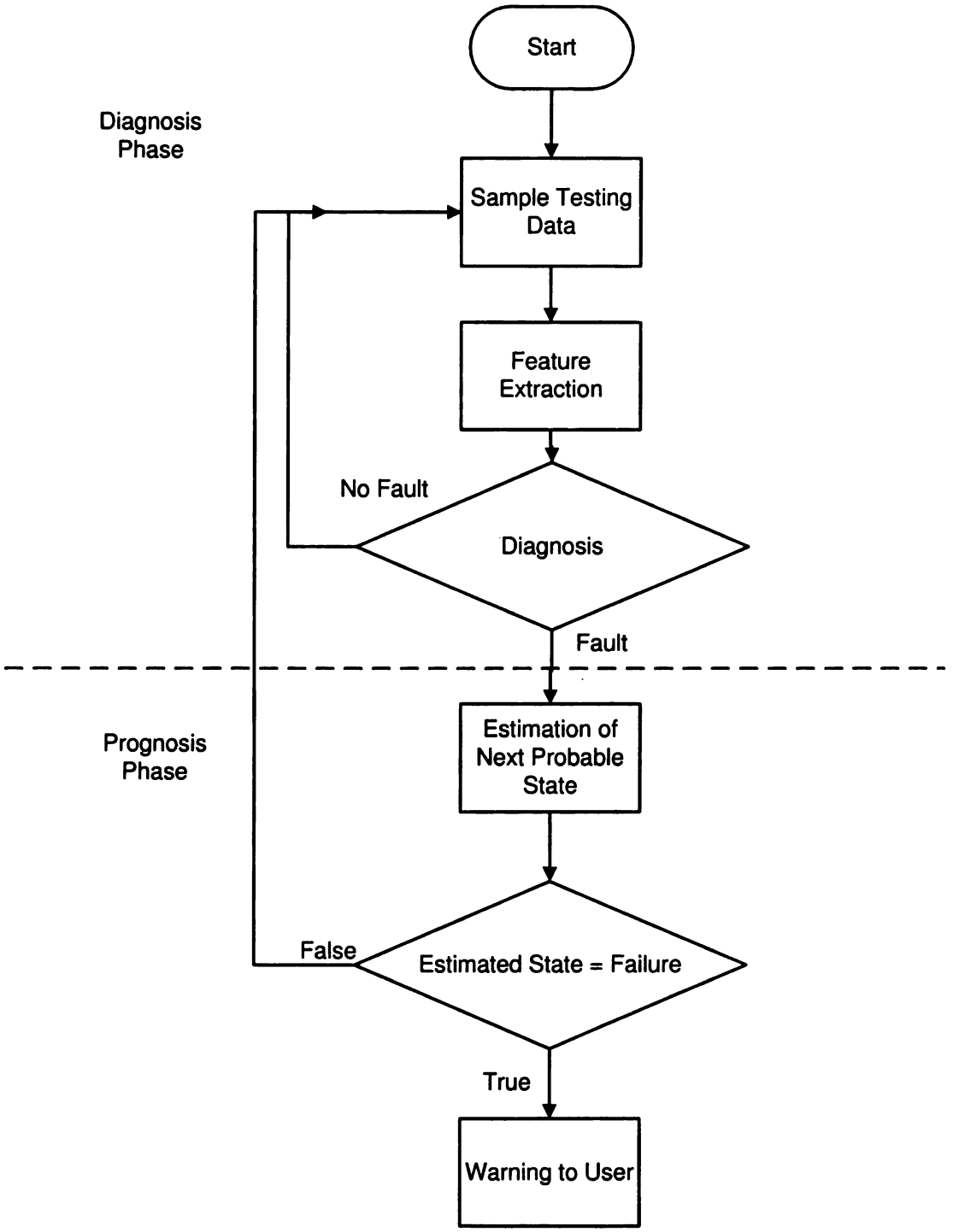

Figure 3.3: Testing block diagram

Chapter 4

Experimental Setup

4.1 Scope and Objective of the Chapter

The first objective of this chapter is to discuss possible faults in electromechanical systems and their manifestation in the signals which can be used for fault analysis. The second objective of the chapter is to present the details of the experiment built. The developed algorithms for fault diagnosis and failure prognosis are generic in nature and can be used for a wide range of applications. However, their working is demonstrated by applying them to the analysis of the health condition of starting systems used in automobiles. This chapter describes the operation of the starting systems, the laboratory experiments, fixtures, control and data acquisition methods, sampled signals, sensors and signal enhancement methods.

4.2 Faults in Electromechanical Systems

A variety of faults can occur in electromechanical systems. These faults can be grouped as electrical and mechanical faults. Generally, the faults which occur in the current path are termed as electrical and ones which are related with mechanical

components are regarded as mechanical faults. Following are the possible faults in electromechanical systems:

1. Electrical faults
    (a) Short circuit
    (b) Open circuit
    (c) Deterioration of brushes
    (d) Brush spring faults
    (e) Deterioration of commutator
2. Mechanical faults
    (a) Damaged gears
    (b) Eccentricity
    (c) Bearing faults

These faults can manifest themselves as mechanical vibrations, acoustic noise, and/or current transients [48, 92, 93, 94, 95] and the fault features can be extracted from these signals. The gear fault in automobile starting systems is selected as a test fault for the demonstration of the developed algorithm. It is a repetitive transient fault which is common in many electromechanical systems. The starting system uses DC machines energized by battery. Although DC machines have been largely replaced in industrial applications by permanent magnet and induction machines, they are still extensively used in auxiliary automotive systems. Besides starting systems, these machines are popular for the wipers, door locks, view mirrors and seat adjustment systems. Despite not being part of the traction system, their malfunction can result in overall failure of the vehicle.

4.3 Operation of Starting Systems

The starter motor cranks the engine when a starting signal is applied by the user through the twist of the key. During cranking, the starter motor turns the flywheel; which translates the rotary motion to linear motion of the pistons through the crank shaft. In one crankshaft revolution, the engine undergoes three cycles of compression and expansion in the cylinders, which causes variable load on the starter motor. The starter motor current is sampled and is shown in Figures 4.1 and 4.2.

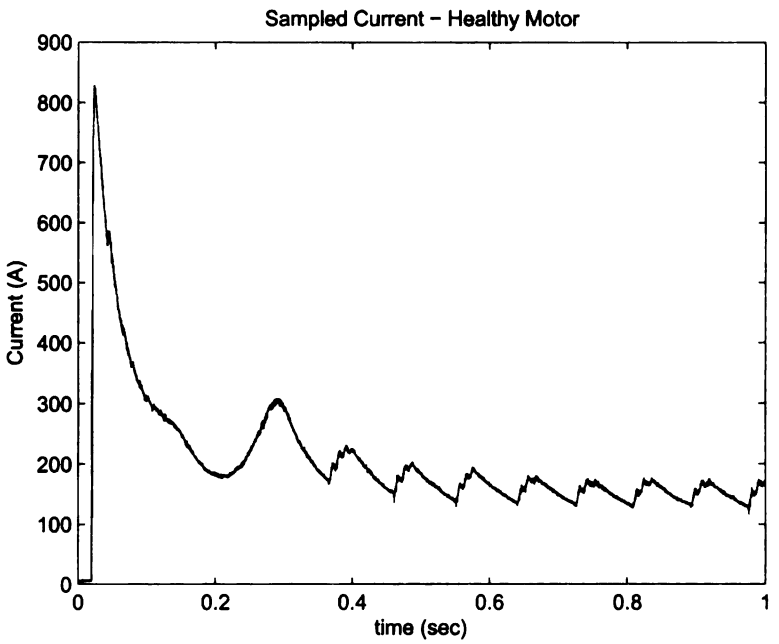

Figure 4.1: Sampled Current

During compression, the load on the flywheel, and in turn on the starter motor increases and a rise in current appears (4.2-region A). This is the time when the starter motor is exerting torque and turning the flywheel. In contrast, during the expansion, the load reduces, and the flywheel is turning mainly due to inertia and does not offer significant load to the starter motor. The motor current decreases causing an increase of starter motor speed (4.2-region B). As the starter gear applies force, a

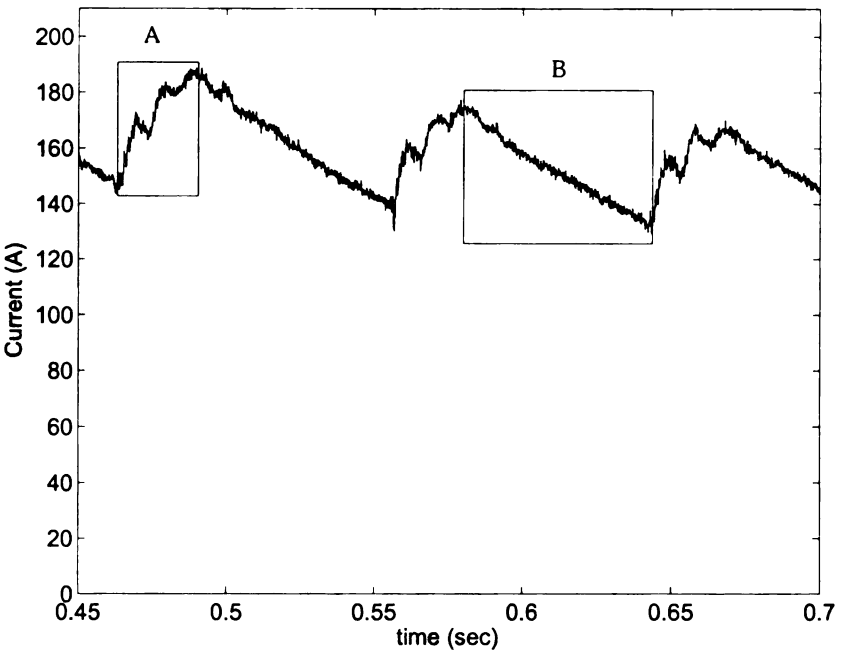

Figure 4.2: Healthy Motor Current - Zoomed

moment comes when contact between the starter and flywheel gears is lost, followed by a knock when contact is re-established. This mechanical transient is translated into speed and torque transients, and subsequently into a short transient of the stator current. Any damage in the gear tooth is reflected in the knock. Figures 4.3~4.7 show schematics of healthy gears and gears with one tooth damaged with different severities. The gear shown in Figure 4.3 is without damaged, called healthy gear. The other four figures show gears with increasing damage severities, which are illustrated by the dotted lines. The gear ratio between the starter motor and flywheel is 15.8:1 and it completes approximately five revolutions during one compression/expansion cycle.

The starter motor provides the initial torque during the starting attempts. The load of the machine is an engine, with widely varying load torque. The load torque depends on the motor conditions, external environment and multiple resonances. The

load on the starting system is very complex both in terms of electrical and mechanical. Therefore its electromechanical model is inadequate. However, despite the nonlinear and variable load, the stator current remains as a good health indicator both in terms of accuracy and in terms of cost and can provide useful insight about the motor operating conditions. Therefore, the signal based techniques are a suitable option for diagnosis and prognosis, which were discussed in Chapter 2.

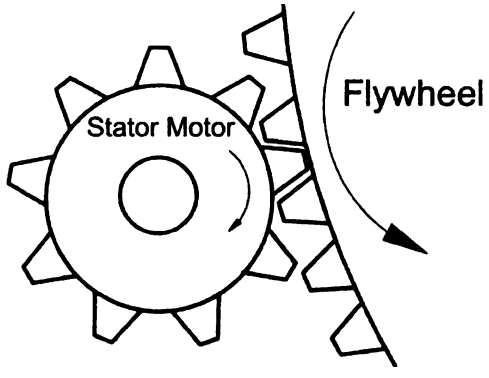

Figure 4.3: Gear with no faults

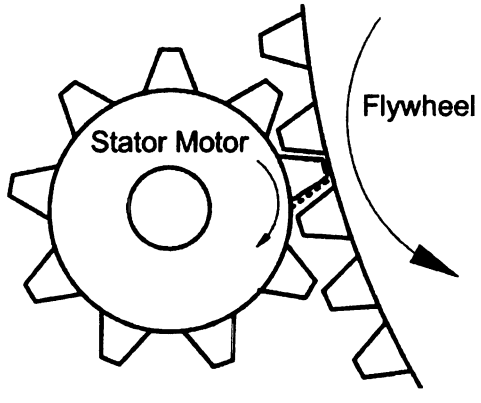

Figure 4.4: Gear fault severity - 1

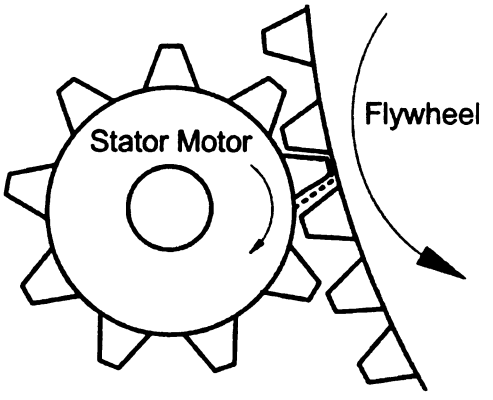

Figure 4.5: Gear fault severity - 2

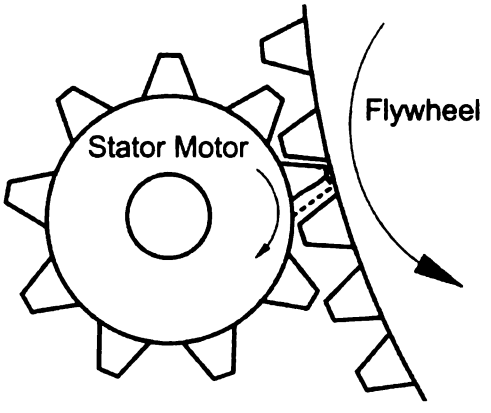

Figure 4.6: Gear fault severity - 3

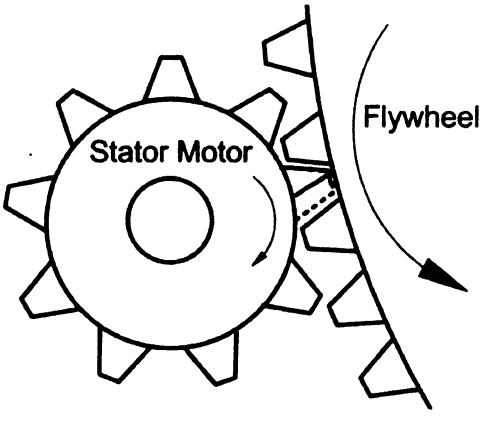

Figure 4.7: Gear fault severity - 4

4.4 Engine and Starter

In automobiles, the starter motor assembly is energized by a 12V battery. It is operated through a relay controlled by the ignition key. The relay and the ignition switch are also energized by the same 12V battery. The starter motor turns a flywheel which cranks the engine. The starter motor is attached to the engine body. In the laboratory, the experimental setup was constructed using a complete engine module, housing the motor assembly. It served two purposes, first to create a realistic operational environment and second to operate the motor under actual load conditions. Commercially available engine cradle with certain modifications was used to hold and stow the engine during the experiments. The engine was mounted on the cradle using rubber pad mounts to dampen the vibrations as much as possible. Figure 4.8 shows the engine placed on the cradle in the laboratory.

4.5 Control of Starter Motor and Data Acquisition

To control the motor and sample the data, a PC running National Instrument Labview software was used. In automobiles, the twist of key sends the energizing signal to the starter relay. In the experiment, this signal is generated by Labview upon the users input. The starting signal is sent from the USB port to the data acquisition card, National Instrument DAQ 6229. It is a high-performance M Series multifunction data acquisition (DAQ) module, capable of sampling at a rate of $250KS/s$ and with 16-bit resolution. The same data acquisition card is used for the sampling of data from the motor. The front end for the motor control was designed using NI Labview. A block diagram of the experimental setup is shown in Figure 4.9.

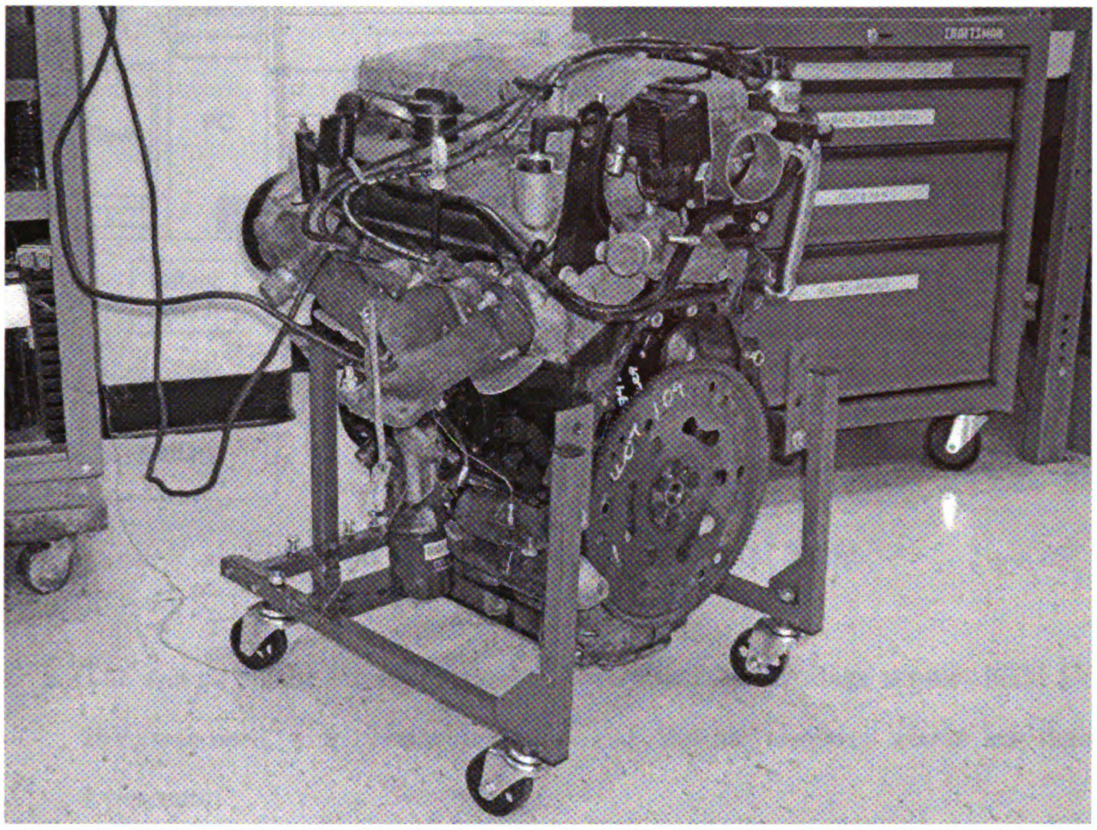

Figure 4.8: Hardware Setup

4.6 Sampled Signals

From the starter motor three signals, motor current, battery voltage and engine vibration, were sampled.

