OPTIMUM MULTI-FAULT CLASSIFICATION OF GEARS WITH INTEGRATION OF EVOLUTIONARY AND SVM ALGORITHMS

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

by

Dhruba Jyoti Bordoloi

(10610303)

Department of Mechanical Engineering

Indian Institute of Technology Guwahati

Guwahati 781 039, India

January 2015

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

CERTIFICATE

It is certified that the work contained in this Thesis entitled **Optimum Multi-Fault Classification of Gears with Integration of Evolutionary and SVM Algorithms** submitted by **Mr Dhruba Jyoti Bordoloi** to the Indian Institute of Technology Guwahati for the award of the degree of Doctor of Philosophy has been carried out under my supervision in the Department of Mechanical Engineering, Indian Institute of Technology Guwahati. This work has not been submitted elsewhere for the award of any other degree or diploma.

January 2015

Dr. Rajiv Tiwari
Professor
Department of Mechanical Engineering,
Indian Institute of Technology Guwahati,
Guwahati– 781039, INDIA

Dedicated to……

My parents

My teachers

&

My family

Acknowledgement

I am indebted to many people who have helped me during the tenure of my doctoral program at IIT Guwahati. I take this opportunity to express my deep gratitude to all of them for their support and help during various phases of my PhD thesis work.

I am highly indebted to my supervisor Prof. Rajiv Tiwari for his guidance and encouragement during the research work. I am grateful to Prof. Tiwari as without his support and advice, it would not have been possible for me to complete my PhD thesis work. He has helped me at various difficult phases of my work with his valuable guidance and moral support. I would like to thank my Doctoral Committee members, Prof. D. Chakraborty (ME), Dr. A. Chakraborty (CE) and Dr. A. Banerjee (ME) for their valuable suggestions and encouragement during annual progress reviews of my PhD thesis work. I am also grateful to the past and present Heads of the Department of Mechanical Engineering, Prof. D. Chakraborty, Prof. P. Mahanta and Prof. A. K. Dass for extending departmental facilities during the tenure of my doctoral program.

I am very much thankful to Mr. Shakun Bansal and Mr. Sachidanand Sahoo for their invaluable help and support during my initial research work and experimentation. I wish to express my sincere thanks to Mr. Rituraj Saikia, Mr. Sanjib Sarma, Mr. Pranjol Paul, Mr. Jiten Basumatary, Mr. Amal Kalita and Mr. Nip Borah for their direct/indirect help during my doctoral program. I also wish to express my sincere thanks to the former PhD students: Dr. Sachin K. Singh, Dr. Mohit Lal and Dr. C. Shravankumar and present PhD students: Mr. Sandeep Singh for their help during various phases of my doctoral program.

I shall always be grateful to my parents and teachers for their love and encouragement. Last but not the least, I thank the members of my family for their constant support, encouragement, patience and sacrifices during this doctoral program.

Finally, I bow my head before God Almighty in deepest gratitude and ask for His blessings.

January 2015

Dhruba Jyoti Bordoloi

IIT Guwahati

Abstract

Of late, the subject of machine condition monitoring as a component of maintenance system became popular worldwide due to the prospective advantages to be gained from reduced maintenance costs, improved productivity, increased machine life time and safer work places. Vibration signals extracted from rotating parts of machineries carry a lot of information within them about the condition of the operating machine. Further processing of these raw vibration signatures, measured at convenient locations of the machine, unravels the condition of components or assemblies under study. Gears in rotating machines are major components of interest for the vibration based condition monitoring. Their failure causes increase in the downtime and the maintenance cost.

Artificial intelligent techniques, like artificial neural networks (ANN), the fuzzy logic system (FLS), the fuzzy neural network and the support vector machine (SVM), have been commonly used in the machine fault diagnosis. However, in the application of SVM for the machine condition monitoring and fault diagnosis still a great deal of work to be explored. The SVM has excellent performance in the generalization, so it can produce high accuracy in the fault classification of the machine condition monitoring and diagnosis. The use of SVM in the expertise and problem orientated domain in machine condition monitoring is growing rapidly.

The main focus of the present thesis is to examine the performance of the multiclass ability of the SVM technique in a gear box by optimizing SVM parameters using the grid-search method (GSM), genetic algorithm (GA) and artificial bee colony algorithm (ABCA). Four fault conditions, i.e. the chipped tooth (CT), the missing tooth (MT) and the worn tooth (WT) along with the normal gear (or no defect, i.e. ND) have been considered.

Time domain and frequency domain vibration signals were collected from the gear box with different fault conditions and at several speeds. The continuous wavelet transform (Mexican hat and Morlet wavelet family) and wavelet packet transform coefficients are extracted from time domain data. The continuous wavelet coefficient are selected based on those have highest value among the coefficients of a range of scale. Similarly, wavelet packet transform coefficients are selected based on those have the maximum energy.

A group of statistical features like the standard deviation, skewness and kurtosis have been extracted from time domain, frequency domain and time-frequency (wavelet) data. The features are divided into two parts; one part is used for optimizing SVM parameters and the other part is used for the fault prediction. The prediction of fault classification has been attempted at the same angular speed as the measured data as well as innovatively at the intermediate and extrapolated angular speed conditions, since it is not feasible to have measurement of data at all speed of interest. Fault predictions are noted in the time, frequency and time-frequency domains.

It is concluded that the SVM has the ability to make perfect classifications if the training data is available for that particular running speed. It is also observed that the prediction accuracy gradually increases with the increase of rotational speed. This is due to the high signal-to-noise level at the high rotation speed due to better manifestation of faults in vibration signals at higher speeds. Fault predictions are better for optimizing by the GA and the ABCA with respect to the GSM technique. The trend of prediction percentage increases with the time, frequency and time-frequency domain features. It is also observed that the Morlet wavelet family showed in an average better prediction.

Finally, to combine fault prediction results from three domains and two SVM classification methods, a voting technique has been proposed. It suggests unified prediction accuracy among the accuracies found out. It is observed that this technique is better than the average predictions obtained from different domains and classification methods, and many time competitive with the best fault prediction results.

Contents

List of Figures

List of Tables

Nomenclature

b	Scalar threshold
c_i	i-th expansion coefficient
e	Vector of all ones
k	Number of classes
$k(x, y)$	Kernel function
q	Probability of selecting the best individual
r	Rank of individual
s	Scale
x	Input vector
y	Indicator vector
w	Boundary
v_i	Weighting factors
v_{ij}	Food position
$[A_A]$	Fault class matrix
$[A_F]$	Fault class matrix using frequency domain data
$[A_O]$	New fault class matrix
$[A_T]$	Fault class matrix using time domain data
$[A_{Wm}]$	Fault class matrix using Mexican hat data
$[A_{Wmo}]$	Fault class matrix using Morlet data
$[A_{Wp}]$	Fault class matrix using wavelet packet data
C	Regularization parameter for C-SVC
C_j	Cumulative probability
E	Entropy

F_i	Fitness value of *i*-th solution
G	Maximum cycle number
K_u	Kurtosis
L	Value of the limit
$P_{initial}$	Randomly generated initial population
P_j	Probability of selection
Q	l by l positive semi-definite matrix
R	Random number
S_d	Standard deviation
S_k	Skewness
T	Translation
T_1	Peak level of a signal
T_2	Root-mean-square value of signal
T_3	Crest factor
α	Optimal for the dual equation
γ	Kernel parameter
$\mu_1 = \overline{x}$	Mean
μ_r	*r*-th moment
ν	Upper bound on the fraction of support vector
ρ	Optimal for the primal equation
σ^2	Variance
ξ	Slack variable

Abbreviations

ABCA	Artificial bee colony algorithm
AI	Artificial intelligent
ANNs	Artificial neural networks
AR	Autoregressive
CBM	Condition-based maintenance
C-SVC	*C*-Support vector classification
CT	Chipped tooth
CV	Cross validation
CWT	Continuous wavelet transform
DAQ	Data acquisition unit
DC	Direct current
DET	Distance evaluation technique
DFT	Discrete Fourier transform
DT	Decision tree
ERM	Empirical risk minimization
ES	Expert system
FFT	Fast Fourier transform
FIT	Fourier integral transform
FNNs	Fuzzy-neural networks
FSE	Fourier series expansion
GA	Genetic algorithm
GMF	Gear meshing frequency
GSM	Grid search method
GST	Grid search technique

HAT	Hospital medical school, London, UK
ICA	Independent component analysis
IEPE	Incorporate integrated electronic piezoelectric
IMF	Intrinsic mode function
KBS	Knowledge-based system
LDA	Linear discriminated analysis
LSI	Large-scale integration
MFS	Machinery fault simulator
MT	Missing tooth
ND	Normal gear or no defect
NFS	Neural-fuzzy systems
PCA	Principle component analysis
PSD	Power spectral density
PSVM	Proximal support vector machines
RBF	Radial basis function
RMS	Root mean square
SRM	Structural risk minimization
SVDD	Support vector data description
SVM	Support vector machine
SVs	Support vectors
UCI	University of California, Irvine
VFD	Variable frequency drive
WPT	Wavelet packet transform
WT	Worn tooth
ν-SVC	ν-Support vector classification

CHAPTER 1

Introduction and Literature Survey

1.1 Introduction

Rotating machineries are very common and widely used in the contemporary industrial world. For example, steam turbines, compressors, pumps, motors and jet engines are the most known and commonly used rotating machines. The breakdown of rotating machines may result in monetary losses and even worse, in the fatality of human beings. That is why damage diagnostics and the maintenance of the rotating machinery during the operation is a very important. It is one of the most difficult engineering tasks on which the durability and safety of machine operation depend. In the world of industrial mechanization and automation, many modern plants have installed flexible computer-controlled automatic and unmanned equipments; and their maintenance cost has been increased substantially. Therefore, maintenance has been historically regarded as a necessary evil by various management functionaries.