1. The motor current was the primary signal for the analysis. In typical starter cycle the initial motor current reaches $750 - 800A$, moreover, the frequencies of the electrical faults are expected to be in the range of kHz. Therefore a sensor having high current tolerance, wide bandwidth, and adequate accuracy was required. LEM Current Transducer HASS 600S was used, which can withstand current up to $\pm 3500A$, has a linear range between $-900A$ and $900A$ and a bandwidth of $50kHz$, with accuracy within 1 percent of full scale.

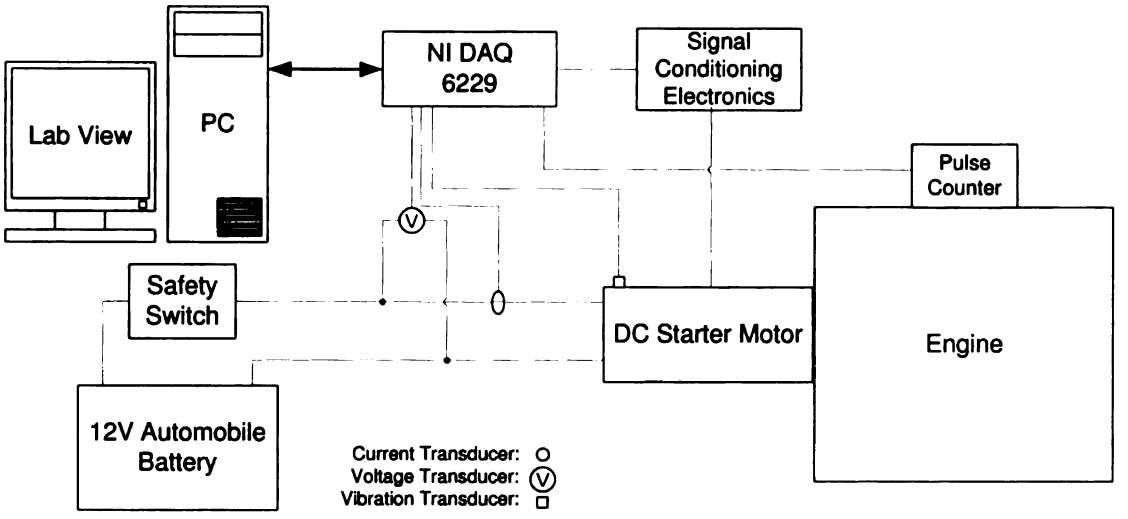

Figure 4.9: Block Diagram Experimental Setup

2. The second signal sampled was the battery voltage. Voltage sensor, LEM LV 25V, was used, which has wide band width and has accuracy within less than 1 percent.

3. The third signal measured was the vibration of the engine. For the measurement of the vibration, series 303A Quartz accelerometer was used. It functions to transfer shock and vibratory motion into high-level, low-impedance (100Ω) voltage signals compatible with readout, recording, or analyzing instruments. It is a small, sensitive ($10mV/g$) sensor, which operates reliably over wide amplitude and frequency ranges under adverse environmental conditions.

4.7 Gear Position Sensor

Information about the precise position of the starter motor pinion and the flywheel was required in order to establish relationship between the fault occurrence instance and the point in the engine cycle. Speed and position sensor4.10 was made, using an optical pulse counter. It gave one pulse for each tooth crossing of the flywheel

and generated a total of 142 pulses per flywheel revolutions. Given the initial position of the pinion, relative positions of the flywheel and starter motor pinion with respect to compression/expansion cycle can be obtained from the output of position sensor. Therefore, the position of pinion was physically noted before collecting each sampled signal. In Figure 4.11 the sampled current with pulses of the optical sensor is exhibited.

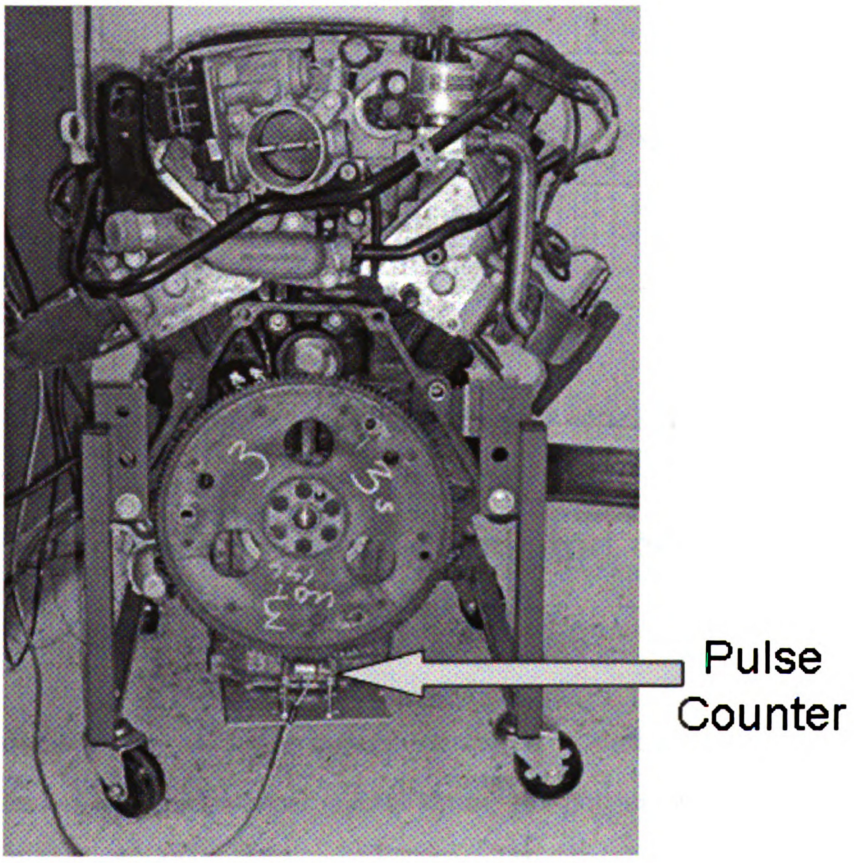

Figure 4.10: Position Sensor

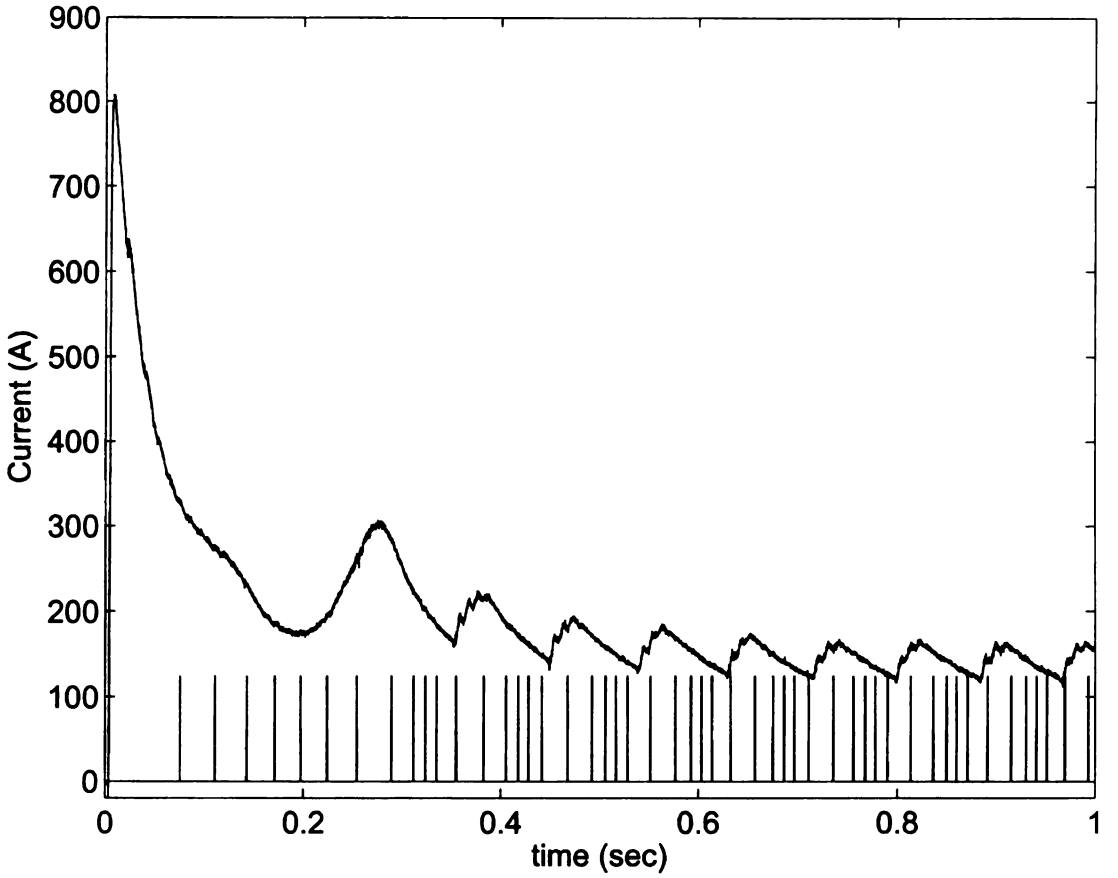

Figure 4.11: Sampled Current with Pulses

4.8 Hardware Optimization and Signal Enhancement

Modifications in the hardware were incorporated to optimize its efficiency and effectiveness. In the initial experimental setup, the output of the current senor was directly feed to the data acquisition system. However, the sensor was capable of handling $\pm 900A$. It produced $3.44V$ for $+900A$, $2.5V$ for $0A$ and $1.56V$ for $-900A$.

In this application, the current was always positive. It reached $700A$ during the initial rise, and remained below $400A$ after that, which corresponded to $3.24V$ and $2.92V$ respectively. The samples collected during the initial spike of the current were

rejected, leaving the samples for the time when the current is below $400A$ $(2.92V)$, which is the maximum current used for analysis. As, the current was not changing

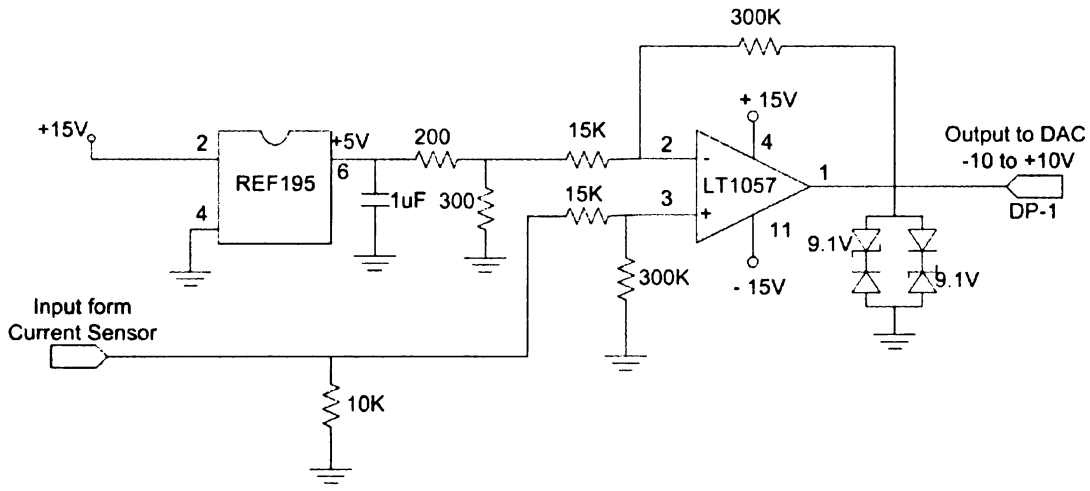

Figure 4.12: Signal Conditioning Electronics

direction and the output was always positive, the range of -10V to 0V was unused, means that the data acquisition card, which has 16 bit resolution for ±10V, was not completely utilized. The following modifications were made for the optimal utilization of hardware:

1. **Voltage Regulator:** The input to the DAQ card was limited according to the useful current range ($\sim 400A$). A zener diode based voltage regulated was placed which sets voltage threshold corresponding to 400A.

2. **Pre Conditioning:** The scale of the input was adjusted in order to utilize the complete range of the DAQ card. The DAQ acceptable range was $\pm 10V$ and the required input voltage range in this experiment was from 2.5V-2.95V. The input was translated to $-10V to + 10V$ respectively by using a differential operational amplifier.

3. **Sensor Utilization:** The current sensor can with stand with much higher

current (approx. $\pm 3500A$). Increasing current through the sensor effected the overall resolution of the systems. The number of turns through the current sensor was increased.

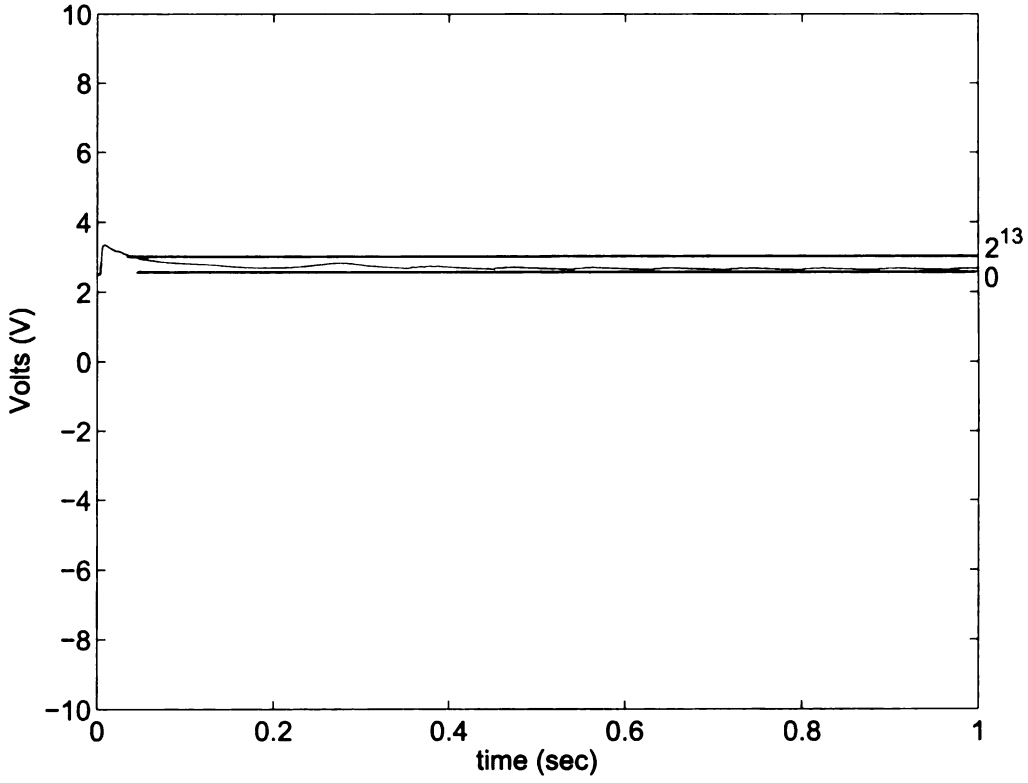

Figure 4.13: Unconditioned Data

The signal conditioning electronics are shown in Figure 4.12. The input to conditioning electronics was fed from non-intrusive Hall effect current sensor. It had $\pm 900A$ linear range. As the sensor can withstand up to $30kA$, the through current was doubled by passing the wire two times around it for maximum utilization of linear range. The output of current sensor was fed to the operational amplifier *LT1057*, which scaled and shifted it. A bidirectional Zener diode was placed at the output of operational amplifier to chop the initial high current and to block the negative

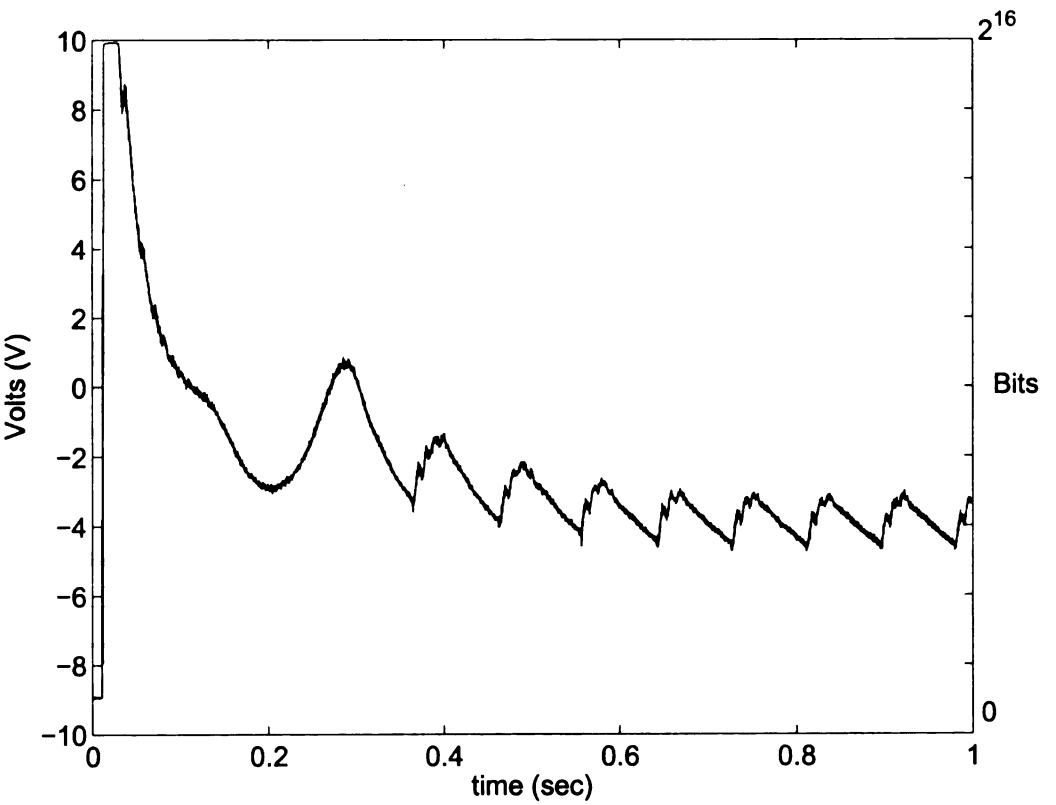

Figure 4.14: Conditioned Data

voltage. The scaling and shifting was performed according to the Eqn 4.1.

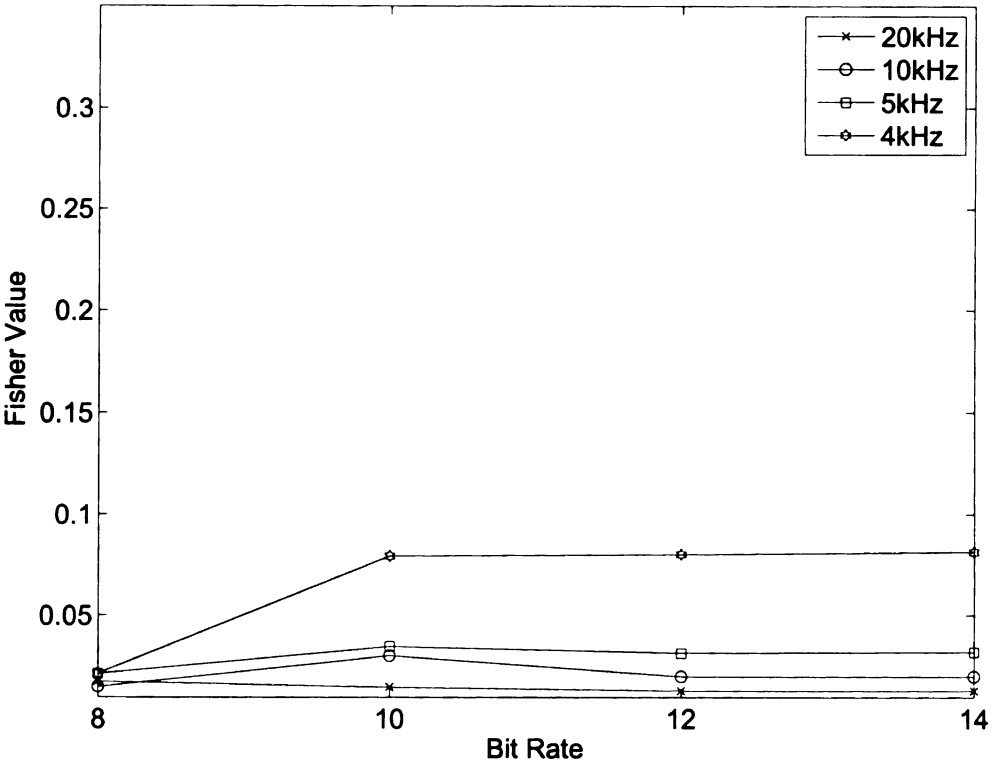

Figure 4.15: Projection using Original Data

$$V_0 = \frac{I_p}{24} - 240 \tag{4.1}$$

where I_p is the motor current and V_0 is the output Voltage. The original and the conditioned signal are shown in Figures 4.13 and 4.14, against the full input scale.

These modifications caused the optimum use of the NI DAQ card resolution. New data was sampled and the effects of down-sampling and bit reduction were observed and compared with the previously sampled ones. It was found that the information in the new data was more robust as the Fisher values were higher for the scaled data. Fisher values of the original data and conditioned data are shown in Figures 4.15 and 4.16.

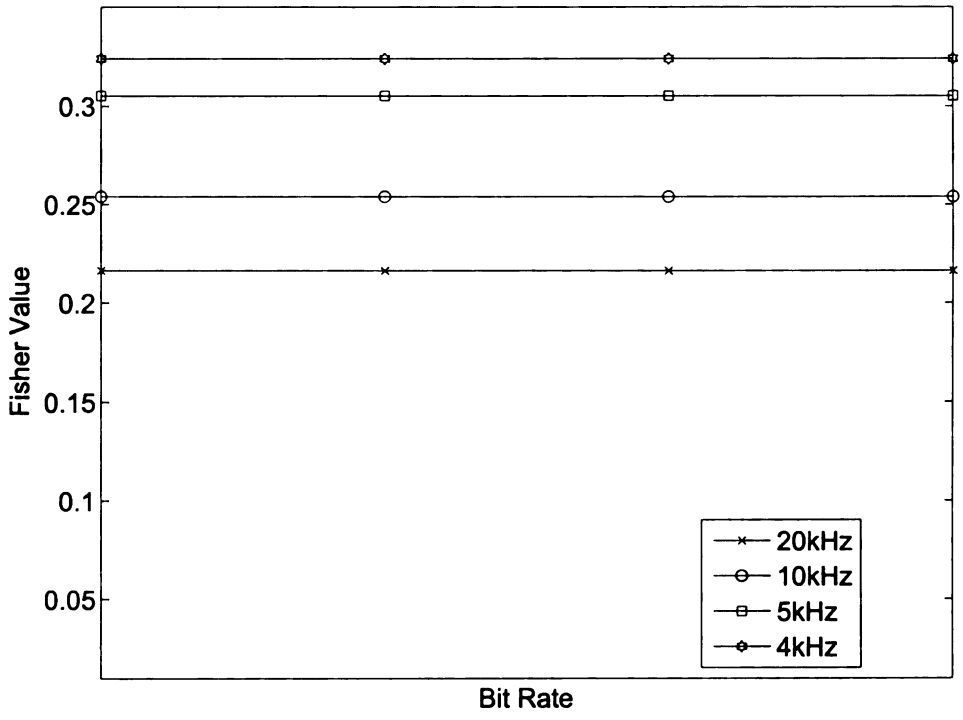

Figure 4.16: Projection using Conditioned Data

From Figure 4.16 it can be observed that the Fisher coefficient values remain within range if the bit rate and sampling frequency are reduced to 8 bits and $4kHz$. This implies use of low cost hardware for sampling. However, the same may not be true for the electric faults where the frequencies are expected to be higher.

4.9 Data Collection

During cranking, the flywheel was turned by starter motor gear pinion. At the start of cranking, the pinion slides out and meshes with the flywheel. The data were collected from healthy starter motors and damaged gear motors. Faults of different intensities were introduced in one tooth of a number of starter motors pinions. The faults were introduced by griding the tooth, which is similar to the griding of tooth due to improper meshing of gears. Figure 4.17 shows the healthy pinion of starter motor, and Figures 4.18 and 4.19 are motors with damaged teeth having different intensity fault levels.

Figure 4.17: Healthy

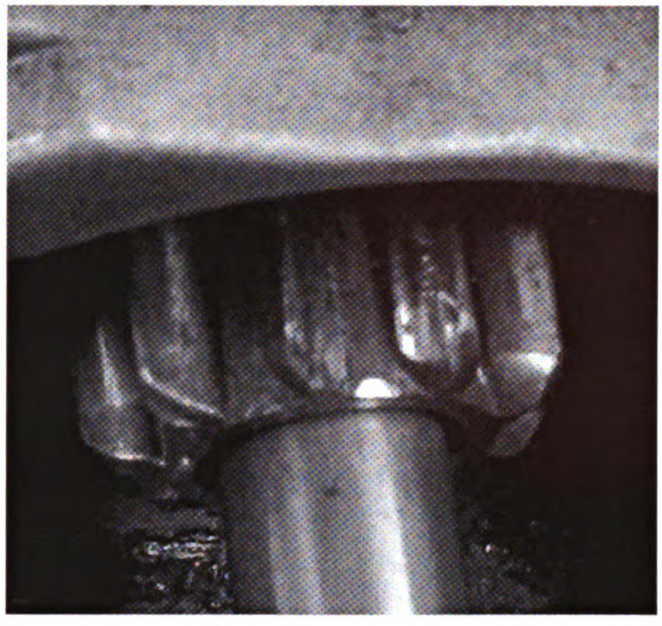

Figure 4.18: Fault Intensity I

The position of the damaged tooth was physically noted for each sample. The sampled current was defined by three separate and concurrent cycles: the compression-expansion cycle of the pistons, the flywheel rotation and the starter motor rotation. Out of these three, the starter motor cycle was of interest as it had the damaged tooth on it. The second cycle which plays a significant role in the analysis is the compression/expansion cycle. Assuming a healthy flywheel, the flywheel cycle had no role in this process.

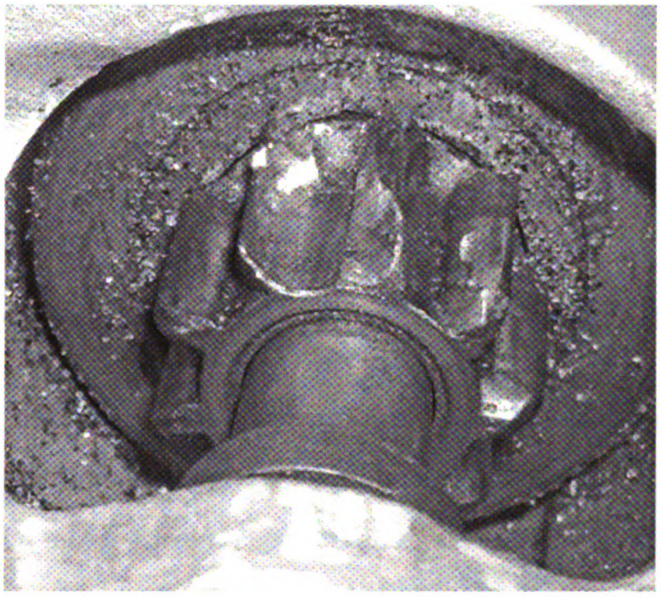

Figure 4.19: Fault Intensity IV

Chapter 5

Feature Extraction and Fault Diagnosis

5.1 Scope and Objective of the Chapter

In this chapter, the objective is to present the feature extraction and diagnostic methods suitable for the analysis of repetitive transient faults. Fault features were extracted from the motor current signature analysis in the time frequency domain. These features were the inputs of the classifiers. In this chapter, the working of classifiers with the consideration to the operational constraints is presented. A method to compute the discriminative strength of the features extracted using four different transformations is also presented.

Diagnosis of faults comprises of detection and categorization. Generally the categorization is performed after the detection of the faults. However, in the present application, the time of the damaged tooth meshing could not be identified precisely. This was due to the complexity of the load, inherent noise of the system and uncertainty in the location of damaged tooth. As the meshing time could not be pinpointed, the training of the algorithms was not very accurate. This made the

detections unreliable. Two groups of the classification algorithms are developed, one with detection of fault event and other without detection of fault events. In Section 5.2, the algorithms are presented which categorize once the detection has been made and in Section 5.3 the algorithms are presented which work without the detection of the faults. In section 5.4 the comparisons of the discriminative strength and execution time of the candidate transforms are made.

5.2 Fault Categorization - Known Meshing Instance of Damaged Tooth

Mostly the categorization algorithms are developed in way that they are executed after the detection of faults and they use only a small portion of measured signal. This make categorization computationally efficient. Categorization algorithms based on all four candidate transforms were developed on the similar approach.

5.2.1 Algorithm Based on the Short Time Fourier Transform

The input to the algorithm was the STFT of the measured motor current, i_m. For this analysis, *nfft*=64, *noverlap*=48, and a 64-point rectangular window is used. The parameters of the STFT are described in Chapter 2. The resultant STFT has 33 frequency bands, however the two outermost bands are discarded. The energy in these bands is far greater than in the inner bands of interest. After the initial high current, the DC component of i_m is approximately 250A. The algorithm is based on the remaining 31 frequency bands. The algorithm has two parts; a detection phase and a classification phase. The detection phase of the algorithm is based on thresholding on the time marginal of the STFT. In this work, data from fifteen experiments on a healthy machine were analyzed. The initial high current region was discarded and

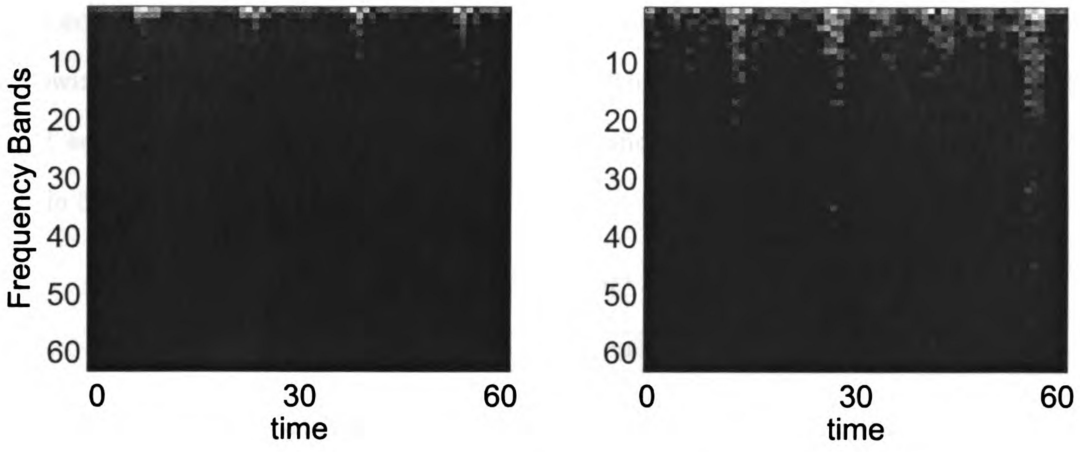

Figure 5.1: Measured currents and spectrograms

each signal was containing 8,192 time samples resulting in 509 STFT samples. The threshold on the time marginal was set to be 25% greater than the largest which was observed in all the healthy STFT samples from the healthy machine data. If the energy in new test data exceeds this threshold, a fault is considered to exist. During the selected time span of the signal, the damaged tooth meshes for four times. The algorithm considers each meshing as separate event and classification of each incident is performed separately.

The classification phase was based on pattern recognition classifiers. This phase is implemented when the criterion for detection is met. Three different classifier are tested in this work, LDC, NNC Euclidean distance and NNC with Mahalanobis distance. Training of certain parameter is required for all the three classifiers.

For the LDC, the weighting coefficients were to be determined. The data used to train the algorithm were 127 STFT samples of the high frequency bands from 15 experiments for each severity level of fault. From each experiment, four meshing events were recorded and total 300 (60 samples per class) were used for training. Following the training of the weighting coefficients, data that had not been used in the training algorithm were tested. Twenty meshing events (data from 5 experiments),

each containing 8,192 time samples resulting in 509 STFT samples, from each of the following operating conditions were tested: Healthy; fault severity 1; fault severity 2; fault severity 3 and fault severity 4. The classification results are presented in the Table 5.1

Table 5.1: STFT - LDC Classifier

	Tested	Correct	Incorrect
Healthy	20	01	19
Severity 1	20	05	15
Severity 2	20	11	09
Severity 3	20	05	15
Severity 4	20	08	12

For the NNC, two distances measures are used, Euclidean distance and Mahalanobis distance. For the Euclidean distance classifier, the means of the training samples of each class were computed. The data used to train the algorithm were 127 STFT samples of the high frequency bands from 60 meshing events (15 experiments). Following the training of the weighting coefficients, data that had not been used in the training algorithm were tested. Twenty meshing events (data from 5 experiments), each containing 8,192 time samples resulting in 509 STFT samples, from each of the following operating conditions was tested: Healthy; fault severity 1; fault severity 2; fault severity 3 and fault severity 4. The classification results are presented in the Table 5.2.

Table 5.2: STFT - NNC Classifier Euclidean Classifier

	Tested	Correct	Incorrect
Healthy	20	09	11
Severity 1	20	04	16
Severity 2	20	01	19
Severity 3	20	05	15
Severity 4	20	03	17

For the NNC, with Mahalanobis distance the variances of the data is also calculated. The data used to compute mean and variances for the algorithm were 127 STFT samples of the high frequency bands from 60 meshing events (15 experiments). The multidimensional variances are computed. Following the training of the weighting coefficients, data that had not been used in the training algorithm were tested. Twenty meshing events (data from 5 experiments), each containing 8,192 time samples resulting in 509 STFT samples, from each of the following operating conditions was tested: Healthy; fault severity 1; fault severity 2; fault severity 3 and fault severity 4. The classification results are presented in the Table 5.3.

Table 5.3: STFT - NNC Classifier Mahalanobis Distance

	Tested	Correct	Incorrect
Healthy	20	08	12
Severity 1	20	02	18
Severity 2	20	04	16
Severity 3	20	01	19
Severity 4	20	16	04

5.2.2 Algorithm Based on the Undecimated Wavelet Transform

For this analysis, the Daubechies D4 mother wavelet was used and decomposition was performed to 6 levels. The input of the algorithm is the measured motor current, i_m. The algorithm has two parts; a detection phase and a classification phase. The detection phase of the algorithm is based on thresholding on the weighted energy in the UDWT.