Nowadays, the role of maintenance is swinging from a “compulsory evil” to a “revenue contributor” and in the direction of a “partner” of organizations to achieve world-class competitiveness (Waeyenbergh and Pintelon, 2002). Now the maintenance concepts have modified from the breakdown maintenance to the preventive maintenance, then to the condition-based maintenance and more recently to the reliability base maintenance. The *breakdown maintenance* is the earliest form of maintenance, where no measures are taken to maintain the equipment until it breaks. It can reinstate the functional operation of failed components by repairing the defect or replacing them with new ones. However, a random stoppage of equipment can limit the machine usage capacity, and has severe impacts on productivity and product quality. Consequently, the breakdown maintenance often results in the high equipment downtime, high cost of restoring equipment, high penalties associated

with the loss of production, high spare part inventory level, and extensive unscheduled repair time (Tsang, 1995).

To prevent disastrous failures and emergency shutdowns, the *preventive maintenance* was introduced in the 1950s. Periodic intervals for the machine inspection and maintenance regardless of the machine's health condition are done in the periodic maintenance. This helps to prevent functional failures by replacing critical components at regular intervals before the end of their expected useful lives. Even though the preventive maintenance reduces the frequency of unplanned breakdown and increases the reliability of equipment, it is costly due to frequent replacements of expensive components before the end of their useful lives and the reduction of the availability of equipments. Furthermore, it may create other unrelated failures due to the removal and replacement of parts, which is inappropriate for specific tolerances and human errors. And because of it the *condition-based maintenance* (CBM) is important. The CBM includes diagnostic and prognostic modules, and it attempts to monitor machinery health based on condition measurements that do not interrupt normal machine operations. It is based on the actual condition and can assess whether the equipment is in need of maintenance or not; and if need arises, finds when the maintenance measures require to be performed. Moreover, by incorporating the prognosis, an alarm level can be set when the predicted values and the actual fault symptom of failure fall within the warning region. This will provide adequate time for system operators to take corrective actions and inspect the condition of the equipment, and conduct a repair on the defect before the catastrophic failure occurs.

Three key components of the CBM (Jardine *et al.,* 2006) are the *data acquisition* (i.e. the collection and storage of machine health information*), data processing* (i.e. the conditioning and feature extraction/selection of acquired data), and *decision making* (i.e. the recommendation of maintenance actions through diagnosis and/or prognosis). The diagnosis

and prognosis are two important aspects in a CBM system. The *diagnosis* is the ability to detect fault, isolate and identify which component has failure, and decide on the potential impact of failed component on the health of system. The *fault detection* is a task to indicate whether something is going wrong in the monitored system; the *fault isolation* is a task to locate the component that is faulty; and the *fault identification* is a task to determine the nature of the fault when it is detected. The *prognosis* is the capability to use available observations to predict upcoming states of machine or forecast the fault before it occurs.

The *data acquisition* is a process of collecting and storing useful data (information) from targeted physical assets (Jardine *et al.*, 2006). The condition monitoring data may be the vibration data, acoustic data, oil analysis data, temperature, pressure, moisture, humidity, weather or environment data, etc. Various sensors, such as micro-sensors, ultrasonic sensors, acoustics sensors, etc. have been utilized to collect different type of data (Austerlitz, 2003; Kirianaki *et al.,* 2002). Data collections fall into three categories like the value type (specific time epoch and single value) for example the oil analysis data, temperature, pressure, humidity, etc.; the waveform type (specific time epoch and time waveform) for example the vibration data, acoustic data, etc.; and the multi-dimensional type (specific time epoch and multi-dimensional) for example the image data (X-ray image, visual image or video, etc.).

The *data processing* can be divided into sub-groups like the feature extraction, the feature selection and the machine fault diagnosis (i.e., the training and testing of data). The feature extraction means transforming existing features into a lower dimensional space (Yang *et al.*, 2005). The data may be a waveform or value type. The waveform data are analyzed in three different forms: time domain, frequency domain and time-frequency domain.

The *time domain analysis* calculates features from time waveform signals as descriptive statistics such as the mean, peak, peak-to-peak interval, standard deviation, crest factor, root mean square (RMS), skewness, kurtosis, etc. The *mean* indicates the average value of a signal and is termed as the mean value. The *standard deviation* is a measure of the effective energy or power content of the vibration signal. The *kurtosis* indicates the flatness or the spikiness of the signal. Its value is very low for good condition and high for faulty condition. The *skewness* is the degree of asymmetry of a distribution around its mean, which means lack of symmetry. The *crest factor* is equal to the peak amplitude of a waveform divided by the RMS value.

The *frequency domain analysis* is based on the transformed signal into frequency domain. The advantages of frequency domain over time domain are its ability to easily identify and isolate certain frequency component of interest. The *time–frequency analysis* investigates waveform signals in both time and frequency domain, and has been developed for non-stationary waveform signals. Traditional time–frequency analysis uses time–frequency distributions, which represent the energy or power of waveform signals in two-dimensional functions of both time and frequency to better reveal fault patterns for more accurate diagnostics.

The *value type data* includes both the raw data obtained via the data acquisition and feature values extracted from raw signals via the signal processing. The value type data look much simpler than the waveform and image data. The feature of this data can be extracted by linear analysis techniques such as principle component analysis (PCA), linear discriminated analysis (LDA), and independent component analysis (ICA). The PCA is a statistical technique that linearly transforms an original set of variables into a substantially smaller set of uncorrelated variables that represents most of the information in the original set of variables (Jolliffe,

1986). The ICA is a technique that transform multivariate random signal into a signal having components that are mutually independent in complete statistical sense (Widodd *et al.,* 2007).

In above discussions, it has been attempted to clarify the importance of the CBM and its different components. Literatures on the machine fault diagnosis and prognosis is huge and diverse primarily due to a wide variety of systems, components and parts. On consideration of the limitation of the thesis the discussion is concentrated to the machine fault diagnosis only.

1.2 Machine Fault Diagnosis

Machine fault diagnosis is a procedure of mapping the information obtained in the measurement space and/or features in the feature space to machine faults in the fault space (Jardine *et al.*, 2006). Several methods have been proposed in order to solve the fault detection and fault diagnosis problems. The most commonly employed solution approaches for the fault diagnosis system include (a) *model-based*, (b) *knowledge-based*, and (c) *pattern recognition-based approaches* (Venkatasubrsmanian, 1944). Generally, analytical model-based methods can be designed in order to minimize the effect of unknown disturbance and perform the consistent sensitivity analysis; knowledge-based methods are used when there is a lot of experience but not enough details to develop accurate quantitative models; and pattern recognition methods are applicable to a wide variety of systems and exhibit real-time characteristics. The application of these methods to the machine fault diagnosis is reviewed as follows.

Model-based methods perform fault diagnosis that is relied on the analytical redundancy in which the consistency between measurements and expected behavior of the process is checked by analytical models. These analytical models could be physical specific or explicit

mathematical model of the monitored machine. Based on this explicit model, residual generation methods such as Kalman filter, system identification, and parity relations are used to obtain signals - so called *residuals* which are indicative of fault presence in the machine (Jardine *et al.*, 2006). Finally, the residuals are evaluated to achieve the fault detection, isolation and identification.

Generally, model-based approaches can be more effective if a suitable and accurate model is built. Simple and fast models such as space vector models may not be very accurate for monitoring purposes whereas detailed models, e.g. based on finite elements take too much computation time. However, explicit mathematical models may not be feasible for complex systems since it would be very difficult or even could not be established for such systems.