Data from fifteen experiments on a healthy machine were analyzed. The initial high current region was discarded and each signal was containing 8,192 time samples. The threshold was set at 12% greater than the largest which was observed in all

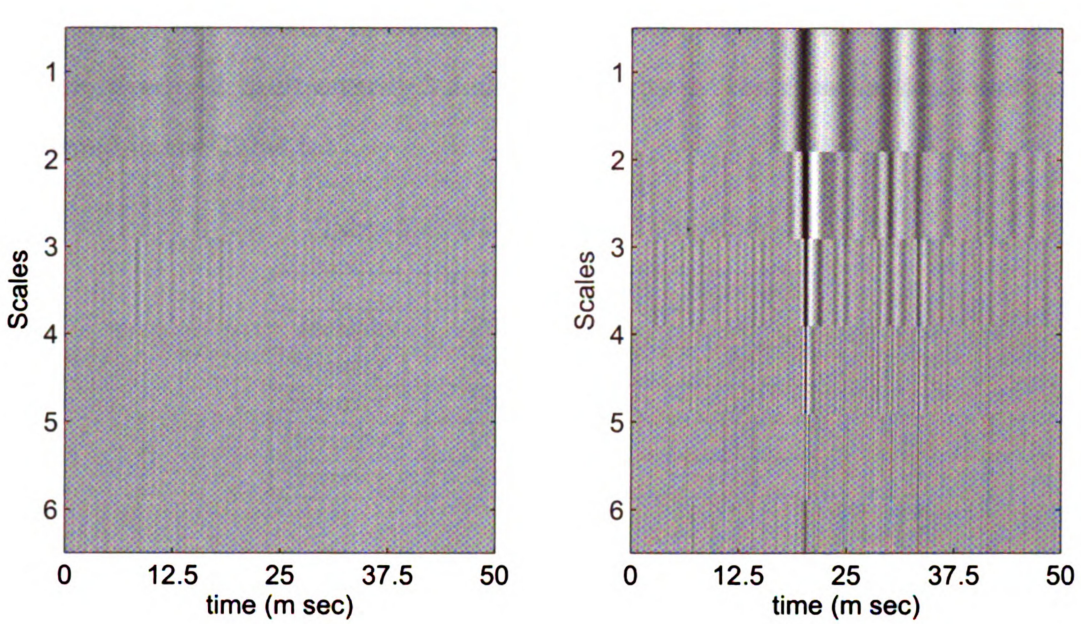

Figure 5.2: Measured currents and UDWT coefficients

samples from the healthy machine data. If the weighted energy in new test data exceeds this threshold, a fault is considered to exist. During the selected time span of the signal, the damaged tooth meshes for four times. The algorithm considers each meshing as separate event and classification of each incident is performed separately.

The classification phase was based on pattern recognition classifiers. This phase is implemented when the criterion for detection is met. Three different classifier are tested, LDC, NNC Euclidean distance and NNC with Mahalanobis distance. Training of certain parameter is required for all the three classifiers.

For the LDC, the weighting coefficients were to be determined. The data used to train the algorithm were 127 UDWT samples of the high frequency bands from 15 experiments for each severity level of fault. From each experiment, four meshing events were recorded and total 300 (60 samples per class) were used for training. Following the training of the weighting coefficients, data that had not been used in the training algorithm were tested. Twenty meshing events (data from 5 experiments),

from each of the following operating conditions were tested: Healthy; fault severity 1; fault severity 2; fault severity 3 and fault severity 4. The classification results are presented in the Table 5.4

Table 5.4: UDWT - LDC Classifier

	Tested	Correct	Incorrect
Healthy	20	20	00
Severity 1	20	12	08
Severity 2	20	03	17
Severity 3	20	08	12
Severity 4	20	08	12

For the NNC, two distances measures are used, Euclidean distance and Mahalanobis distance. For the Euclidean distance classifier, the means of the training samples of each class were computed. The data used to train the algorithm were 127 UDWT samples of the high frequency bands from 60 meshing events (15 experiments). Following the training of the weighting coefficients, data that had not been used in the training algorithm were tested. Twenty meshing events (data from 5 experiments), from each of the following operating conditions was tested: Healthy; fault severity 1; fault severity 2; fault severity 3 and fault severity 4. The classification results are presented in the Table 5.5.

Table 5.5: UDWT - NNC Classifier Euclidean Classifier

	Tested	Correct	Incorrect
Healthy	20	03	17
Severity 1	20	08	12
Severity 2	20	0	20
Severity 3	20	10	10
Severity 4	20	14	06

For the NNC, with Mahalanobis distance the variances of the data is also calculated. The data used to compute mean and variances for the algorithm were 127

UDWT samples of the high frequency bands from 60 meshing events (15 experiments). The multidimensional variances are computed. Following the training of the weighting coefficients, data that had not been used in the training algorithm were tested. Twenty meshing events (data from 5 experiments), from each of the following operating conditions was tested: Healthy; fault severity 1; fault severity 2; fault severity 3 and fault severity 4. The classification results are presented in the Table 5.6.

Table 5.6: UDWT - NNC Classifier Mahalanobis Distance

	Tested	Correct	Incorrect
Healthy	20	0	20
Severity 1	20	0	20
Severity 2	20	2	18
Severity 3	20	0	20
Severity 4	20	20	00

5.2.3 Algorithm Based on the Wigner Transform

For this analysis, the input to the algorithm is the measured motor current i_m. The input signal is decomposed in sixteen bands. In Wigner transform the kernel function is 1 which no effect on time frequency distributions. As the information in the band 1 to 8 is same as contained in bands 9 to 16, only the bands 1 to 8 are analyzed.

The algorithm has two parts; a detection phase and a classification phase. The detection phase of the algorithm is based on thresholding on the weighted energy in the WVD.

Data from fifteen experiments on a healthy machine were analyzed. The initial high current region was discarded and each signal was containing 8,192 time samples. The threshold was set at 25% greater than the largest which was observed in all samples from the healthy machine data. If the weighted energy in new test data exceeds this threshold, a fault is considered to exist. During the selected time span of

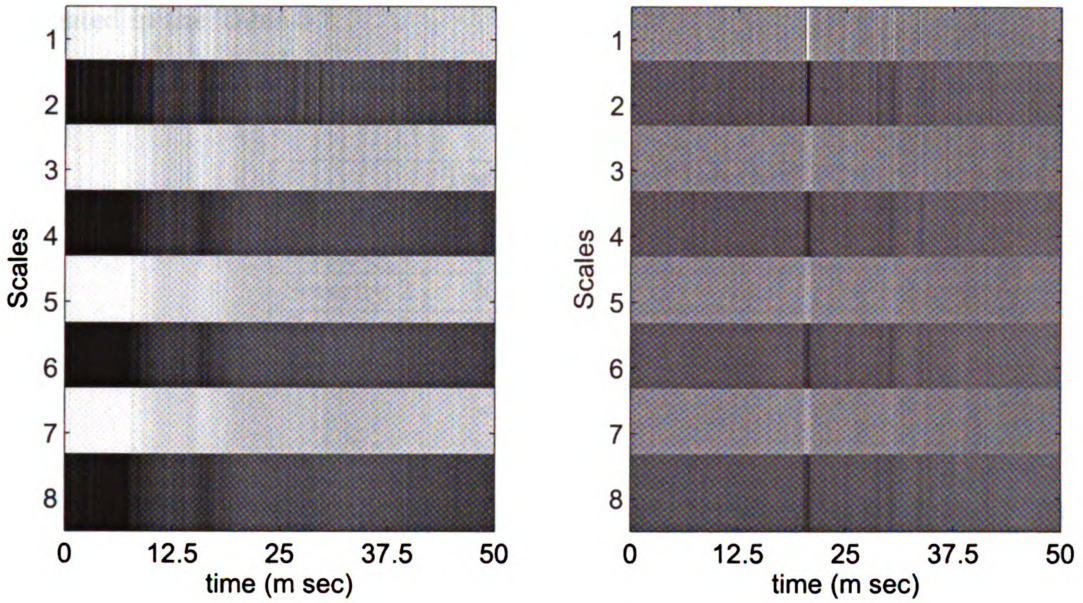

Figure 5.3: Measured currents and WVD coefficients

the signal, the damaged tooth meshes for four times. The algorithm considers each meshing as separate event and classification of each incident is performed separately.

The classification phase was based on pattern recognition classifiers. This phase is implemented when the criterion for detection is met. Three different classifier are tested, LDC, NNC Euclidean distance and NNC with Mahalanobis distance. Training of certain parameter is required for all the three classifiers.

For the LDC, the weighting coefficients were to be determined. The data used to train the algorithm were 127 WVD samples of the high frequency bands from 15 experiments for each severity level of fault. From each experiment, four meshing events were recorded and total 300 (60 samples per class) were used for training. Following the training of the weighting coefficients, data that had not been used in the training algorithm were tested. Twenty meshing events (data from 5 experiments), from each of the following operating conditions were tested: Healthy; fault severity 1; fault severity 2; fault severity 3 and fault severity 4. The classification results are

presented in the Table 5.7

Table 5.7: WVD - LDC Classifier

	Tested	Correct	Incorrect
Healthy	20	19	01
Severity 1	20	12	08
Severity 2	20	07	13
Severity 3	20	06	14
Severity 4	20	12	08

For the NNC, two distances measures are used, Euclidean distance and Mahalanobis distance. For the Euclidean distance classifier, the means of the training samples of each class were computed. The data used to train the algorithm were 127 WVD samples of the high frequency bands from 60 meshing events (15 experiments). Following the training of the weighting coefficients, data that had not been used in the training algorithm were tested. Twenty meshing events (data from 5 experiments), from each of the following operating conditions was tested: Healthy; fault severity 1; fault severity 2; fault severity 3 and fault severity 4. The classification results are presented in the Table 5.8.

Table 5.8: WVD - NNC Classifier Euclidean Classifier

	Tested	Correct	Incorrect
Healthy	20	18	02
Severity 1	20	09	11
Severity 2	20	05	15
Severity 3	20	06	14
Severity 4	20	10	10

For the NNC, with Mahalanobis distance the variances of the data is also calculated. The data used to compute mean and variances for the algorithm were 127 WVD samples of the high frequency bands from 60 meshing events (15 experiments).

The multidimensional variances are computed. Following the training of the weighting coefficients, data that had not been used in the training algorithm were tested. Twenty meshing events (data from 5 experiments), from each of the following operating conditions was tested: Healthy; fault severity 1; fault severity 2; fault severity 3 and fault severity 4. The classification results are presented in the Table 5.9.

Table 5.9: WVD - NNC Classifier Mahalanobis Distance

	Tested	Correct	Incorrect
Healthy	20	16	04
Severity 1	20	04	16
Severity 2	20	08	12
Severity 3	20	11	09
Severity 4	20	08	12

5.2.4 Algorithm Based on the Choi-Williams Transform

For this analysis, the input to the algorithm is the measured motor current i_m. The input signal is decomposed in sixteen frequency divisions. In the CWD, the smooth of the signal is controlled by the parameter σ. For the higher values of σ, smooth is less and the distribution approaches Wigner transform. The information in the frequency divisions 1 to 8 is same as contained in 9 to 16, therefore only the lower ones were used for the analysis.

The algorithm has two parts; a detection phase and a classification phase. The detection phase of the algorithm is based on thresholding on the weighted energy in the CWD. Data from fifteen experiments on a healthy machine were analyzed. The initial high current region was discarded and each signal was containing 8,192 time samples. The threshold was set at 25% greater than the largest which was observed in all samples from the healthy machine data. If the weighted energy in new test data exceeds this threshold, a fault is considered to exist. During the selected time span of

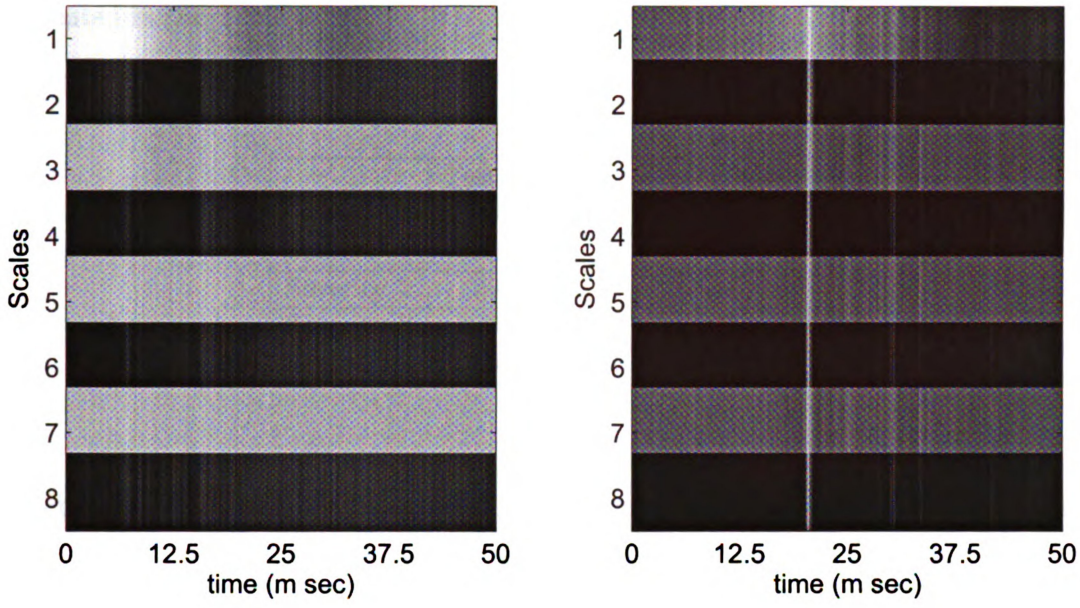

Figure 5.4: Measured currents and CWD coefficients

the signal, the damaged tooth meshes for four times. The algorithm considers each meshing as separate event and classification of each incident is performed separately.

The classification phase was based on pattern recognition classifiers. This phase is implemented when the criterion for detection is met. Three different classifier are tested, LDC, NNC Euclidean distance and NNC with Mahalanobis distance. Training of certain parameter is required for all the three classifiers.

For the LDC, the weighting coefficients were to be determined. The data used to train the algorithm were 127 CWD samples of the high frequency bands from 15 experiments for each severity level of fault. From each experiment, four meshing events were recorded and total 300 (60 samples per class) were used for training. Following the training of the weighting coefficients, data that had not been used in the training algorithm were tested. Twenty meshing events (data from 5 experiments), from each of the following operating conditions were tested: Healthy; fault severity 1; fault severity 2; fault severity 3 and fault severity 4. The classification results are

presented in the Table 5.10

Table 5.10: CWD - LDC Classifier

	Tested	Correct	Incorrect
Healthy	20	20	00
Severity 1	20	12	08
Severity 2	20	10	10
Severity 3	20	07	13
Severity 4	20	12	08

For the NNC, two distances measures are used, Euclidean distance and Mahalanobis distance. For the Euclidean distance classifier, the means of the training samples of each class were computed. The data used to train the algorithm were 127 CWD samples of the high frequency bands from 60 meshing events (15 experiments). Following the training of the weighting coefficients, data that had not been used in the training algorithm were tested. Twenty meshing events (data from 5 experiments), from each of the following operating conditions was tested: Healthy; fault severity 1; fault severity 2; fault severity 3 and fault severity 4. The classification results are presented in the Table 5.11.

Table 5.11: UDWT - NNC Classifier Euclidean Classifier

	Tested	Correct	Incorrect
Healthy	20	18	02
Severity 1	20	09	11
Severity 2	20	05	15
Severity 3	20	04	16
Severity 4	20	10	10

For the NNC, with Mahalanobis distance the variances of the data is also calculated. The data used to compute mean and variances for the algorithm were 127 CWD samples of the high frequency bands from 60 meshing events (15 experiments).

The multidimensional variances are computed. Following the training of the weighting coefficients, data that had not been used in the training algorithm were tested. Twenty meshing events (data from 5 experiments), from each of the following operating conditions was tested: Healthy; fault severity 1; fault severity 2; fault severity 3 and fault severity 4. The classification results are presented in the Table 5.12.

Table 5.12: UDWT - NNC Classifier Mahalanobis Distance

	Tested	Correct	Incorrect
Healthy	20	19	01
Severity 1	20	11	09
Severity 2	20	09	11
Severity 3	20	08	12
Severity 4	20	09	11

5.2.5 Classification Efficiency of the Classifiers

The categorization accuracy of the classifiers is shown in Figure 5.5. The LDC produces more accurate results as compared to the other classifiers. The performance of the classifier is better for the features which are extracted using CWD. The features extracted using WVD caused second best performance of the classifier. The categorization results were low for the features extracted using STFT. The classification results indicate that the CWD can efficiently represent transient repetitive faults.

5.3 Fault Categorization - Unknown Meshing Instance of Damaged Tooth

The results of the fault identification show that LDC is suitable for this type of faults. However, the problem of detection that a fault may be present is more complex. The

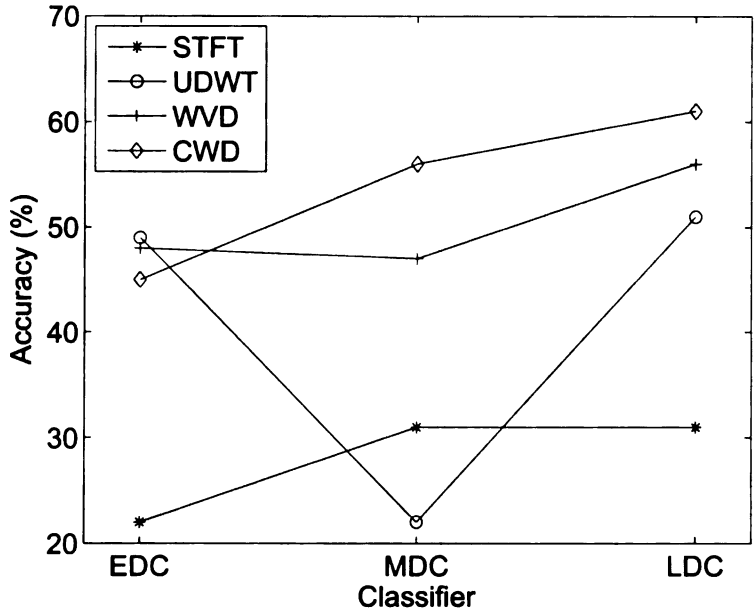

Figure 5.5: Classifiers' Accuracy

threshold energy of the transformed signals was used for this purpose, and to initialize the testing for a fault. However, the same is not accurate in the present case due to the complicated nature of operation of the starter motor, non linear load, the fault type and the similar energy levels of healthy and damaged cases. For the accurate classification, the extracted features should be discriminatively representative of their classes. During the training, the fault features are extracted from the motor current for the duration in which the damaged tooth is meshing. To collect data for the training of algorithms, the correct instance of the meshing of damaged tooth is mandatory, which a difficult task. This makes the identification of the meshing instance ambiguous, which can be one of the major reasons for the lower classification accuracy. A heuristic approach is used for the fault categorization in the absence of known starting instance.