A knowledge-based system (KBS) or expert system (ES) for the fault diagnosis is performed based upon the evaluation of on-line monitored data according to a rule set which is determined by the expert knowledge. This knowledge includes locations of input and output process variables, patterns of abnormal process conditions, fault symptoms, operational constraints, and performance criteria. The operators' and engineers' intelligence related to the specific process systems can be implemented into this approach. Their knowledge can help to recognize potential faults based on previous experiences. This approach can reduce difficulties on the exact numeric information and automates the human intelligence for the process supervision.

Compared with model-based approaches, knowledge-based approaches are particularly suitable for large industrial plants since those non-linear real plants are extremely difficult to model and the linear approximation of the model results in large errors. In addition, knowledge-based approaches are able to reduce the complexity when implementing the

corresponding safety system, and make it flexible, easy to understand and follow. Combining knowledge-based fault diagnosis methods with real-time process variables monitoring will definitely improve the efficiency and reliability of detecting fault behavior and overall effectiveness of the system.

The *pattern recognition* is the assignment of a label to a given input value. Traditionally, the *pattern recognition* is manually done by auxiliary graphical tools such as the power spectrum graph, phase spectrum graph, cepstrum graph, autoregressive spectrum, spectrogram, and wavelet phase graph. However, the manual pattern recognition requires expertise in the specific area of diagnosis application. Thus, highly trained and skilled human resources are necessary. Therefore, automatic pattern recognition is extremely desirable. This can be achieved by classification of signals based on the information and feature extracted from signals.

Pattern recognition approaches, which include artificial intelligent (AI) techniques, have been increasingly applied to machine diagnosis and have shown superior performance over conventional approaches. However, it is not easy to apply AI techniques in practice due to the lack of efficient procedures for obtaining training data and specific knowledge that are necessary for training the models. Thus far, most of the applications in the literatures just used experimental data for training models. Popular AI techniques for machine diagnosis are artificial neural network (ANNs), fuzzy logic systems, fuzzy-neural networks (FNNs), and neural-fuzzy systems (NFS). A review of recent developments in applications of AI techniques for the induction machine stator fault diagnostics was provided by Siddique *et al.* (2003).

Hill and Baincs (1988) discussed the design of an expert system (ES) for the measured data. The ES is a computer program that is able to process the input data to diagnose a problem. The artificial intelligent system has been increasingly applied to the machine condition diagnosis, and has been shown its ability over conventional approaches. In the condition monitoring practice, knowledge from domain specific experts is usually inaccurate and reasoning on knowledge is often imprecise. Therefore, measures of uncertainties in knowledge and reasoning are required for ES to provide more robust problem solving. Unremarkably used uncertainty measures are probability, fuzzy member functions in fuzzy logic theory and belief functions in belief networks theory.

An example of applying the fuzzy logic to the machine fault classification was given in Mechefske (1998) to classify frequency spectra representing various faults in rolling element bearings. Du and Yeung (2004) introduced an approach, so-called the fuzzy transition probability, which combined transition probability with the fuzzy set, to monitor progressive faults. The fuzzy logic is also incorporated with other techniques such as neural networks and ES for the fault diagnostic application. For example, Zhang *et al.* (2003) developed an FNN for the fault diagnosis of rotary machines to improve the recognition rate of the pattern recognition, especially in the case that the sample data are similar. Lou and Loparo (2004) employed an adaptive neural-fuzzy inference system as a diagnostic classifier for the bearing fault diagnosis. Liu *et al*. (1996) applied the fuzzy logic and ESs to build a fuzzy ES for the bearing fault detection. Chang *et al*. (1995) built a system for decision-making support in a power plant using both rule-based ES and fuzzy logic. Genetic algorithms (GAs), which are the most ordinarily used type of evolutionary algorithms (EA), have been applied to machine diagnostics. Examples of ANN incorporating GA and other EAs for machine fault classification and diagnostics reported by several researchers (Chen *et al.* 2003; Huang and Huang, 2002; Yan and Ma, 2004). A technique called the *support vector machine* (SVM) is a

new general machine learning tool based on the structural risk minimization principle. It has received much consideration in recent times due to its high accuracy and good generalization capabilities. The use of SVM and its extension for the machine fault diagnosis were summarized in a review paper (Widodo and Yang, 2007).

In the review paper by Widodo and Yang (2007), they discussed how the SVM has gained popularity in the machine learning. They indicated that though numerous methods had been developed based on intelligent systems such as the ANN, fuzzy expert system, condition-based reasoning, and random forest. However, the use of SVM for the machine condition monitoring and fault diagnosis is still rare. The SVM has excellent performance in generalization so it can produce high accuracy in the classification for machine condition monitoring and diagnosis. Based on literatures until 2006, they concluded that the SVM in the machine condition monitoring and diagnosis field was tending to develop towards expertise-orientated and problem-oriented domain. Finally recommended, the ability to continually change and obtain a new novel idea for machine condition monitoring and diagnosis using SVM would be future works.

The above discussion has been related to different machine fault procedures and their trends based on available literatures. At the end the discussion is concluded with the use of the SVM in recent scenario for the fault detection and diagnostics. Though the SVM method has been used in many situations, however, still there is much more scope to explore its use for machine faults diagnostics.

1.3 Feature-based diagnosis

The SVM performs the machine condition monitoring and diagnosis using its excellence ability in the classification process. Based on the input data vectors that consist of

representation of the fault in machine, the SVM recognizes these patterns. Usually, each fault produces special features that can be considered as patterns. So the main task of SVM here is to recognize and classify these patterns as accurate as possible related to faults. The SVM is motivated to represent patterns in a high dimension typically much higher than the original feature space. With an appropriate nonlinear mapping using a kernel function to a sufficiently high dimension, data from two or more categories can always be separated by a hyperplane (Duda, 2001).

The primary desire of combining neural networks with fuzzy systems is to extend knowledge based expert systems by an adaptive component which realizes learning like capability. The learning component has to facilitate the bi-directional knowledge exchange between the neural network and fuzzy system in order to enable a permanent adaptation and tuning of the rule base with less effort (Ayoubi and Isermann, 1997).

For a good classification, data pre-processing is a very important step. A good data pre-processing will reduce the noise in the data and retains as much information as possible (Bishop, 1995). From a clean data, features (condition indicators) can be calculated and considered as patterns for fault diagnosis purposes. In the classification process for the fault diagnosis, when the number of objects in the training set is too small for the number of feature used, most of classification procedures cannot find good classification boundaries. This is called *curse of dimensionality* (Duda, 2001). Moreover, by a good pre-processing, the number of features per object can be reduced such that the classification problem can be solved satisfactorily. The fault diagnosis routine that uses data represented as feature is called *feature-based diagnostic*. Recently, the feature-based diagnostic procedure has been employed for the fault diagnosis of machine. Researchers who used features method for the fault diagnosis are reported now.

Tax *et al.* (1999) employed the feature-based procedure from the power spectrum, envelope spectrum, autoregressive modelling, music spectrum and classical spectrum for the failure detection of a small submersible pump. They tried to find the best representation of data features such that the target class can best be distinguished from the outlier class. The support vector data description (SVDD) was proposed to accomplish their work for finding the smallest sphere containing all target data.

Other authors used statistical features based on moments, cumulants and other statistical features of the time data series and the spectral of vibration data for the fault detection are reported (Jack and Nandi, 2002; Samanta *et al.*, 2003; Samanta, 2004; Xu and Wang, 2005; Ren *et al.*, 2005; Sugumaran *et al.*, 2007; Hu *et al.*, 2007). Yang *et al.* (2000, 2004, 2005, 2007) used statistical features of time and frequency domain for the fault detection in the rotating machinery and cavitations of the butterfly valve. In the case of induction motor, they acquired data of the vibration and the stator current signal. Yuan and Chu (2006, 2007) performed the fault diagnosis of the turbo-pump rotor using data features that were acquired from frequency bands of the secondary vibration signal. The frequency of secondary signal was divided into 9 bands then the frequency amplitudes on each band and their average value were calculated as features. Sun *et al.* (2004) employed statistical features, which comes from the acoustic emission signal for the wear detection in the machine tool. They also used the cutting parameter such as the cutting speed, depth of cut and feed rate as additional features. Cho *et al.* (2005) carried out the tool break detection using features from cutting forces and power consumption in the end milling machine. The other application was reported by Han *et al.* (2004), and they conducted the hot spot detection in the power plant using features from the data temperature, which were acquired by thermocouple and infrared sensors. Moreover, Ramesh *et al.* (2003) conducted a prediction of the thermal error in machine tools using features from temperature sensors.

In the feature-based diagnosis process, after defining features (statistical features) from original data, a huge dimensionality problem of features occurs. It cannot be avoided because of not all features are useful and optimal for the classification process. The existence of irrelative features tends to degrade the performance of classifier. One of solutions to resolve this problem is by performing the feature extraction, which can extract optimal features and all at once reduce the dimensionality of features. Basically, the feature extraction means mapping process of data from higher dimension into low dimension space. Many methods have been proposed to perform dimensionality reduction using the linear and nonlinear techniques as reported by Fodor (2002). In the machine condition monitoring and fault diagnosis research area, the feature extraction using the component analysis was reported as follows by using linear methods: the principal component analysis (PCA) (Widodo *et al*., 2006, 2007; and Yuan and Chu, 2007), and using independent component analysis (ICA) (Widodo *et al*., 2006, 2007). Moreover, nonlinear feature extraction using the kernel PCA and kernel ICA was also performed by Widodo and Yang (2007). The other techniques called rough set theory (RST) was conducted for extracting optimal features and reduce dimension of features by Xu *et al.* (2005) and Fang (2006). In their research, RST was employed to pre-process the data for eliminating redundant information and reducing the sample dimension.

Moreover, some experts suggested use of the feature selection after defining features set from the original data. The techniques, which are addressed to the feature selection, are the genetic algorithm (GA) and the distance evaluation technique (DET). In machine fault diagnostics area, researchers who employed the GA technique were Jack and Nandi (2002), Samanta *et al.* (2003, 2004), and Li *et al*. (2005). In addition, Hu *et al*. (2007) and Yang *et al*. (2004, 2005) were reported successful performances of feature selections using the DET.

From aforementioned discussions, it can be observed that the SVM as the feature-based diagnosis method have been widely used in many applications of machine condition monitoring and fault diagnosis. Most of results from the feature-based technique have been relatively satisfactory according to papers reviewed. It means that the feature-based procedure is a recommended method when the recognition and classification process are performed. The research is still continuing on different machine components for its faults diagnosis using the SVM. On consideration of the focus of the present thesis, the review of literatures is concentrated only on the gearbox fault diagnosis in the following section.

1.4 Gear Box Condition Monitoring and Fault Diagnosis

A faulty gear system could result in serious damage if defects occur to one of gears during operating condition. Early detection of defects, therefore, is crucial to prevent the system from malfunction that could cause severe damage or entire system halt. The major causes for gear failure can be an error in design or an application error or it is likely that there is a manufacturing error or any combination of the above. Design errors may be due to causes like improper gear geometry, use of wrong materials, quality, lubrication and other specifications. Application errors can be due to problems like vibration, mounting and installation, cooling and maintenance, while manufacturing errors can be in the form of inaccuracies in machining or improper heat treatments. Conventional methods for processing the measured data contain the time domain, frequency domain and time-frequency domain techniques. These methods have been widely employed for gear failure detection also. The most commonly used method for rotating machines is the vibration analysis. Measurements can be taken on machine bearing casings with seismic or piezo-electric transducers to measure casing vibrations, and on the vast majority of critical machines, with eddy current transducers that directly observe rotating shafts to measure the radial (and axial) vibration of the shaft. The first basic principle

of vibration monitoring is that a change in the force and/or mass-stiffness-damping relationship will be reflected in the vibration signature of the machine.

The vibration condition monitoring as an aid to fault diagnosis was examined by several researchers. Smith (1980) showed how the vibration characteristics can be used to recognize rotor faults. Stewart (1976) and Taylor (1995) also included the information of the actual data analysis process - how the measured data should be processed for the diagnosis. Downham (1976) gave a broad but a comprehensive coverage of advances in malfunction diagnosis in rotary machinery by vibration analysis; he also discussed developments during past five decades. Thomas (1984) outlined the vibration monitoring strategy for a 500MW turbo generator. He also discussed about the financial benefit from such a scheme. Smalley *et al.* (1996) presented a method to assess the severity of vibration (in terms of probability of damage by the analysis of vibration signals) and its related cost using the net present value method.

It is generally thought that vibration is a symptom of a gearbox condition. Vibration generated by gearboxes is complicated in its structure but gives a lot of information. It may be said that vibration is a signal of a gearbox condition. The use of vibration based gear fault detection technique is well established in industry (Cameron & Stuckey, 1994, Taylor, 1995, Leblane *et al*., 1990). The vibration analysis can be carried out using Fourier transform techniques like Fourier series expansion (FSE), Fourier integral transform (FIT) and discrete Fourier transform (DFT) (Collacott, 1977). After the development of large-scale integration (LSI) and the associated microprocessor technology, the fast Fourier transform (FFT) analyzers became cost effective for general applications. It is well known that the most important components in gear vibration spectra are the gear meshing frequency (GMF) and its harmonics, together with sidebands due to modulation phenomena. The increment in the number and amplitude of such

sidebands may indicate a fault condition. Moreover, the spacing of the sidebands is related to their source. The raw signatures acquired through a vibration sensor needed further processing and classification of the data for any meaningful surveillance of the condition of the system being monitored.

Samanta (2004) carried out the gear fault detection using the SVM combined with the GA. The time-domain vibration signal of a rotating machine with the normal and defective gears were processed for the feature extraction. Extracted features from the original signal were used as inputs to the SVM classifier. In this research, the GA was performed in the feature selection and optimizing the radial basis function kernel parameters.

Xuan et al. (2005) used the power spectral density (PSD) of vibration signals of the gearbox casing to construct spectral features. The GA was used to reduce the feature dimension from original signals. The classifier was based on SVMs with the multi-class classification ability. Prediction results showed the effectiveness of the approach. Liu *et al.* (2005) proposed a weighted SVM with the GA based parameter selection for the binary SVM parameter selection. The experimental analysis showed that the method was feasible and efficient. Rojas and Fernandez-Reyes (2005) described a GA approach for adjusting multiple parameters in SVM kernels with weighted RBF kernels using the HAT (Hospital Medical School, London, UK) proteomic dataset. Hung and Wang (2006) optimized the *C*-SVC parameters and feature subset simultaneously, without degrading the SVM classification accuracy with the GA based approach. The proposed method performs the feature selection and the parameter setting in an evolutionary way. They used eleven real-world datasets from UCI (University of California, Irvine) database.

Saravanan *et al.* (2008) deals with the effectiveness of wavelet-based features for fault diagnosis using the SVM and proximal support vector machine (PSVM) algorithms. The statistical feature vectors from Morlet wavelet coefficients were classified using the decision tree (DT) and predominant features were fed as the input for the training and testing the SVM and PSVM algorithms, and their relative efficiency in classifying faults in the bevel gear box was compared. They concluded that the PSVM had slightly better classification capability than the SVM. In 2009, Saravanan *et al.* used statistical features from vibration signals, and features that discriminate different fault conditions of the gearbox were selected using the DT. The rule set for the fuzzy classifier was obtained using the DT. A fuzzy classifier was built and tested with representative data. Saravanan *et al.* (2010) subsequently presented the effectiveness of wavelet-based features for the fault diagnosis of a gear box using the ANN and the PSVM. The statistical feature vectors from Morlet wavelet coefficients were classified using DT and predominant features were fed as the input for the training and testing of the ANN and the PSVM. Their relative efficiency in classifying faults in the bevel gear box was compared and found that the average efficiency of the ANN was more compared with the PSVM.

Cheng *et al.* (2008) decomposed vibration signals into a finite number of intrinsic mode functions, then the autoregressive (AR) model of each intrinsic mode functions (IMF) component was established. Finally, the corresponding AR parameters and the variance of remnant were regarded as the fault characteristic vectors and used as input parameters to the SVM classifier to classify the working condition of gears. They analyzed results of three kinds of gear vibration signals among which one was normal and the other two were the gear with a crack and the gear with a broken tooth as faults, respectively. They concluded that the gear fault diagnosis approach based on the IMF and AR model with the SVM could be applied to classify the gear fault effectively even in case of smaller number of samples.

Tiwari *et al.* (2009) presented an approach based on the SVM technique to detect and classify multiple gear-fault conditions. In this work, samples of vibration signals of gearbox casing were recorded under various monitoring conditions and formulated a procedure for classifying the data. Two approaches for feature extractions were adopted for analyses in frequency domain. Cases of 0% errors during classifications were found in some instances. Overall classification efficiency of the SVM was compared with the reported efficiency of ANN and one case of 100% classification as opposed to 93% using ANN was found. Ali *et al.* (2009) extracted features from the vibration signal by using the signal analysis in time, frequency and time-frequency domain. These features were taken from statistical characteristics of signals in different domains. To remove fault-irrelevant features, a feature selection technique, the Euclidian distance, was used to pick five superior features. Experimental results showed the approach to detect abnormalities of the gearbox. At the same time, the category and severity of faults were indicated accurately. Gao *et al.* (2010) presented a method that combined the wavelet lifting, SVM and rule-based reasoning to diagnose gearbox faults. Gearbox vibration signals were initially processed by the wavelet packet decomposition. Then, energy coefficients of each frequency band were calculated and used as input vectors to the SVM to recognize the normal and faulty gearbox patterns. The precise analysis from the wavelet lifting scheme was then utilized to obtain the machine fault feature frequency. Finally, based on the fault feature frequency, the existing diagnostic knowledge and rules (fuzzy) were used for logical reasoning to establish a knowledge base to identify fault types.