Compared to expansion cycle, meshing of bad tooth during compression cycle

is more significant and prominent. In the proposed method, the fault recognition is preformed only during the compression cycle, using window sizes of progressively increasing length. The algorithm first searches for the start of the compression cycle, using the slope information of the current.

5.3.1 Recognition of the Commencement of Compression

During the compression the torque requirements are high, the motor current increases and the current waveform has a positive slope. To recognize this positive slope, a watchdog is used, which calculates the consecutive slopes. If it finds the expected number of consecutive positive slopes, the compression part of the cycle is assumed to have commenced. Once the system is in compression, the recognition phase starts.

5.3.2 Fault Recognition - Compression Cycle

Although, this phase commences when the system is in compression, the exact starting time of fault commencement is not known, if a fault exists at all. No analytical method or well defined criterion is available. Therefore, a heuristic method is developed and the following points are considered while defining the method:

1. *Repeated Incidences*: Typically during normal operation, the vehicle starts in less than a second of cranking. During this period, after an initial high current, there are approximately six to seven compression cycles and the starter motor completes approximately the same number of turns. If we misclassify a bad tooth as a good one (false negative), there are fairly good chances of detecting it correctly elsewhere.

2. *Error Cost*: False indication of fault (false positive) is assumed to be a more costly error [51]. Therefore only misclassification of bad tooth during few cycles is acceptable.

According to the proposed algorithm, the fault is sought using windows of different time lengths during the compression cycle, that is, first a window of smaller time length, w_1 is used, as shown in Figure 5.6. If consecutive incidences are classified as fault, they are investigated using more windows, of higher lengths. If all three windows identify the incidence as a fault, it is recorded as a fault, otherwise it is assumed a false indication. Current measurements of one second duration were

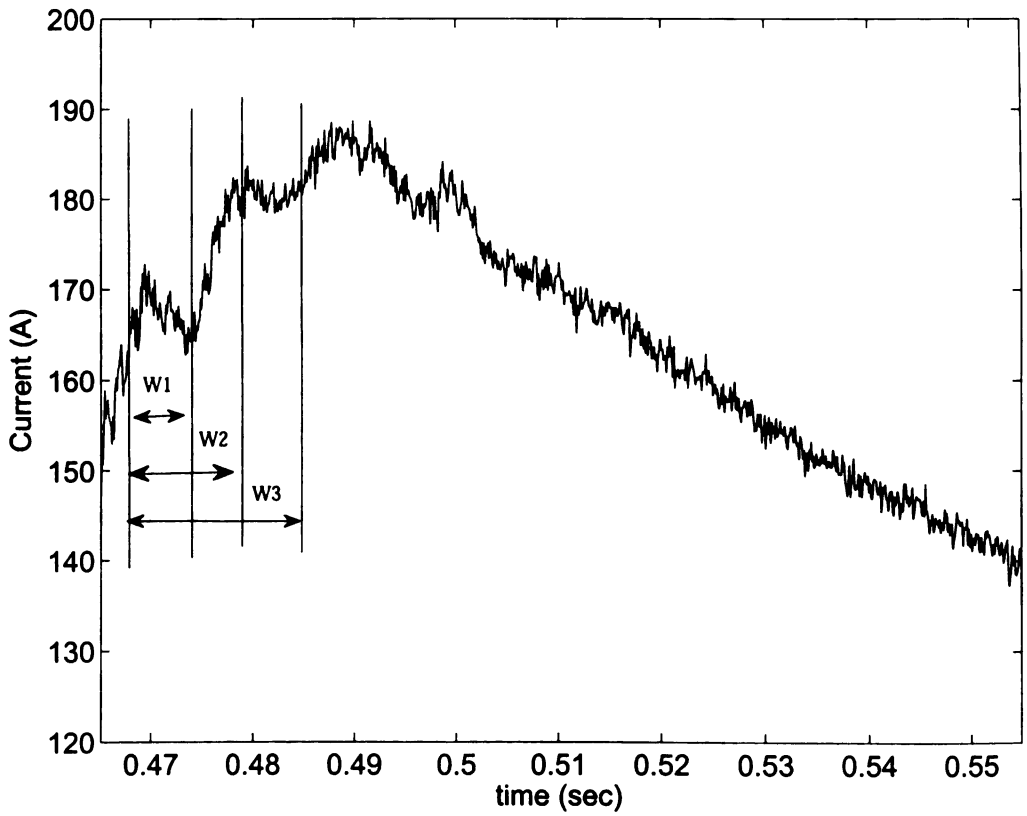

Figure 5.6: Motor Current with different length windows

sampled from healthy and defective motors. Each signal had seven compressions. Fifteen samples from each class were used for training the algorithm coefficients and five were used for testing. The results of the testing algorithm are shown in Table 5.13.

Table 5.13: Results of the Classification Algorithm identification

Class	32 Time Samples	64 Time Samples	128 Time Samples	Decision
Healthy 1	H	-	-	Healthy
Healthy 2	D	D	D	Defective
Healthy 3	D	H	H	Healthy
Healthy 4	H	-	-	Healthy
Defective 1	D	D	H	Healthy
Defective 2	D	D	D	Defective
Defective 3	D	D	D	Defective
Defective 4	D	D	D	Defective

It can be observed from Table 5.13 that by using three different window lengths, w_1, w_2 and w_3, the number of false positives is minimized. The initial window length is $w_1 = 32$ time samples, which is increased to $w_2 = 64$ and $w_3 = 128$ time samples. Each incidence is first investigated using the shorter window. If it is identified as faulty, for the required number of times, then further investigation is carried out using wider time window. If the algorithm recognizes the incidence as faulty using each window for required numbers, only then is the incidence recognized as a fault. The algorithm can be made more robust on the supervisory level, by introducing more stringent conditions to avoid false positives, which is considered a much costlier error.

5.4 Transform Discrimination Power

The Fisher ratio is used as figure of merit to measure the discriminative strength of the features generated by the distributions. It compares the within class spread of features of the same category and between class scatter of features from different categories as mentioned in Chapter 2. Ideally, the within class spread should be as low as possible and the between class scatter should be as high as possible. The Fisher

ratio is computed for all the four candidate transforms. From 20 sampled currents signals, four meshing events were recorded for each class making at total samples size of eighty meshing events per class. Fisher ratios were computed and are presented in the Table 5.14.

Table 5.14: Fisher discrimination ratio for transforms

Transform	STFT	UDWT	WVD	CWD
Fisher Value	0.0201	0.2899	2.3750	2.5040

From the computed Fisher values, it can be seen that the CWD has better discrimination strength as compared to any other transform. However, in terms of computational efforts, it is not efficient. In Table-5.15, the execution time for each transform is given. The CWD needs longer time as compared to the others.

Table 5.15: Execution time for transforms

Transform	STFT	UDWT	WVD	CWD
Time (sec)	0.0201	0.2899	2.3750	2.5040

As the CWD has better discrimination strength, it is expected that it will produce better results during diagnosis. It is the nature of application, which is the deciding factor in selection of the suitable transform. The CWD can be selected, provided the computational cost is not a concern. The discrimination strength of CWD depends on the value of smoothing function, σ. If the value of σ is selected high (near 10 or -10), the CWD approaches the Wigner distribution. The discrimination strength of CWD is maximum when the smooth function is set to 0.1. Fisher ratios of CWD for different values of σ are shown in the Table-5.7

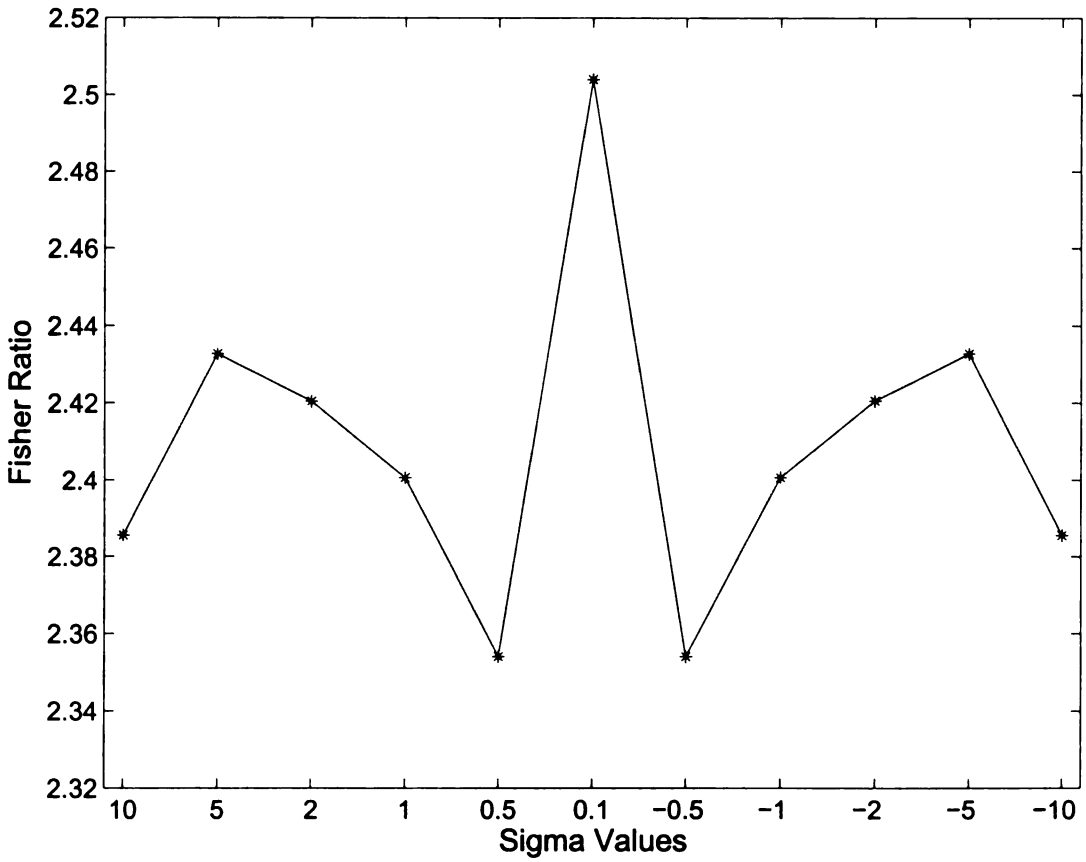

Figure 5.7: Fisher ratio for CWD for different values of σ

Chapter 6

Prognosis

6.1 Scope and Objective of the Chapter

During the prognosis phase, RUL of the system is computed. In case of continuously increasing severity event, the developed algorithm can compute the RUL in terms of time, however, in case of discontinuous event, the RUL cannot be specified as time. It can be implicitly computed in terms of probability of the failure. This approach enable users to set the threshold depending upon the criticality of the system. In this work, the develop prognosis algorithm is tested by applying it on the starter motor data. State motor operation is not continuous, therefore the RUL is computed in terms of probability of the failure.

This chapter describes the reasons of the selection of HMM as the prognosticator in section6.2. For the model three parameters needs to be defined, state transition probabilities(matrix $\boldsymbol{A}$), state-dependent observation densities (matrix $\boldsymbol{B}$) and initial state probabilities (matrix $\boldsymbol{\pi}$). The computation of these parameters becomes a challenging task if the available data is limited. The proposed methodologies developed to compute the model parameters using limited data are presented in section 6.3. In section 6.4, the algorithm for prognosis is explained. In section 6.5, shows how the

proposed methods for parameter computation are applied on the data sampled from the electromechanical system. The working of the algorithm is demonstrated using example and the future state probabilities are calculated. The result obtained from the parameter computation methods and the probabilities computed are presented in section 6.7.

6.2 Selection of Prognosticator

Generally, the candidate prognosticators are Kalman filter, particle filter or hidden Markov model. Kalman filter predicts the future state of the linear systems in the presence of Gaussian noise, however, if the system is a nonlinear one, extended Kalman filter(EKf) is the alternate option. However, in case of severe nonlinearities and non Gaussian noise, extended Kalman filter does perform good. On the other hand, although the particle filter method can work with non linear systems, it needs large of number of states to in order to converge. A more suitable option for prognosis of starting systems, is the hidden Markov model.

6.3 Model Parameter Calculation

The components of HMM need to be trained before it could be used for prediction. The training and testing phases are shown as part of Fig. 3.2 and Fig. 3.3 respectively. Parametric probability densities are defined for the state dependent observation densities using the experimental data. For the state transition probabilities a matching pursuit decomposition based method is presented.

6.3.1 State-dependent Observation Density B

The state dependant observation density **B**, $\left(b_j(O_t) = p(O_t|S_t = i)\right)$ is defined as the probability of observing a feature, given that the machine is in fault severity state S_i. According to the Bayes rule, this probability is given as (6.1):

$$P(O|S_i) = \frac{P(S_i|O) \times P(O)}{P(S)} \tag{6.1}$$

These probabilities could not be computed in closed form. It was assumed that the probabilities are Gaussian and were defined as Eqn-6.2.

$$P(O|S_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(\frac{O - \mu_{O|S_i}}{\sigma_{O|S_i}}\right)^2 \tag{6.2}$$

The statistics were obtained from the experimental data. The output of training of the LDC classifier was a set of coefficients ($\alpha \in \Re^{k+1}$) for each class, which gives the maximum discrimination for the corresponding class. The sets of coefficients were used as projection hyperplanes, and each training sample, $x^j = \{x_1^j \cdots x_k^j\}, x^j \in \Re^k$, was projected to all C planes as:

$$D_c(x^j) = x_1^j\alpha_{1c} + x_2^j\alpha_{2c} + \ldots + x_k^j\alpha_{kc} + \alpha_{k+1,c} \quad c = 1, 2, \ldots, C \tag{6.3}$$

The statistics for the probabilities densities were computed from these projections. For each of C classes, mean $\mu_{O|S}$ and standard deviations $\sigma_{O|S}$, were calculated corresponding to the projections on each of the C planes, generating a set of $C \times C$ means and $C \times C$ variances.

6.3.2 State Transition Probabilities Matrix A

In order to determine the state transition probabilities, large amounts of historical/collected training data were needed. Following were the possible options:

- *Large Scale Testing:* carried out on large samples of machines and allowing faults to naturally develop to failure over a long period of time.

- *Fatigue Analysis* of the meshing faces of starter motor gear and flywheel was an alternative. It may include finite element analysis, stress analysis and/or non destructive testing.

- *Online Estimation:* The transition probabilities can be estimated online during the operation of the motor. Transition probabilities could be estimated if they are considered similar to earlier transitions.

However, in the case of highly reliable electrical machines, the first two options were not practical. The third one could only be used during the actual operation of the machine, where faults develop naturally. For the laboratory setup, a mix of experiments and heuristic methods were used to estimate these probabilities.

The State Transition Probability $\mathbf{A}$ $\left(a_{ij} = p(x_{t+1} = j | x_t = i)\right)$, has on the diagonal the probabilities of self state transitions (SST) and at off diagonal other state transition probabilities (OST). For the calculation of OST probabilities, a method based on Matching Pursuit Decomposition (MPD) was developed.

6.3.2.1 Matching Pursuit Decomposition

In general, any signal $f \in X$ can be represented by a linear combination (finite or infinite) of signals g_k, provided that the set of signals g_k, forming the dictionary D, spans the space, means is complete. A signal can be represented as a sum of

predefined signals $g_k, k \in \{0, 1, 2 \cdots, K\}$, (in general $\|g_k\| = 1$), as

$$f = \sum_{k=0}^{M-1} a_k g_k \quad k \in \{0, 1, 2....M\} \tag{6.4}$$

where g_k is the k^{th} atom, a_k is the k^{th} weighting coefficient and M is the number of dictionary atoms. Equation 6.4 decomposes the signal f as a linear combination of M signals. The dictionary may contain more elements than necessary to span the space X, that is, M may be larger than the dimension of the signal space. The key issue is how to obtain the coefficients α_k and the atoms g_k. Usually, a signal approximation is generated by using a number of atoms smaller than M. In general, this approximation improves as the number of atoms used increases. Usually, the dictionary must contain a very large number of structures g_k to enable the decomposition of 6.4 to coherently identify the different phenomena composing a signal. The dictionary should contain distinct structures for any phenomena that would be present in the systems, meaning that the dictionary may be quite large. That is why K is, in general, larger than the dimension of the signal space. In this case, the dictionary is called redundant or overcomplete.

Matching Pursuit (MP) Decomposition is an adaptive approximation algorithm, which decomposes the signal in terms of the elements, i.e., atoms (g_k), of an over-complete dictionary. It was introduced by Mallat and Zhang[96]. At each step, the MP chooses the atom in the dictionary that best represents the signal (the atom with largest inner product with the signal). The chosen atom is then scaled and subtracted from the signal, and the process is repeated, representing the signal by progressive approximations. The atoms are selected by a iterative greedy algorithm, which chooses the atom in a dictionary that best represents the signal.

$$g_{\gamma_k} = \arg max_{\gamma_i} \langle R^k f, g_{\gamma_i} \rangle \tag{6.5}$$

where R^k is the residue at the kth step and is initially set as the signal f. The chosen atom is then scaled and subtracted from the previous residue.

$$R^{k+1} = R^k - \langle R^{k+1}, g_{\gamma_{k+1}} \rangle g_{\gamma_{k+1}} \tag{6.6}$$

The process is repeated, representing the signal by progressive approximations. The outcome of the MP decomposition is:

$$f = \sum_{k=0}^{M-1} \langle R^k f, g_{\gamma_k} \rangle g_{\gamma_k} + R^K f \tag{6.7}$$

At each step of the MP, two pieces of information must be stored, the coefficient α_k and the index γ_k defining the atom and more information about the signal is extracted. In this way we can arranged the complete set of dictionary atoms. The arrangement of dictionary atoms depends on the signal being represented by it. The information contained in the positioning of atoms, in MP decomposition of signals is manipulated to get the state transition probabilities.

6.3.2.2 Computation of State Transition Probabilities by Matching Pursuit Decomposition

The time frequency representation of two signals differs by the ordering of the atoms, reflecting the dissimilarity between them. This ordering was used to obtain a probability estimate. The following observations were made when analyzing the decomposed samples:

- The samples of same class had very similar ordering of the atoms.
- The samples from two different classes, which were closely matched, had small variation in ordering of atoms.

- The samples from two different classes, having large difference in fault severity, showed large variations in atoms' ordering.

Based on the above observations, a method to calculate OST probability was developed, using the relative atom's position. Samples from each class were decomposed and the mean of the decomposition is calculated as class representative. The relative difference (Δ) in the ordering of atoms of the representative decompositions was used for transition probability estimation. Δ was defined as (6.8):

$$\Delta_{ij} = \sum_{k=0}^{M} \frac{|\gamma_k^i - \gamma_k^j|}{\gamma_k^i} \; 1 \leq i, j \leq N \tag{6.8}$$

where γ_k is the rank of atom and N is the number of states.

The variances of each state were used to estimate SST by projecting them on a inverse unit scale through a sigmoid function transform to the range of 0-1. The sum of transition probabilities to other states except itself is *1-self transition probability.* The transitional probabilities are computed as

$$a_{ij} = \begin{cases} SST_{ii} & i = j, \\ \frac{(1-SST_{ii})}{\Delta_{ij}} & i \neq j \end{cases} \tag{6.9}$$

6.3.3 Initial State Distribution Vector π

The initial state probability for the faults should be obtainable from the manufacturer or the repair facilities. However, such data was not easily available. Reasonable probabilities were assumed to illustrate the methodology. The initial state probabilities for a new starter motor should be assumed to be 1 for healthy state and 0 for all the rest. The fault prognosis method then would then track the evolution of a fault. For two reasons different probabilities set was selected in this case: first, we did use used starter motors, and secondly in the sequence of observations we used for testing, an

initial distribution as mentioned above, only delays the transition to a faulty state and the manifestation of the prognosis of failure. In practice, many more observations will be taken before a healthy motor fails.

6.4 Algorithm for Future State Probability Estimation

Intuitively, if there is no additional information available, the machine condition will be the predicted based on the initial state probabilities only. However, if somehow, information about the present state of operation of the machine is available, the initial probabilities can be update in light of new measurements and more likely probabilities of the present state can be obtained. In order to make prediction for the future state, from the updated present state, information about the transition chances from one state to other is required. If this information is available, we can predict the future state of operation. The components of HMM, matrix A, B and π, provide these required inputs. A prognosis algorithm is developed based on HMM model.

The goal of prognosis is to assess the time to failure (remaining useful life) or the next probable state. In this work an algorithm is developed to estimate the latter one at each time sample. We define $\delta_t(i)$ as the normalized forward probability at time t for each state S_i. The state transition probabilities, a_{ij} and $\delta_t(i)$ are used to predict the probabilities of states at time $t+1$. The transition probability to state S_j at the time instance $t+1$ is given by:

$$\mathrm{P}[\mathrm{q_{t+1}}=\mathrm{S_j}|\lambda] = \sum_{\mathrm{i}=1}^{\mathrm{j}} \mathrm{P}[\mathrm{q_t}=\mathrm{S_i}|\lambda]\mathrm{a_{ij}} = \sum_{\mathrm{i}=1}^{\mathrm{j}} \delta_{\mathrm{t}}(\mathrm{i})\mathrm{a_{ij}} \tag{6.10}$$

where λ is the set of model parameters. The most probable state at time $t+1$ is the one that has the highest probability. The predicted state probabilities are updated

using state dependent observation $b_t(j)$ at each time step. The algorithm works as follows:

- Initialization : In the initialization phase the initial state probabilities are update based on the freshly acquired information. δ contains the update probabilities about the machine condition at the present time. The HMM is doubly statistical, therefore the states are probabilistically estimated from the information acquired from motor current.

$$\delta_1(i) = \pi_i b_i(O_1) \quad 1 \leq i \leq N \tag{6.11}$$

$$q_1(i) = 0 \quad 1 \leq i \leq N \tag{6.12}$$

- Recursion : In the recursion phase the future state probabilities are estimated. This estimate is made based on the probabilities of current state and transition probabilities. The transition probabilities were computed before hand from the collected data.

$$q_t(j) = \arg\max \sum_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] \quad 2 \leq t \leq T, \; 1 \leq j \leq N \tag{6.13}$$

Finally the state which has the largest probability at future time is the most likely state.

$$\delta_t(j) = \sum_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(O_t) \quad 2 \leq t \leq T, \quad 1 \leq j \leq N \tag{6.14}$$

6.5 Prognosis - Implementation

The proposed method and algorithm were tested using the sampled data from the experimental setup. The model parameters were computed and the algorithm working

was demonstrated by illustrated examples.

6.5.1 Matrix A Calculation

To calculate the state transition probability, data were analyzed from fifteen experiments each on a healthy machine and four machines with damaged gear of increasing intensity. The initial high current region was discarded and each signal was containing 8,192 time samples. Four meshing of the damaged tooth occur during this period and each meshing is considered as a separate event. The motor current representing each event was decomposed by MPD using a Gabor dictionary of 3905 atoms. It generated 60 sets of 3905 atoms per class, one for each sample. Had the faults been reversible, there would be five possible transitions from each state, including self transitions. However, the severity progression of the gear faults is unidirectional, leaving only the left-right transitions possible and an upper triangular transition probability matrix. The initial estimates of state transition probabilities are calculated using (6.9) and are given in Table 6.1.

Table 6.1: State Transition Probabilities Matrix (**A**)

Class	1	2	3	4	5
1	0.5063	0.2435	0.1481	0.0800	0.0221
2	0.0000	0.4935	0.2967	0.1751	0.0347
3	0.0000	0.0000	0.4542	0.3709	0.1749
4	0.0000	0.0000	0.0000	0.5829	0.4171
5	0.0000	0.0000	0.0000	0.0000	1.0000

6.5.2 Matrix B Calculation

Matrix **B** describes the probability distributions of the fault observations. It was assumed that these distributions are Gaussian. The statistics to calculate them were obtained using the experimental data. The diagnosis phase gave five sets of LDC

weighting coefficients, one for each class.The features of the samples from each class were projected on the corresponding set of LDC weighting coefficients planes. The statistics of the distributions are given in Tables 6.2 and 6.3. The mean of the projection of the samples on their own corresponding planes is on the diagonal and it is higher as compared to the other means. The overlapping variances reflect the chances of the outliers to transition from one class to other. The histograms of the projection of sample data of class 3 on all the LDC planes are shown in Fig.6.1. The mean is shown as vertical line in each subplot, which is largest for the projection on planes 3.

Table 6.2: Means of the projection on each plane

	P_1	P_2	P_3	P_4	P_5
1	**0.0759**	-0.0008	0.0087	-0.0052	0.0012
2	0.0148	**0.0793**	0.0213	0.0042	0.0143
3	0.0582	0.0449	**0.1284**	0.0533	0.0525
4	0.0665	0.0537	0.0379	**0.2726**	0.0223
5	0.2117	0.2161	0.2326	0.2552	**0.4383**

P_i is the LDC plane of fault i

Table 6.3: Variances of the projections of the samples from each class on LDC planes

	P_1	P_2	P_3	P_4	P_5
1	0.0081	0.0084	0.0103	0.0140	0.0138
2	0.0088	0.0066	0.0112	0.0101	0.0143
3	0.0231	0.0294	0.0336	0.0410	0.0358
4	0.1134	0.1282	0.1934	0.1712	0.2363
5	0.1753	0.1669	0.2522	0.2631	0.2612

P_i is the LDC plane of fault i

Once a new sample is obtained at time instance t, the state dependent observation probabilities, $b_i(O_l)$ (where i is the number of the state), were calculated.

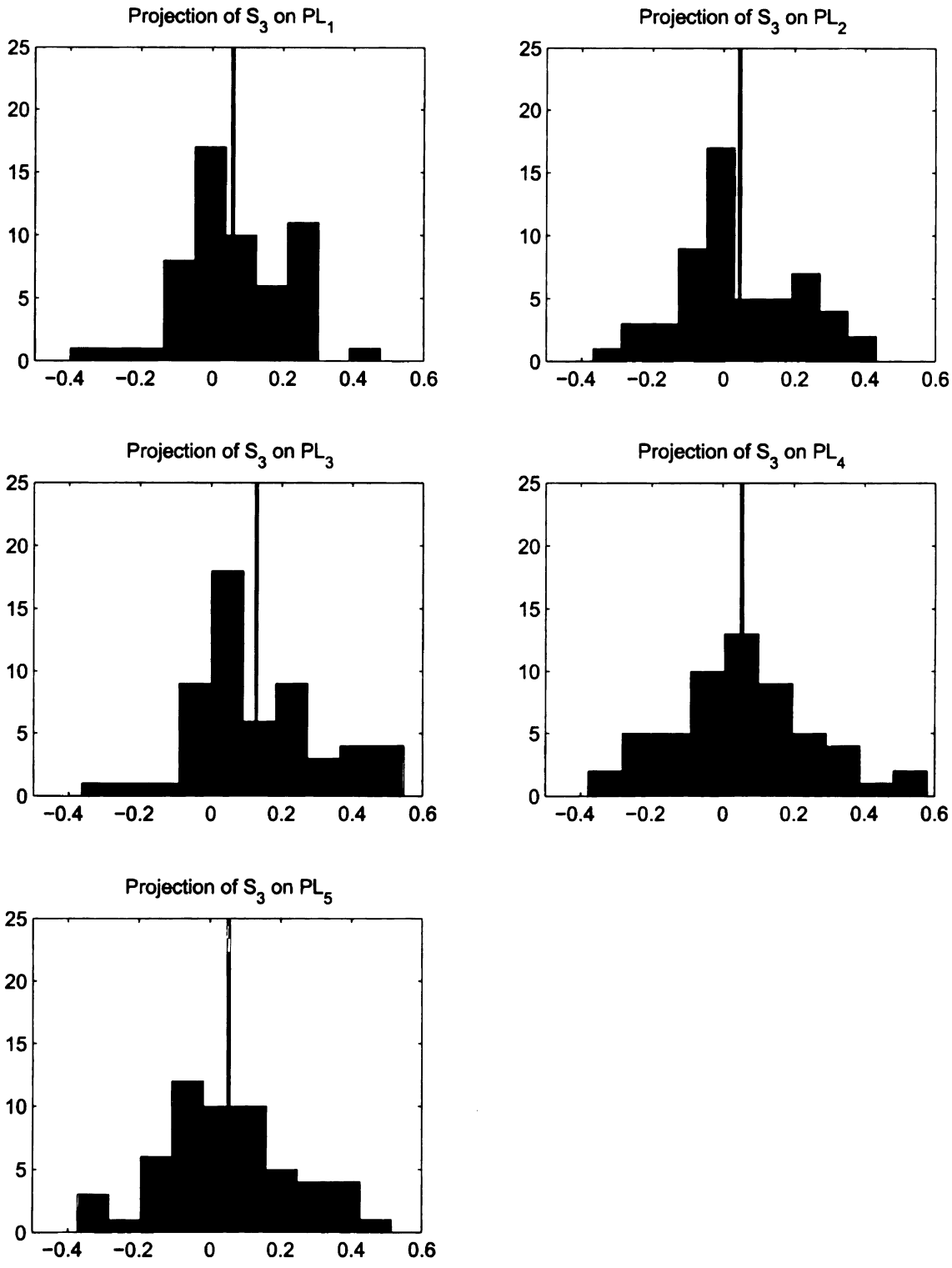

Figure 6.1: Projections of Class 3 on LDC Planes

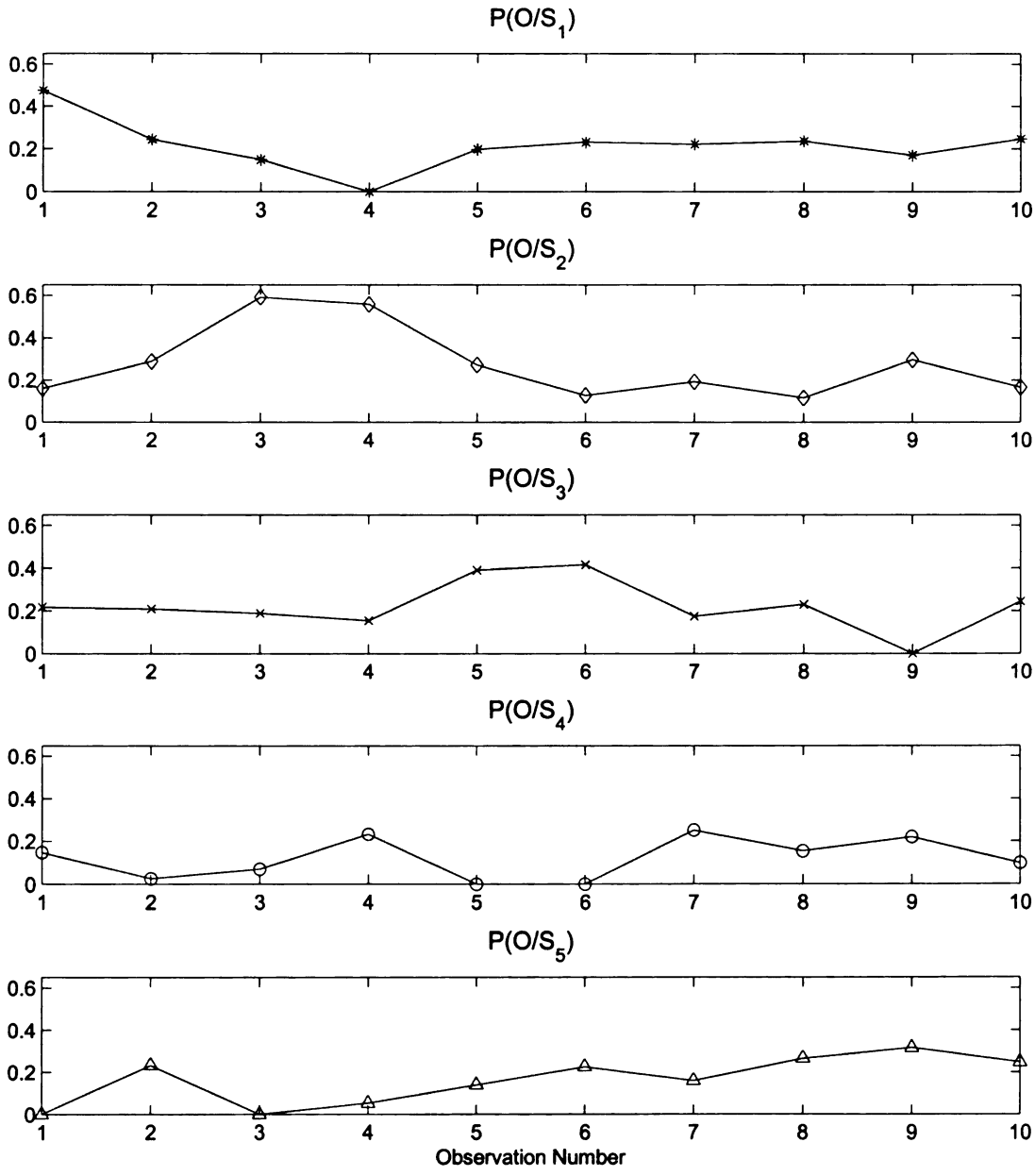

Figure 6.2: State Dependent Observation Probabilities -Example I

6.5.3 Initial State Distribution Vector π

The possible options for the initial state probabilities were usually obtained from the manufacturer, repair facilities or experimentally by large scale sample analysis. Here,

in order to demonstrate the proposed method, arbitrary values of the initial state probabilities were assumed and are given in Table 6.4.

Table 6.4: Assumed Initial State Probabilities (π)

Class	1	2	3	4	5
Probability	0.6	0.2	0.1	0.1	0

6.6 Examples of Future State Probability Estimation

Although the prime objective of prognosis is to estimate the remaining useful life, the developed algorithm computes the probability of all the failure states during the next starting attempt, thus implicitly giving the remaining useful life. It is not practical to test the algorithm using the data acquired from the field, where fault inception and progression are natural and slow, nor was it possible to 'run to failure' a large number of starter motors in the laboratory. Therefore, data from the machines with artificially created faults arranged in order of increasing severity, were used to test and evaluate the algorithm. The algorithm was tested using two data sets:

6.6.1 Illustrative Example I

To emulate the development of a fault, data was sampled from a number of starter motors. These samples were ordered corresponding to increasing fault severity: S_1, S_2, S_2, S_2, S_3, S_3, S_4, S_5, S_5, S_5. The values of the state-dependent observation probabilities are calculated for each sample (shown in Fig. 6.2). The first value was used to initialize the algorithm. The estimated next state sequence and failure state probabilities are shown in Fig. 6.3 and 6.4 respectively.