Zamanian and Ohadi (2010) used the exact wavelet analysis to minimize effects of overlapping and distortion in the case of gearbox faults. The gearbox was considered in both the healthy and chipped tooth gear conditions. The SVM with the RBF was used in extracted features from the exact wavelet analysis for the classification. The efficiency of this classifier was then evaluated with other signals acquired from the test setup. They claimed that in

comparison with the continuous wavelet transform, the exact wavelet transform had the better ability in the feature extraction in lieu of more computational effort. Samadzadegan *et al.* (2010) indicated the potential of artificial intelligence methods for optimization of *C*-SVM parameters (GA-based approach). The proposed method was compared with the grid search method, a traditional method for the parameter setting, by conducting experiments using different benchmark data sets.

The present literature review on the machine condition monitoring and diagnosis focuses on to the gearbox using the support vector machine (SVM). It surveys articles using a keyword index on the machine condition monitoring and machine fault diagnosis using the SVM. Since SVM has gained popularity in machine learning; however, researchers who applied SVM in machine condition monitoring and diagnosis are now gradually increasing because of its effectiveness. Until 2006, it was observed that the SVM in the machine condition monitoring and diagnosis was tending towards development of the expertise-orientation and problem-oriented domain. The ability to continually change and obtain a new novel idea for machine condition monitoring and diagnosis using the SVM is yet to be explored in detail. After the review of literatures on the machinery fault analysis, following points are summarized

- Machinery encounters several faults like unbalance, mechanical looseness, misalignment, cracks in shaft, oil whirl/whip, electrical problem, faulty bearing and faulty gearbox, and these have been addressed in detail by several researchers.
- The gearbox in a machinery system is a crucial component, and early and accurate detection of faults of a gear box prevents the system from severe damage or entire system halt. The main focus of the present literature survey was centered on gearbox only.

- Several types of faults in gearbox have been considered in literatures like chipped tooth, missing tooth, worn tooth, etc.
- Normal gearbox vibration signals have dominance of periodic signals. However, faulty gearbox vibration signals have contribution of both the periodic and random signals. Time, frequency and time-frequency domain have been used to capture the dynamics of faulty gearbox.
- For satisfactorily handling the signals carrying non-stationary or transient components, the time-frequency domain signal has been used. In which the study of time varying spectra of the gearbox can be studied due to different fault conditions.
- Expert systems (ES) can be developed for the robust fault classification with the use of extracted features from vibration signal. Several machine learning algorithms have been used, for example the ANN, SVM, fuzzy logic, nero-fuzzy and fuzzy-SVM.
- Though the SVM is built primarily for the binary classification, it may used in the multi-fault classification of gear box. The prediction of multi-faults in gearbox is still rare.
- Indices (RMS value, peak level value, and crest factor) and overall vibration level do not provide any diagnostic information but may have limited application in fault detection in simple safety of critical accessory components. Statistical moments (for example 1^{st} moment is the mean, 2^{nd} moment is the variance and so on) of vibration signals are capable to identify fault conditions.
- All the statistical features are not equally significant for fault classifications. Few are effective, which can be selected by different feature selection techniques (for example the distance evaluation and decision tree techniques).

- Discussions and implementations of different version of SVM algorithms (*C*-SVC and ν-SVC) are limited except of the *C*-SVC. Different versions of SVM algorithm may have different classification performances for variety of faults in the gearbox.
- SVM algorithms contain several parameters to be chosen appropriately during training phase. Selections of optimum SVM parameters are one of the crucial components for better predictions.
- Different optimization techniques have been utilized for the optimization of SVM parameters. There are few literatures available on the GA based optimization of *C*-SVC parameters. There is a still scope to optimize parameters of different versions of the SVM by optimization tools (Grid search method, GA, ABCA, etc.).
- Different optimization tools have different algorithms to reach a converging point. Due to inappropriate selection of bounds and constraints, they may approach to a local convergence. To find satisfactory optimal results a comparison study is necessary among the different optimal values. The GA and GSM are already reviewed in the literature survey.
- Vibrations of gearbox are very much dependent on the load level and speed. The fault classification in the gearbox has been attempted by researchers for different level of loading conditions and at different speed conditions.
- The fault classification attempts have been attempted basically at the same rotational speed at which the training and testing have been performed. The prediction of faults at other than rotation speeds, for which features are not available i.e. at interpolated and extrapolated speeds, has not been discussed in literatures at all.
- Most of literatures attempted fault classifications based on the time, frequency and time-frequency domain data, individually. An integrated approach for fault

classification with all three domain data is expected to give more reliable and efficient predictions.

From the above discussions and observations the motivation of the present work has been formed. The aim and objective of this work described briefly in the following section.

1.5 Aim and Objective of the Present Work

Based on observations and discussions of previous section from available literatures, it could be concluded that among the different machine elements gear is one of crucial elements that need more investigations on multi-fault classifications. The classification of gear faults based on vibration is very common. Statistical features could be extracted from the vibration signature in the time, frequency and time-frequency domains. The SVM has still a great potential and very few attempts have been made for multi-fault clarification for gears. The main objectives of the present investigation are as follows:

- Fault conditions of gear would be simulated in a gearbox with the introduction of faulty pinions with faults comprising of the cheeped tooth, missing tooth, worn tooth and a healthy pinion.
- The comparison of classification ability of two versions of the SVM classifier (i.e. *C*-SVC and ν-SVC) would be attempted for the gearbox.
- SVM parameters would be optimized by the genetic algorithm (GA), artificial bee colony algorithm (ABCA) and conventional grid search method (GSM) for confirming the near optimum predictions.

- The prediction of classifier would be demonstrated at the same rotational speed as well as at the interpolation and extrapolation speeds, at which vibration data features are not available.
- The calculation of overall prediction accuracies would be demonstrated from the prediction results found from the different domain and classifiers.

1.6 Organization of the Thesis

In the present thesis, Chapter 1 gives an introduction & literature survey on the condition based maintenance and faults diagnosis. Different aspects of the condition based system, with reference to expert system and feature generation by highlighting the gearbox fault identifications, has been explained in this chapter. Literatures related to features based fault classification and fault classifiers have been presented. The SVM classifier and its classification performed by different researchers with focus on the gearbox fault identification are compiled in this chapter. Finally, scopes of work have been presented and the present work is drawn from that.

In Chapter 2, the experimental set-up is discussed and different faulty gears are presented. Details of the data acquisition system hardware used and the software implemented are described. The collection of experimental vibration data from the experimental set-up is presented.

In Chapter 3, details of the classification methodology (SVM) used for the classification are elaborated. Optimization techniques (grid search method, genetic algorithm, artificial bee colony algorithm) used for optimization of the classifier parameters is discussed. Finally, the algorithm used for the implementation of the optimized classifier for the classification of faults is discussed.

In Chapter 4, the basis of selection of statistical features is discussed first. Next, the statistical features are generated from the time domain vibration data. Fault classifications of gearbox faults at the same rotational speed, and at the interpolation and extrapolation rotational speeds using time domain features are elaborated. Prediction accuracies of fault results are presented and discussed.

In Chapter 5, the generation of frequency domain statistical features from the FFT data are discussed. The classification ability of the classifier using frequency domain statistical features is presented at the same rotational speed as well at the interpolation and extrapolation rotational speeds. The comparative study of time domain and frequency domain results is presented.

In Chapter 6, the generation of time-frequency (wavelet) domain data from time domain data are discussed here. The procedure for calculation of continuous wavelet transform (CWT) and wavelet packet transform (WPT) coefficients from time domain vibration data are discussed in detail. The procedure also explains statistical features extracted from the chosen wavelet CWT and WPT coefficients. The ability of classifier using these wavelet statistical features is demonstrated.

In Chapter 7, results on fault classification obtained from the time, frequency and time-frequency domains are compared and discussed. Finally, a novel procedure is presented by combining overall prediction results of three domains.

The overview of the present work, major conclusions and contributions, applicability and limitations, and future scopes of work are provided in Chapter 8. References and publications from the present work are included at the end of the thesis.

CHAPTER 2

Experimental Setup and Data Acquisition

2.1 Introduction

Experimental setup for a bevel gearbox is described and instrumentation for measurement of vibration is presented. The procedure of vibration data collection from the healthy and faulty gears for fault classifications has been presented in the present chapter.