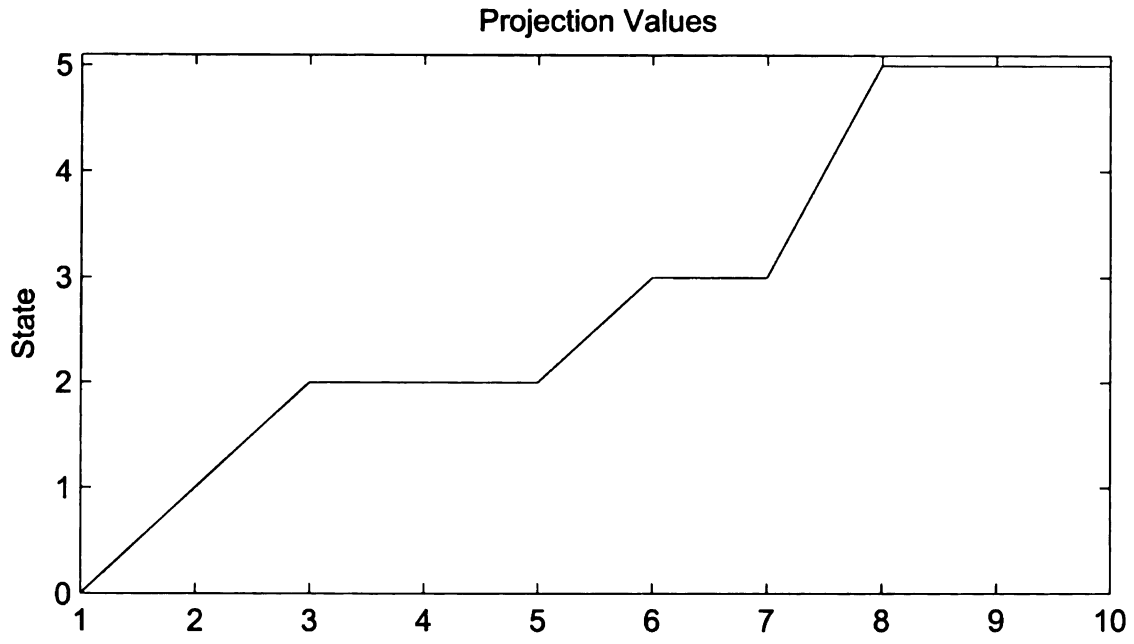

Figure 6.3: Probable Next State - 10 Samples, Example I

6.6.2 Illustrative Example II

The algorithm was also tested using artificial data, sequenced in way that the fault severity was in increasing order. Ten samples from each of the five classes were generated by adding white noise to the actual observations. The observed values of $b_t(i)$ were used as mean of the white noise and variance of the noise was set to be 30% of $b_t(i)$. The estimated next state sequence and failure state probabilities are shown in Fig. 6.5 and Fig. 6.6 respectively.

6.7 Prognosis - Results

6.7.1 Training of HMM Parameters

From the computed values, it can be observed that the self transition probabilities are higher compared to other state transition probabilities, which means that is generally more likely for the motor to continue operating in the same class. The other obser-

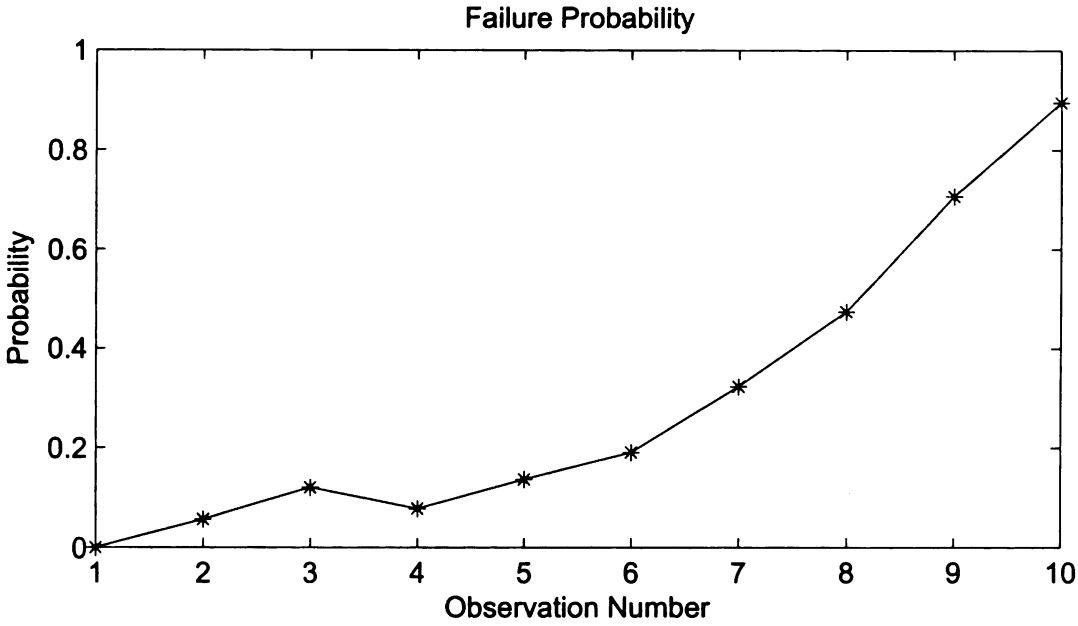

Figure 6.4: Failure State Probability - 10 Samples, Example I

vation is that the higher the difference in the intensity of faults in consecutive states, the lower the transition probability is, which conforms to the general understanding of fault progression.

The statistics for the state dependent observation probability densities are computed from the projections of the training samples on the LDC planes.The mean of projections is the highest when the samples are projected on the corresponding plane, which concurs with the expectations.

6.7.2 Examples

The proposed algorithm is illustrated by two examples, in which the fault state of the machine during the next starting attempt is predicted. The algorithm selects the most probable next state as the predicted state. Creating a large number of naturally developing sequences of machine gear faults was not practical. Therefore, sequences of increasing severity of artificially induced physical faults were generated to test the

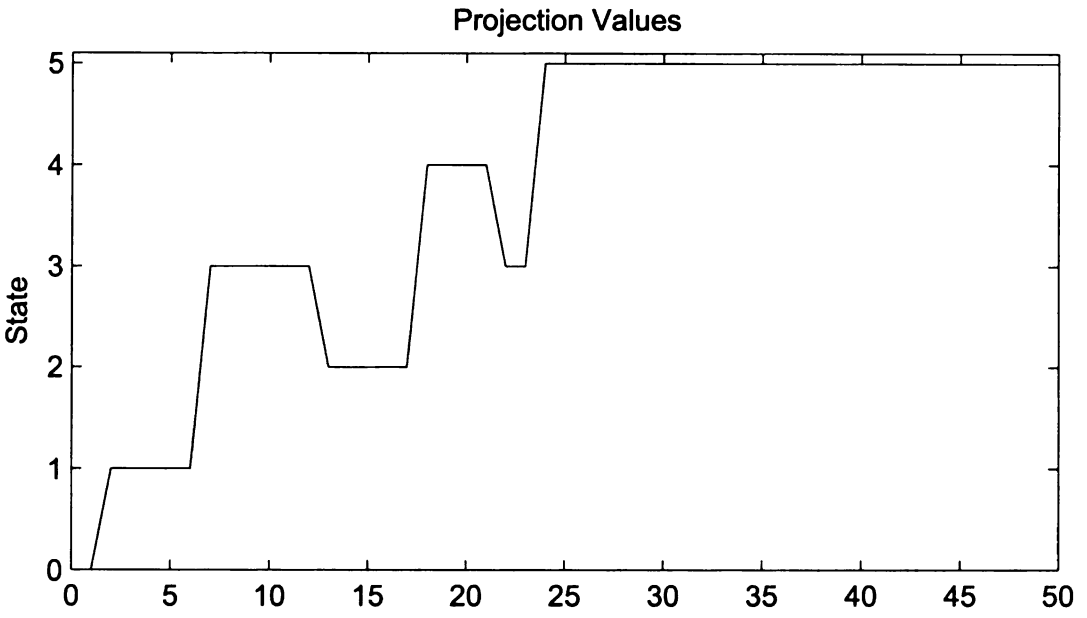

Figure 6.5: Probable Next State, Example II

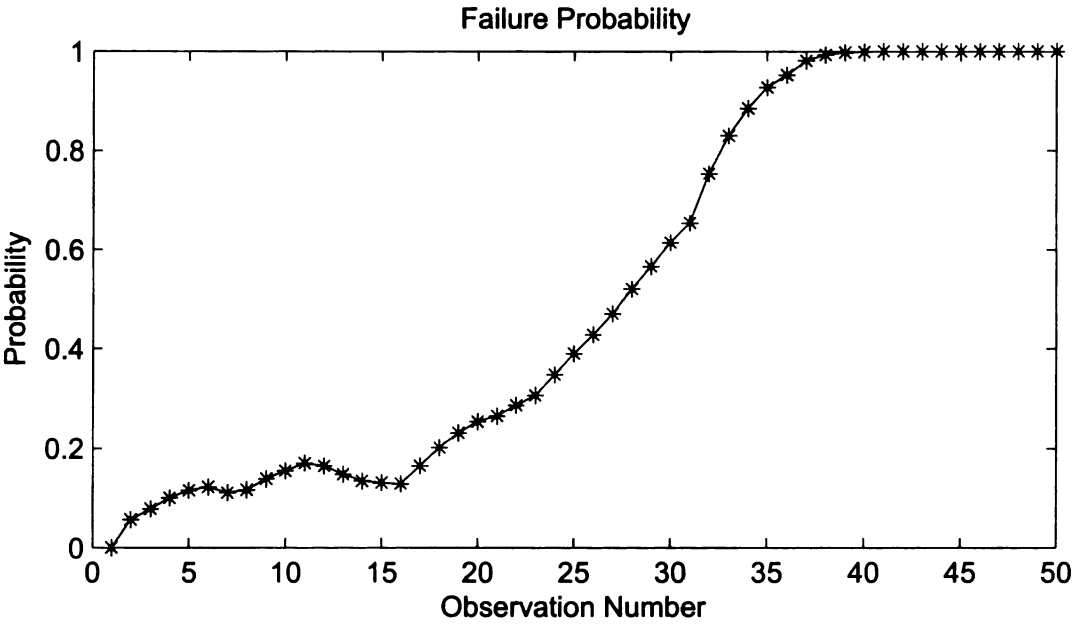

Figure 6.6: Failure State Probability, Example II

proposed method. The predicted next probable states are shown in Fig. 6.3 and 6.5 for the two examples. The prediction accuracy of the next probable states is 70% and

the rest 30% are predicted with a difference of only one state. This error might be due to the very small difference in fault severities of different classes.

It was expected that when starting from a healthy motor the probability of reaching the failure state in the next start (state 5) would be low, and towards the end of the sequence this probability would be the highest. The estimated results show this trend. Figures 6.4 and 6.6 show failure state probabilities for two examples, supporting the expected trend.

Chapter 7

Conclusions and Suggestions for Future Work

7.1 Conclusions

This thesis presented the theoretical foundation and implementation of a complete framework for the diagnosis and prognosis of complex electromechanical systems. In this work, non-intrusive signal based methods were used. The analysis was performed using the fault features, extracted in the time frequency domain from the motor current. Diagnosis and prognosis of the repetitive and transient faults was performed.

Different time frequency distributions were compared. The candidate transforms were the Short Time Fourier Transform, Un-Decimated Wavelet Transform, Wigner Transform and Choi-Williams Transform. The Fisher discriminant ratio of the fault features generated by respective transforms were used as the figure of merit. Choi-Williams transform had the highest Fisher ratio, however, the computation time required for this transform was high. The un-decimated wavelet transform had lower Fisher ratio however, it needed much less computational time.

A general framework for diagnosis was presented, in which fault categorization

methods were discussed with and without explicit detection of fault instances. Different pattern recognition classifiers were employed, and classification accuracy and computation cost were calculated. The linear discriminant classifier, Euclidean distance classifier and Mahalanobis distance classifier were tested. It was observed that the classification accuracy of linear discriminant classifier was better, though it was computationally inferior than the distance classifiers. The multiple discriminant analysis based distance classifiers were also analyzed, however, it is sensitive to training data, which made it unsuitable for the application as a supervised learning method. The classifiers were not only used for the categorization, but also for the collection of statistics for the prediction phase. The output of the training phase of the linear discriminant classifier was a set of discriminating planes, LDC planes. On these planes, the extracted features were projected and the statistics were computed to be used during the prognosis phase.

A failure prognosis algorithm was presented, based on the statistical modeling method, Hidden Markov Model. Methods to train the HMM elements were developed for the case of sparse data set. The state transition probabilities were estimated from the relative order of atoms in matching pursuit decomposition of the measured motor current signals. For the decomposition of signals, Gabor dictionary was used. The selection of the appropriate dictionary is important for the proper representation of signals. The other element of HMM is the group of the state dependent observation densities, which were defined as parametric densities. A method was presented to compute these parameters which uses the experimental observations and the output of training of the classifier. The training data was projected on the corresponding linear discriminant classifier planes and statistics were computed from the projections.

An algorithm was presented for the estimation of the next state probability using the model parameters. The algorithm calculates the most probable sequence of states from a sequence of observations and the previous state.

The proposed methods of diagnosis and prognosis were experimentally validated by analyzing faults in the gears of automobile DC starter motors. The selected fault was repetitive and transient. An experimental setup was built and sample data was collected. The working of the algorithm was explained by illustrative examples.

The methods presented are generic, and address the problem of transient fault detection, categorization and prediction with sparse data, using heuristic methods and reasonable assumptions. The test fault is of mechanical nature. However, the sample methodology can be applied for the analysis of electrical faults.

7.2 Future Work

The fault analysis method was developed on the component level. In many applications, system level diagnosis and prognosis is required. The developed prognosis method can be implemented for the system level maintenance by using some polling system or time slot sharing topology. It needs supervisory level routines for such applications.

The selection of the dictionary is of significance in the computation of the state transition probabilities, A dictionary which truly represents the underlying transient phenomenon, can provide accurate probabilities.

The proposed method is validated by analyzing the mechanical faults of DC machines. It is expected that for electrical faults in DC machines and for electrical/mechanical faults in other types of electrical machines, a similar approach is applicable, with suitable modifications. However, electrical faults are generally transient and manifest themselves in high frequencies, raising demand for much higher sampling frequency. The same diagnosis and prognosis algorithm could be applied to electrical fault analysis.

Implementation of the algorithm for online fault diagnosis and prognosis can be

accomplished by using DSP. The algorithm needs to be developed in C or C++ or DSP programming language.

In the presented work, each meshing event of the damaged tooth is categorized. After the initial high current, there are four meshing incidences of the damaged tooth in one sample of the measured motor current. Categorize of the features extracted from the complete motor current sample, after the initial high current, can be an alternate option. Initial results show that the categorization is better if the spectrum energy density analysis of the complete motor current sample is performed. The fault diagnosis and failure prognosis algorithms based on the spectrum energy density analysis can be developed.

Although the algorithms developed in this work were used for analysis of mechanical faults associated with DC machine system, they can be implemented for the fault and failure analysis of induction machines and synchronous machines. However, some modifications in the algorithms might be required and transformations of the measured signals might be suitable to address the issues of power frequencies and frame of references.

The sampling frequency was set to be 20kHz and the sample size was 16 bits. In practical systems, this high sampling frequency and required memory might not be possible. The effects of lower sampling frequency and lesser resolution were discussed in Chapter 4, but it needs to be further explored.

In this work, HMM based algorithm is presented for prognosis. However, the field of fault prognosis in electromechanical systems is relatively new, the question of which signal processing method is preferable, is not settled. A comparative analysis of different prognosis methods in terms of lower or higher cost, more or less accuracy, easier or more difficult to implementation method, might be of interest. The candidate methods can be HMM, extended Kalman, neurofuzzy and particle filters etc.

BIBLIOGRAPHY

[1] B. A. Paya, I. I. Esat, and M. N. M. Badi, "Artificial Neural Network Based Fault Diagnostics of Rotating Machinery Using Wavelet Transforms as a Preprocessor," *Mechanical Systems and Signal Processing*, vol. 11, pp. 751–765, Sep 1997.

[2] S. Bachir, S. Tnani, J.-C. Trigeassou, and G. Champenois, "Diagnosis by parameter estimation of stator and rotor faults occurring in induction machines," *IEEE Transactions on Industrial Electronics,*, vol. 53, no. 3, pp. 963 – 973, june 2006.

[3] R. Beguenane and M. El Hachemi Benbouzid, "Induction motors thermal monitoring by means of rotor resistance identification," in *IEEE International Electric Machines and Drives Conference Record*, 18-21 1997, pp. TD2/4.1 –TD2/4.3.

[4] S.-B. Lee, T. Habetler, R. Harley, and D. Gritter, "An evaluation of model-based stator resistance estimation for induction motor stator winding temperature monitoring," *IEEE Transactions on Energy Conversion,*, vol. 17, no. 1, pp. 7 –15, Mar 2002.

[5] G. Stone, "Advancements during the past quarter century in on-line monitoring of motor and generator winding insulation," *IEEE Transactions on Dielectrics and Electrical Insulation,*, vol. 9, no. 5, pp. 746 – 751, Oct 2002.

[6] W. Zhao-yang and Q. Yu-lin, "Study and realizing of method of ac locating fault in distribution system," in *Asia-Pacific Power and Energy Engineering Conference (APPEEC),*, 28-31 2010, pp. 1 –5.

[7] F. Briz, M. Degner, A. Zamarron, and J. Guerrero, "Online stator winding fault diagnosis in inverter-fed ac machines using high-frequency signal injection," *IEEE Transactions on Industry Applications,*, vol. 39, no. 4, pp. 1109 – 1117, july-aug. 2003.

[8] T. Assaf, H. Henao, and G.-A. Capolino, "Simplified axial flux spectrum method to detect incipient stator inter-turn short-circuits in induction machine," in *IEEE*

International Symposium on Industrial Electronics, vol. 2, 4-7 2004, pp. 815 – 819.

[9] Y.-J. Lee and Y.-H. Ju, "An assessment of insulation condition for generator rotor windings," in *International Conference on Condition Monitoring and Diagnosis, CMD*, 21-24 2008, pp. 543 –545.

[10] W. Zhaoxia, L. Fen, Y. Shujuan, and W. Bin, "Motor fault diagnosis based on the vibration signal testing and analysis," in *Third International Symposium on Intelligent Information Technology Application, IITA*, vol. 2, 21-22 2009, pp. 433 –436.

[11] W. Shuting, L. Yonggang, L. Heming, and T. Guiji, "The new diagnosis method of rotor winding inter-turn short circuit fault and imbalance fault based on stator and rotor vibration characteristics," in *Proceedings of the Eighth International Conference on Electrical Machines and Systems, ICEMS*, vol. 3, 29-29 2005, pp. 2207 –2210 Vol. 3.

[12] A. Sadoughi, M. Ebrahimi, and E. Razaei, "A new approach for induction motor broken bar diagnosis by using vibration spectrum," in *International Joint Conference SICE-ICASE,*, 18-21 2006, pp. 4715 –4720.

[13] C. C. Yeh, B. Mirafzal, R. J. Povinelli, and N. A. O. Demerdash, "A condition monitoring vector database approach for broken bar fault diagnostics of induction machines," in *International Electric Machines and Drives Conference*, May 2005, pp. 29–34.

[14] O. Moseler and R. Isermann, "Application of Model-Based Fault Detection to a Brushless DC Motor," *IEEE Transactions on Industrial Electronics*, vol. 47, pp. 1015–1020, Oct 2000.

[15] X. Q. Liu, H. Y. Zhang, J. Liu, and J. Yang, "Fault Detection and Diagnosis of Permanent-Magnet DC Motor Based on Parameter Estimation and Neural Network," *IEEE Transactions on Industrial Electronics*, vol. 47, pp. 1021–1030, Oct 2000.

[16] M. Dai, A. Keyhani, and T. Sebastian, "Fault Analysis of a Trapezoidal Back-EMF Permanent Magnet Brushless DC Motor Using Finite Element Method," *IEEE Transactions on Energy Conversion*, 2001 (in print).

[17] C. Combastel, S. Lesecq, S. Petropol, and S. Gentil, "Model-Based and Wavelet Approaches to Induction Motor On-Line Fault Detection," *Control Engineering Practice*, vol. 10, pp. 493–509, May 2002.

[18] S. Kia, H. Henao, and G.-A. Capolino, "Digital signal processing for induction machines diagnosis - a review," in *33rd Annual Conference of the IEEE Industrial Electronics Society, IECON*, Nov. 2007, pp. 1155–1162.

[19] C. Kral, T. Habetler, and R. Harley, "Detection of mechanical imbalances of induction machines without spectral analysis of time-domain signals," *IEEE Transactions on Industry Applications,*, vol. 40, no. 4, pp. 1101 - 1106, july-aug. 2004.

[20] A. Cash, T. G. Habetler, and G. B. Kliman, "Insulation Failure Prediction in AC Machines Using Line-Neutral Voltages," *IEEE Transactions on Industry Applications*, vol. 34, pp. 1234–1239, Nov–Dec 1998.

[21] H. Nejjari and M. E. Benbouzid, "Monitoring and Diagnosis of Induction Motors Electrical Faults Using a Current Park's Vector Pattern Learning Approach," *IEEE Transactions on Industry Applications*, vol. 36, pp. 730–735, May–Jun 2000.

[22] A. Bellini, F. Filippetti, C. Tassoni, and G.-A. Capolino, "Advances in diagnostic techniques for induction machines," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 12, pp. 4109–4126, Dec. 2008.

[23] A. Bellini, F. Filippetti, G. Franceschini, C. Tassoni, R. Passaglia, M. Saottini, G. Tontini, M. Giovannini, and A. Rossi, "On-Field Experience With Online Diagnosis of Large Induction Motors Cage Failures using MSCA," *IEEE Transactions on Industry Applications*, vol. 38, pp. 1045–1053, Jul.–Aug. 2002.

[24] W. le Roux, R. G. Harley, and T. G. Habetler, "Rotor Fault Analysis of a Permanent Magnet Synchronous Machine," *International Conference on Electrical Machines*, 2002.

[25] H. S. Liu, B. Y. Lee, and Y. S. Tarng, "Monitoring of Drill Fracture from the Current Measurement of a Three-Phase Induction Motor," *International Journal of Machine Tools and Manufacture*, vol. 36, pp. 729–738, Jun 1996.

[26] J. Altmann and J. Mathew, "Multiple Band-Pass Autoregressive Demodulation for Rolling-Element Bearing Fault Diagnosis," *Mechanical Systems and Signal Processing*, vol. 15, pp. 963–977, Sep 2001.

[27] R. B. Randall, J. Antoni, and S. Chobsaard, "The Relationship between Spectral Correlation and Envelope Analysis in the Diagnostics of Bearing Faults and other Cyclostationary Machine Signals," *Mechanical Systems and Signal Processing*, vol. 15, pp. 945–962, Sep 2001.

[28] D. M. Yang, A. F. Stronach, P. MacConnell, and J. Penman, "Third-Order Spectral Techniques for the Diagnosis of Motor Bearing Condition using Artificial Neural Networks," *Mechanical Systems and Signal Processing*, vol. 16, pp. 391–411, Mar–May 2002.

[29] A. Belšak, M. P. Virtič, B. Aberšek, and J. Flašker, "Vibrations Diagnostics of a Gear Unit," *Key Engineering Materials*, vol. 251–252, pp. 457–462, 2003.

[30] I. Yesilyurt, "The Application of the Conditional Moments Analysis to Gearbox Fault Detection - a Comparative Study Using the Spectrogram and Scalogram," *NDT & E International*, vol. 37, pp. 309–320, Jun 2004.

[31] H. Zheng, Z. Li, and X. Chen, "Gear Fault Diagnosis based on Continuous Wavelet Transform," *Mechanical Systems and Signal Processing*, vol. 16, pp. 447–457, Mar–May 2002.

[32] A. R. Webb, *Statistical Pattern Recognition*, 2nd ed. John Wiley & Sons, 2002.

[33] S. Poyhonen, M. Negrea, A. Arkkio, H. Hyotyniemi, and H. Koivo, "Fault diagnostics of an electrical machine with multiple support vector classifiers," in *Proceedings of the 2002 IEEE International Symposium on Intelligent Control,*, 2002, pp. 373 - 378.

[34] R. R. Schoen, T. G. Habetler, F. Kamran, and R. G. Bartheld, "Motor Bearing Damage Detection Using Stator Current Monitoring," *IEEE Transactions on Industry Applications*, vol. 31, pp. 1274–1279, Nov–Dec 1995.

[35] R. R. Schoen, B. K. Lin, T. G. Habetler, J. H. Schlag, and S. Farag, "An Unsupervised, On-Line System for Induction Motor Fault Detection Using Stator Current Monitoring," *IEEE Transactions on Industry Applications*, vol. 31, pp. 1280–1286, Nov–Dec 1995.

[36] S. F. Farag, R. G. Bartheld, and T. G. Habetler, "An Integrated On-Line Motor Protection System," *IEEE Industry Applications Magazine*, vol. 2, pp. 21–26, Mar–Apr 1996.

[37] M. E. H. Benbouzid and G. B. Kliman, "What stator current processing based technique to use for induction motor rotor faults diagnosis," *IEEE Power Engineering Review,*, vol. 22, no. 8, pp. 62 –62, aug. 2002.

[38] S. Rajagopalan, T. Habetler, R. Harley, J. Restrepo, and J. Aller, "Non-stationary motor fault detection using recent quadratic time-frequency representations," in *Conference Record of the 2006 IEEE Industry Applications Conference, 41st IAS Annual Meeting*, vol. 5, 2006, pp. 2333–2339.

[39] S. Rajagopalan, T. Habetler, R. Harley, J. Aller, and J. Restrepo, "Diagnosis of rotor faults in brushless dc (bldc) motors operating under non-stationary conditions using windowed fourier ridges," in *Conference Record of the 2005 Industry Applications Conference, Fourtieth IAS Annual Meeting*, vol. 1, 2005, pp. 26–33.

[40] S. Rajagopalan, J. M. Aller, J. A. Restrepo, T. G. Habetler, and R. G. Harley, "Detection of rotor faults in brushless dc motors operating under nonstationary conditions," *IEEE Transactions on Industry Applications*, vol. 42, pp. 1464–1477, 2006.

[41] S. Rajagopalan, T. G. Habetler, R. G. Harley, T. Sebastian, and B. Lequesne, "Current/voltage-based detection of faults in gears coupled to electric motors," *IEEE Transactions on Industry Applications*, vol. 42, pp. 1412–1420, 2006.

[42] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, March 1998.

[43] W. J. Wang, "Wavelets for Detecting Mechanical Faults with High Sensitivity," *Mechanical Systems and Signal Processing*, vol. 15, no. 4, pp. 685–696, Jul 2001.

[44] W. G. Zanardelli, E. G. Strangas, H. K. Khalil, and J. M. Miller, "Wavelet-Based Methods for the Prognosis of Mechanical and Electrical Failures in Electric Motors," *Mechanical Systems and Signal Processing*, pp. 411–426, Mar 2005.

[45] W. Zanardelli, E. Strangas, H. Khalil, and J. Miller, "Comparison of wavelet-based methods for the prognosis of failures in electric motors," in *Power Electronics in Transportation*, 2002, pp. 61–67.

[46] W. G. Zanardelli, E. G. Strangas, H. K. Khalil, and J. M. Miller, "The Use of Wavelet Analysis and the Nearest Neighbor Rule for the Prognosis of Failures in Electric Motors," *IEEE International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives*, pp. 591–596, Sep 2001.

[47] W. G. Zanardelli, "The Use of Wavelet Analysis for the Prognosis of Failures in Electric Motors," Master's thesis, Michigan State University, 2000.

[48] W. Zanardelli, E. Strangas, and S. Aviyente, "Identification of intermittent electrical and mechanical faults in permanent-magnet AC drives based on time-frequency analysis," *IEEE Transactions on Industry Applications*, vol. 43, pp. 971–980, 2007.

[49] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms, A Primer.* Prentice Hall, 1998.

[50] L. Cohen, "The scale representation," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3275–3292, Dec 1993.

[51] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Wiley-Interscience, Oct 2000.

[52] J. Cusido, L. Romeral, J. Ortega, J. Rosero, and A. Garcia Espinosa, "Fault detection in induction machines using power spectral density in wavelet decomposition," in *IEEE Transactions on Industrial Electronics*, vol. 55, no. 2, Feb. 2008, pp. 633–643.

[53] J. Rosero, L. Romeral, J. Ortega, and E. Rosero, "Short-circuit detection by means of empirical mode decomposition and wignerville distribution for pmsm running under dynamic condition," in *IEEE Transactions on Industrial Electronics*, vol. 56, no. 11, Nov. 2009, pp. 4534–4547.

[54] J.-R. Ruiz, J. Rosero, A. Espinosa, and L. Romeral, "Detection of demagnetization faults in permanent-magnet synchronous motors under nonstationary conditions," *IEEE Transactions on Magnetics*, vol. 45, no. 7, pp. 2961–2969, July 2009.

[55] S. Kia, H. Henao, and G.-A. Capolino, "Diagnosis of broken-bar fault in induction machines using discrete wavelet transform without slip estimation," *IEEE Transactions on Industry Applications*, vol. 45, no. 4, pp. 1395–1404, July-Aug. 2009.

[56] A. Bellini, A. Yazidi, F. Filippetti, C. Rossi, and G.-A. Capolino, "High frequency resolution techniques for rotor fault detection of induction machines," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 12, pp. 4200–4209, Dec. 2008.

[57] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, pp. 174–188, 2002.

[58] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-gaussian bayesian state estimation," *IEE Proceedings of Radar and Signal Processing*, vol. 140, no. 2, pp. 107–113, Apr 1993.

[59] M. Orchard and G. Vachtsevanos, "A particle filtering-based framework for real-time fault diagnosis and failure prognosis in a turbine engine," in *Mediterranean Conference on Control & Automation*, June 2007, pp. 1–6.

[60] F. Zidani, D. Diallo, M. El Hachemi Benbouzid, and R. Nait-Said, "A fuzzy-based approach for the diagnosis of fault modes in a voltage-fed PWM inverter induction motor drive," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 2, pp. 586–593, Feb. 2008.

[61] G. Zhang, D. Das, R. Xu, and M. Pecht, "IDDQ trending as a precursor to semiconductor failure," in *International Conference on Prognostics and Health Management*, Oct. 2008, pp. 1–7.

[62] Y. Zhang, G. Gantt, M. Rychlinski, R. Edwards, J. Correia, and C. Wolf, "Vehicle design validation via remote vehicle diagnosis: A feasibility study on battery management system," in *International Conference on Prognostics and Health Management*, Oct. 2008, pp. 1–6.

[63] H. Razik, M. de Rossiter Correa, and E. da Silva, "A novel monitoring of load level and broken bar fault severity applied to squirrel-cage induction motors using a genetic algorithm," *IEEE Transactions on Industrial Electronics*, vol. 56, no. 11, pp. 4615–4626, Nov. 2009.

[64] T. Wang, J. Yu, D. Siegel, and J. Lee, "A similarity-based prognostics approach for remaining useful life estimation of engineered systems," in *International Conference on Prognostics and Health Management*, Oct. 2008, pp. 1–6.

[65] F. Rufus, S. Lee, and A. Thakker, "Health monitoring algorithms for space application batteries," in *International Conference on Prognostics and Health Management*, Oct. 2008, pp. 1–8.

[66] X. Lou and K. A. Loparo, "Bearing fault diagnosis based on wavelet transform and fuzzy inference," *Mechanical Systems and Signal Processing*, vol. 18, no. 5, pp. 1077 – 1095, 2004.

[67] W. Q. Wang, M. F. Golnaraghi, and F. Ismail, "Prognosis of machine health condition using neuro-fuzzy systems," *Mechanical Systems and Signal Processing*, vol. 18, no. 4, pp. 813 – 831, 2004.

[68] L. Rabiner and B. Juang, "An Introduction to Hidden Markov Models," in *IEEE ASSP Magazine*, vol. 3, 1986, pp. 4–16.

[69] H. Yun, A. Smith, and H. Silverman, "Speech recognition HMM training on reconfigurable parallel processor," in *Proceedings of the 5th IEEE Symposium on FPGA-Based Custom Computing Machines.* IEEE Computer Society, 1997, p. 242.

[70] B.-W. Min, H.-S. Yoon, J. Soh, Y.-M. Yang, and T. Ejima, "Hand gesture recognition using hidden Markov models," in *IEEE International Conference on Systems, Man, and Cybernetics*, vol. 5, Oct 1997, pp. 4232–4235 vol.5.

[71] L. Heck and J. McClellan, "Mechanical system monitoring using hidden Markov models," in *International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, Apr 1991, pp. 1697–1700 vol.3.

[72] H. Ocak, K. A. Loparo, and F. M. Discenzo, "Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: A method for bearing prognostics," in *Journal of Sound and Vibration*, vol. 302, no. 4-5, May 2007, pp. 951–961.

[73] R. Rammohan and M. Taha, "Exploratory investigations for intelligent damage prognosis using hidden Markov models," in *IEEE International Conference on Systems, Man and Cybernetics*, vol. 2, 2005, pp. 1524–1529.

[74] W. Zhou, D. Chakraborty, N. Kowali, A. Papandreou-Suppappola, D. Cochran, and A. Chattopadhyay, "Damage classification for structural health monitoring using time-frequency feature extraction and continuous hidden Markov models," in *The Forty-First Asilomar Conference on Signals, Systems and Computers*, Nov. 2007, pp. 848–852.

[75] F. Camci and R. Chinnam, "Hierarchical HMMs for autonomous diagnostics and prognostics," in *International Joint Conference on Neural Networks, IJCNN*, 2006, pp. 2445–2452.

[76] L. Atlas, M. Ostendorf, and G. Bernard, "Hidden Markov models for monitoring machining tool-wear," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 6, 2000, pp. 3887–3890.

[77] L. Abdesselam and G. Clerc, "Diagnosis of induction machine by time frequency representation and hidden Markov modelling," in *IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, SDEMPED*, Sep. 2007, pp. 272–276.

[78] P. Baruah and R. B. Chinnam, "HMMs for diagnostics and prognostics in machining processes," *International Journal of Production Research*, vol. 43, pp. 1275–1293(19), March 15, 2005.

[79] L. Cohen, *Time-Frequency Analysis.* Prentice Hall, 1995.

[80] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. Academic Press, 1999.

[81] P. Dutilleux, "An Implementation of the "algorithme à trous" to Compute the Wavelet Transform," in *Wavelets: Time-Frequency Methods and Phase Space, Proceedings of the International Conference.* Springer-Verlag, 1987.

[82] T. Y. Young and T. W. Calvert, *Classification, Estimation and Pattern Recognition.* American Elsevier Publishing Co., Inc., 1974.

[83] S. S. H. Zaidi, W. G. Zanardelli, S. Aviyente, and E. G. Strangas, "Comparative study of time-frequency methods for the detection and categorization of intermittent faults in electrical drives," in *IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, SDEMPED*, 2007, pp. 39–45.

[84] V. N. Vapnik, *The nature of statistical learning theory*, 2000.

[85] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications.* Artech House Publishers, Feb 2004.

[86] A. J. Haug, "A tutorial on bayesian estimation and tracking techniques applicable to nonlinear and non-gaussian processes," MITRE Corporate, Mclean Virginia USA, Tech. Rep. MTR 05W0000004, Jan 2005.

[87] A. Soto, "Self adaptive particle filter," in *International Joint Conference on Artificial Intelligence*, 2005.

[88] A. Smith, A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*, 1st ed. Springer, Jun 2001.

[89] W. Thomson and M. Fenger, "Current signature analysis to detect induction motor faults," *IEEE Industry Applications Magazine*, vol. 7, pp. 26–34, 2001.

[90] R. D. Neelam Mehala, "Motor current signature analysis and its applications in induction motor fault diagnosis," in *International Journal of Systems Applications, Engineering & Development*, vol. II, 2007.

[91] R. Obaid, T. Habetler, and R. Tallam, "Detecting load unbalance and shaft misalignment using stator current in inverter-driven induction motors," in *IEEE International Electric Machines and Drives Conference, IEMDC*, vol. 3, 2003, pp. 1454–1458.

[92] S. Kia, H. Henao, and G.-A. Capolino, "Analytical and experimental study of gearbox mechanical effect on the induction machine stator current signature," *IEEE Transactions on Industry Applications*, vol. 45, no. 4, pp. 1405–1415, July-Aug. 2009.

[93] S. H. Kia, H. Henao, and G.-A. Capolino, "Torsional vibration assessment using induction machine electromagnetic torque estimation," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 1, pp. 209–219, Jan. 2010.

[94] S. Kia, H. Henao, and G.-A. Capolino, "Torsional vibration effects on induction machine current and torque signatures in gearbox-based electromechanical system," *IEEE Transactions on Industrial Electronics*, vol. 56, no. 11, pp. 4689–4699, Nov. 2009.

[95] A. Mohanty and C. Kar, "Fault detection in a multistage gearbox by demodulation of motor current waveform," in *IEEE Transactions on Industrial Electronics*, vol. 53, no. 4, June 2006, pp. 1285–1297.

[96] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," in *IEEE Transactions on Signal Processing*, vol. 41, 1993, pp. 3397–3415.