2.2 Experimental Set-up and Experimentation

Experiments were performed on a Machinery Fault Simulator™ (MFS). This could be used for the simulation of a range of machine faults like in the gearbox, shaft misalignments, rolling bearing damages, resonances, reciprocating mechanism effects, motor faults, pump faults, etc. The actual set-up of the MFS used for the experiment with instrumentation and a schematic diagram is shown in Figure 2.1 and Figure 2.2 and it consisted of the following units

1. Experimental setup (Make Spectra Quest, USA)
2. Tri-axial accelerometer (Make Meta Mess-und Frequenztechnik, Germany)
3. Constant DC power source unit (Make Scientific MES-Technik PVT. LTD., India)
4. Data acquisition system (Make National Instruments, USA)

These are described in detail in subsequent subsections.

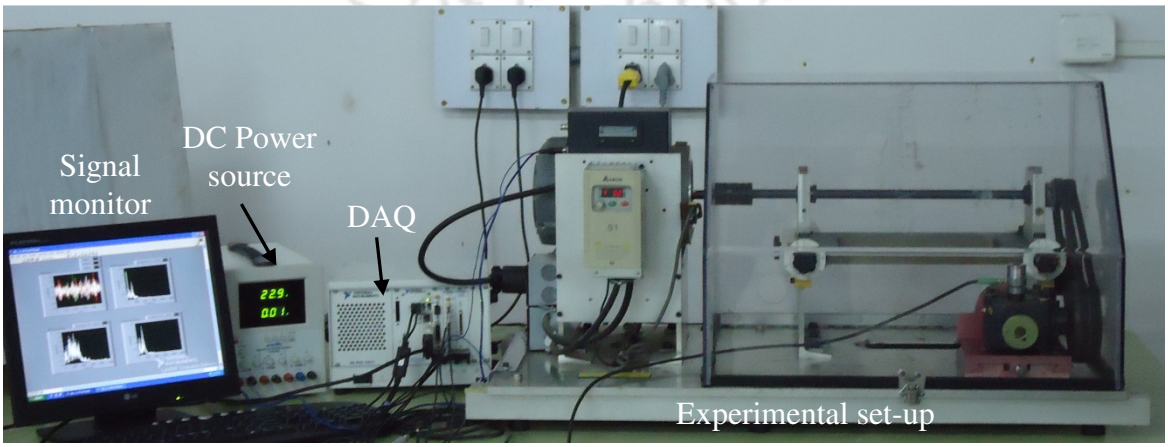


Figure 2.1 Experimental set-up with instrumentation

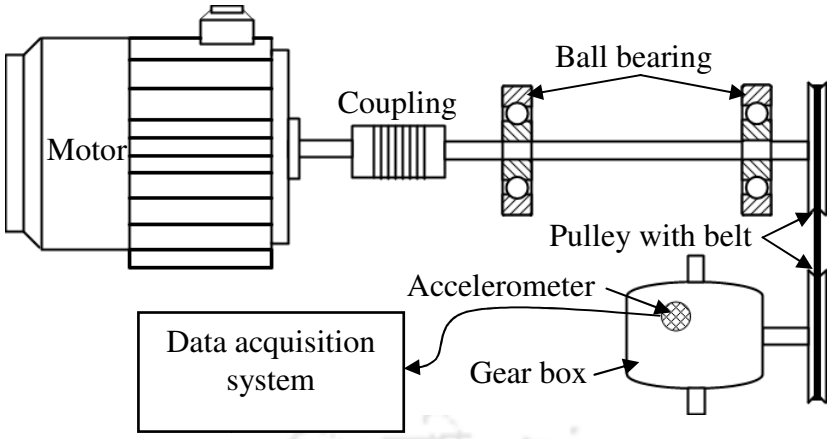


Figure 2.2 A schematic diagram of the experimental set-up

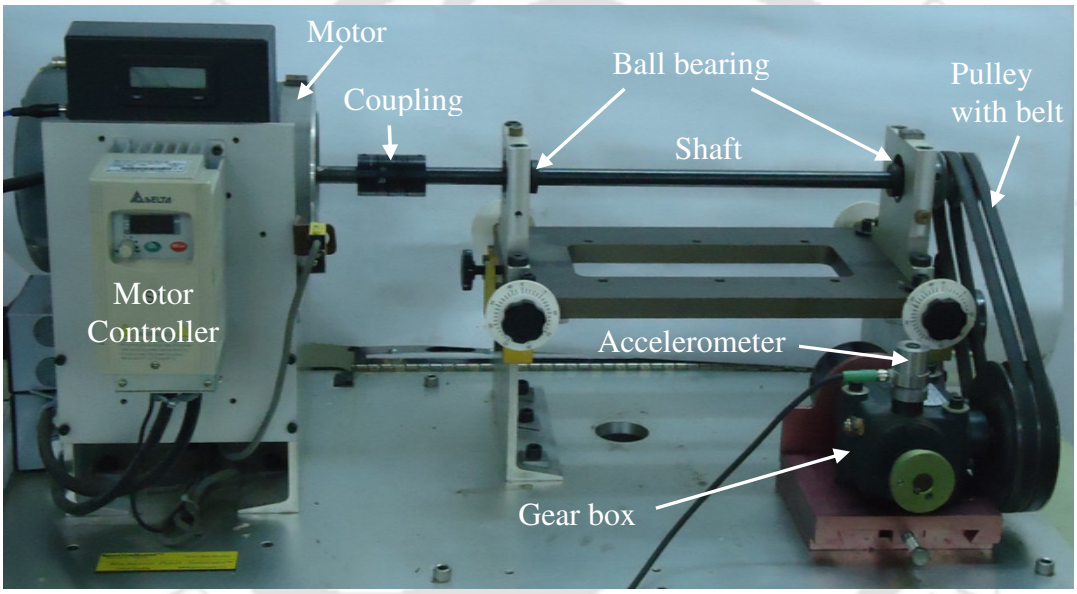


Figure 2.3 Experimental set-up on MFS

2.2.1 Experimental Setup

In the MFS experimental setup (illustrated in Figure 2.3), 3-phase induction motor was connected to a rotor through flexible coupling. The rotor was mounted on two bearings. Bearing housings were rigidly mounted on a rigid plate and with channels in turn it was mounted on a rigid base plate. Through a pulley and belt mechanism the rotor was connected to the gearbox. The motor speed and, hence, the gearbox speed could be controlled by a variable frequency drive (VFD). One photovoltaic sensor (as illustrated in Figure 2.4) was

mounted near the coupling, connecting the motor shaft with the main shaft, to measure the rotational speed of the shaft. The rotational speed is not the same of the rotational speed of the motor spindle and a reduction of speed of 2.6 times achieved with the help of a pulley and belt arrangement.

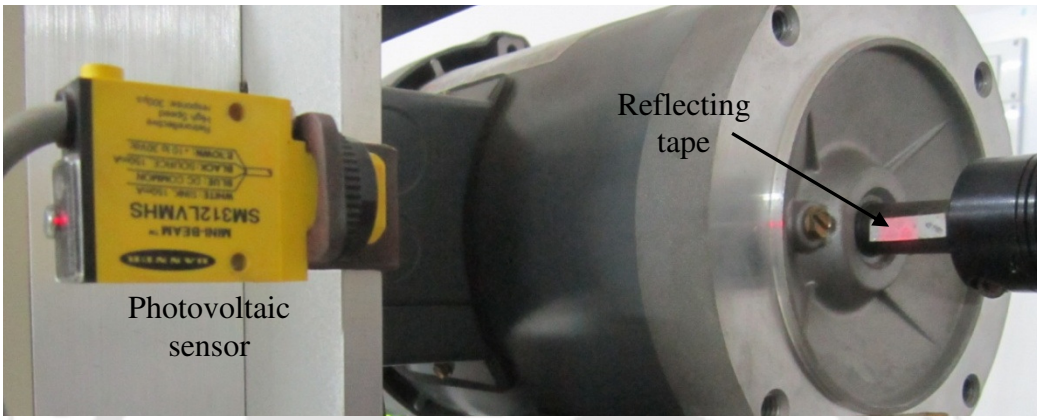


Figure 2.4 A photovoltaic sensor and a reflecting tape

The gear box and its assembly were mounted on the base plate and are illustrated in Figure 2.5. The gear-box was consisted of single-stage bevel gears with 18 numbers of teeth on the pinion and 27 numbers of teeth on the gear. The shaft connected with the pinion was used as driving shaft (input shaft) by means of connecting the main rotor shaft by belt and pulley arrangement. The other end of the bevel gear was the output shaft. In the study of faults in gears, three different types of faulty pinion gears namely the chipped tooth (CT), missing tooth (MT) and worn tooth (WT) along with a normal gear (or no defect i.e., ND) were used (illustrated in Figure 2.6). The normal pinion could be interchanged with other faulty pinions and mounted in the appropriate position in the gearbox. The gearbox was lubricated for smooth running.

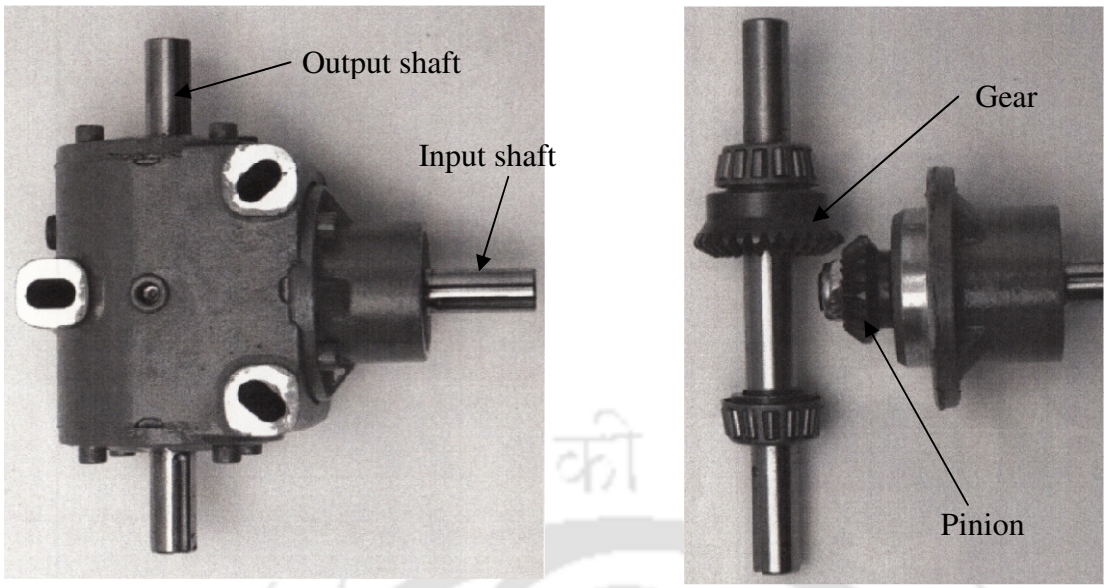


Figure 2.5 The gearbox and its assembly (left) gear box (right) pinion and gear pair

(a) (b) (c)

Figure 2.6 A range of of bevel gear with (a) chipped tooth, (b) missing tooth and (c) worn gear

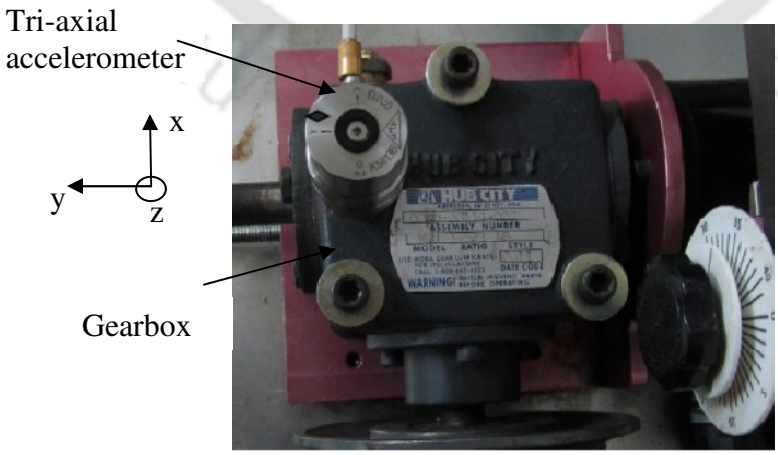


Figure 2.7 An accelerometer mounted on the gearbox

2.2.2 Tri-axial Accelerometer

A real time data in time domain was measured using a tri-axial accelerometer (sensitivity: *x*-axis 100.3 mV/g, *y*-axis 100.7 mV/g, *z*-axis 101.4 mV/g) mounted on the top of the gearbox housing (as illustrated in Figure 2.7). The accelerometer was connected to the data–acquisition hardware.

2.2.3 Constant DC Power Source Unit

The photovoltaic sensor (illustrated in Figure 2.4) required DC power supply of the range ±24 V/2A to operate and it was provided through a constant DC power supply. The constant DC power supply source offered high-resolution, high-power voltage and current outputs for the present set up. The constant DC power supply source is illustrated in Figure 2.8.

Figure 2.8 A constant DC power source

2.2.4 Data Acquisition System

The data-acquisition hardware consisted of National Instruments make NI PXI-8108 Core 2 Duo 2.53 GHz Controller with NI PXI-4472, 8 channels, 24 bits, 102.4 kS/s sampling rate for the accelerometer module. These are mounted in a single PXI chassis which are illustrated in

Figure 2.9. The controller of the data acquisition unit consists of the input and output device connection facilities, like the key board, mouse, monitor, USB devices, LAN connectors jack and hard drive. The controller runs on the Windows XP operating software. The NI PXI-4472 is an 8-channel dynamic signal acquisition module for making high-accuracy frequency-domain measurements. Input channels incorporate integrated electronic piezoelectric (IEPE) signal conditioning for accelerometers and microphones. The eight PXI-4472 input channels simultaneously digitize input signals over a bandwidth from DC to 45 kHz.

The accelerometer was connected by wire with the PXI-4472 in three different channels for three direction measurement. The NI LabVIEW data acquisition software was pre-loaded in the system (PXI-8108) and it was programmed for collecting the real time data from accelerometer in time and frequency domain.

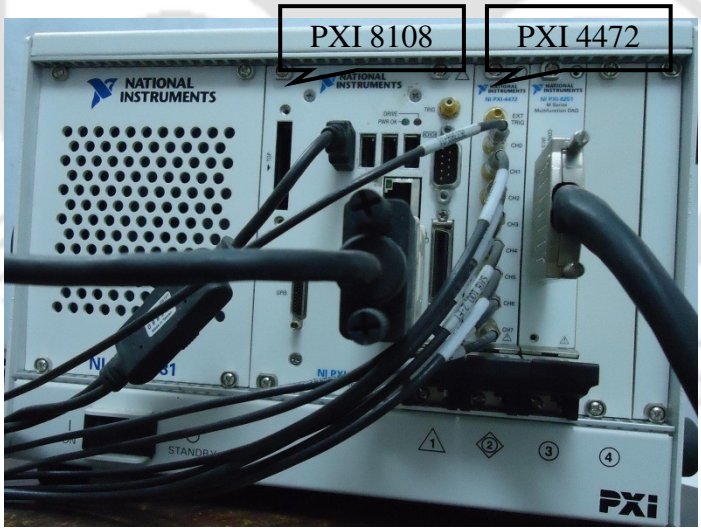


Figure 2.9 PXI-8108 with PXI-4472 data acquisition unit

2.3 Measurement Procedure

Measurements were taken for the rotational speed of 10 Hz to 30 Hz in intervals of 2.5 Hz for each of four fault conditions. For a measurement set in time domain 2,000 samples were

collected at the rate of 20,000 samples per second. In the similar way 300 sets of readings were taken. Altogether, 2000×300 data (sample) points were collected for each of three directions (i.e., *x*, *y* and *z* directions). The data-acquisition system had facility to convert time domain signal to frequency domain also. For a measurement set in frequency domain, 1,000 samples were collected. Altogether, 1000×300 data points were collected for each of the three orthogonal directions. The data were stored in the hard drive of the system (PXI-8108) at individual speeds and for various gear faults in the form of Microsoft excel file.

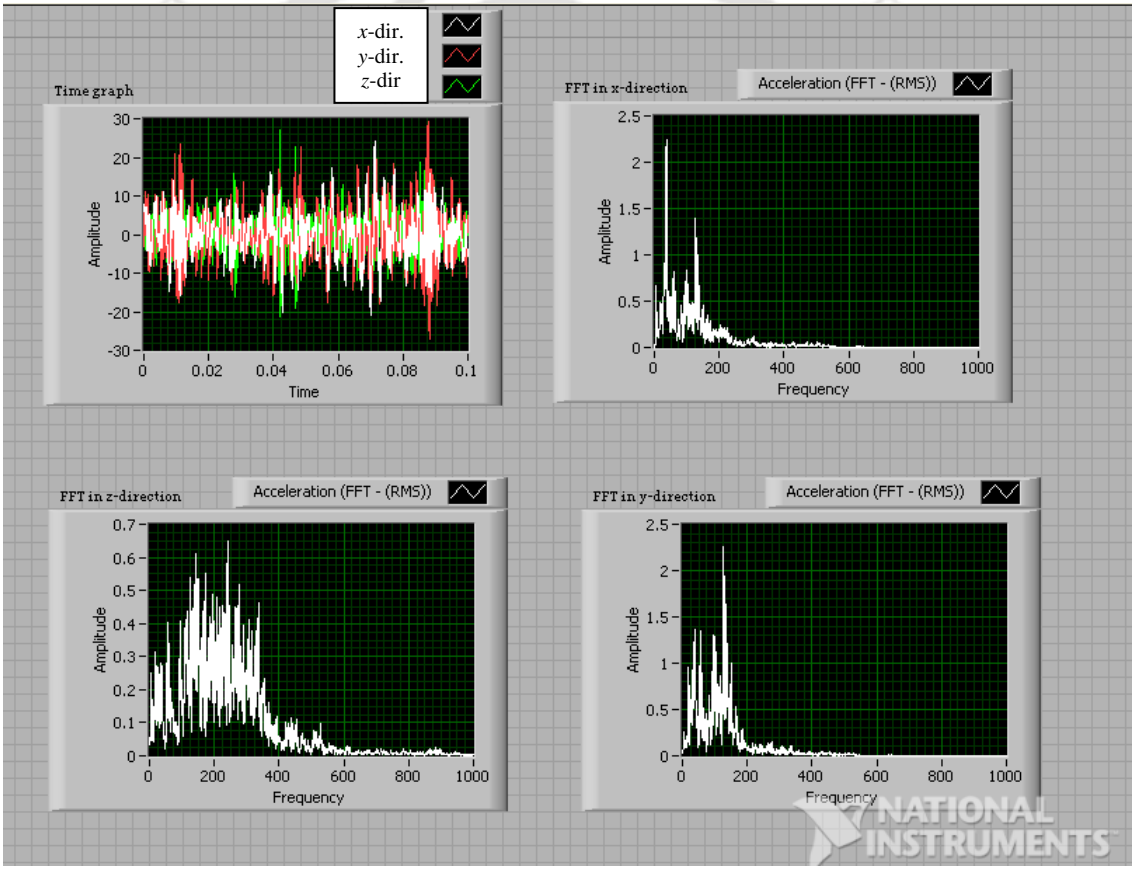
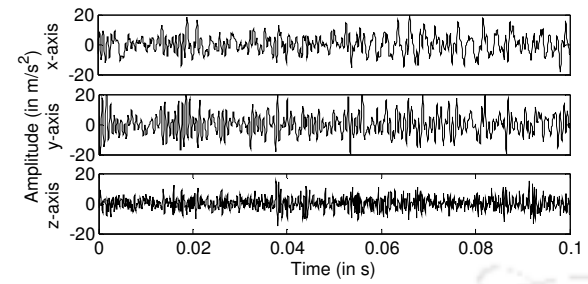


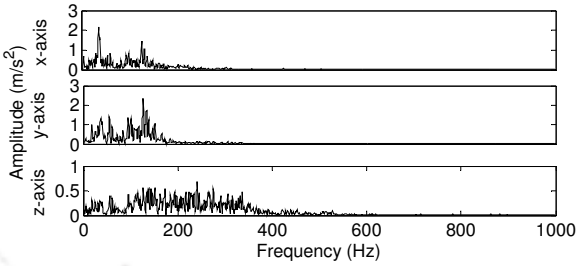
Figure 2.10 A typical screenshot of LabVIEW data collection (time in s and frequency in Hz)

Figure 2.10 illustrates a typical screen shot of the data collection by LabVIEW. Four windows in the graph indicate all time domain signals in three orthogonal directions, the FFT in *x*, *y* and *z*-axis direction, respectively. The respective graphs are specified on the right hand top

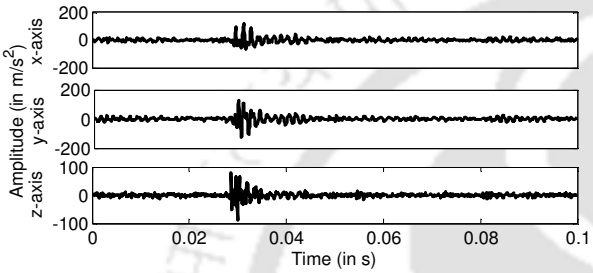
corner in the screenshot. The time domain graph with three different colors (white, red and green) indicates the *x*, *y* and *z*-axis direction response.



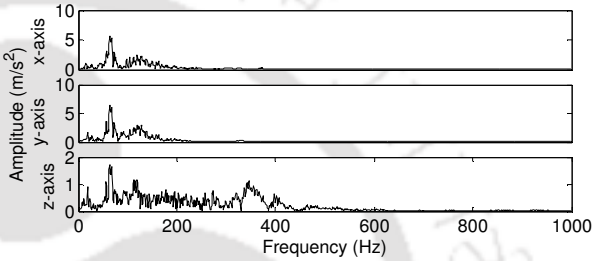
(a) CT case in time domain



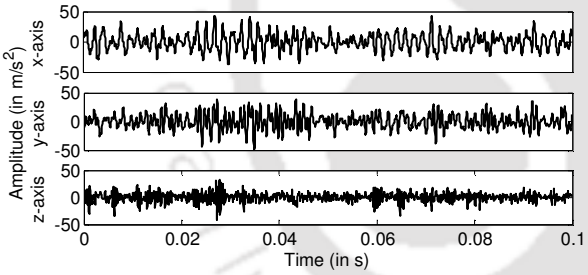
(e) CT case in frequency domain



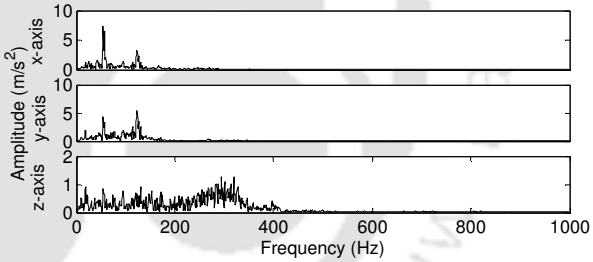
(b) MT case in time domain



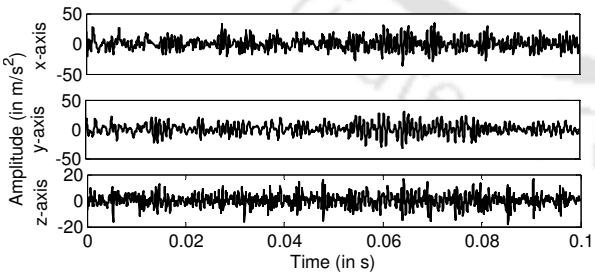
(f) MT case in frequency domain



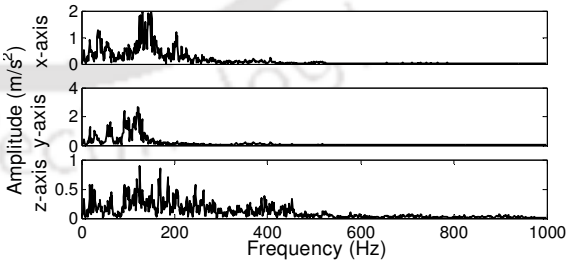
(c) WT case in time domain



(g) WT case in frequency domain



(d) ND case in time domain



(h) ND case in frequency domain

Figure 2.11 Responses in three orthogonal directions at 30 Hz rotational speed in (a)-(d) time domain and in (e)-(h) frequency domain

Figure 2.11 (a)-(d) illustrates the typical plots for time and Figure 2.11 (e)-(h) illustrates the frequency domain responses at 30 Hz rotational speed for the CT gear fault.

Excel files were downloaded from the hard drive and copied into the PC where the classification code was running. Respective excel files were read by the MATLAB code for calculation of statistical features.

2.4 Summary

Experimental data collection procedure was presented in the present chapter for the faulty and no-faulty gears at different rotational speed. A laboratory test rig was identified and experiment was performed. The faulty and healthy gear pinions were interchanged into the gear box for the generation of vibration signatures. The gear box is rotated by a motor by means of a belt and pulley arrangement. Vibration signatures were measured by a tri-axial accelerometer that was mounted on the top of the gear box. The vibration data in the time and frequency domains were collected through a DAQ at different rotational speeds and stored for further processing. The classification methodology of SVM and optimization of its parameters will be discussed in Chapter 3. The procedure of classification of data (features) will also be included in the following chapter.

CHAPTER 3

Fault Classification Methodology by the SVM

3.1 Introduction

The data set prepared from gearbox in the previous chapter will be used for the fault classification. The classification methodology depending upon the SVM classifier will be discussed in this chapter. Two variations of the SVM classifier will be utilized for the classification. There is a scope to tune SVM parameters for optimizing fault classification prediction accuracies. Procedures will be presented for optimizing SVM parameters by two optimizing tools, i.e. genetic algorithms (GA) and artificial bee colony algorithm (ABCA). Along with these the conventional grid search technique (GST) will be discussed for the tuning of SVM parameters.

3.2 SVM Classifier

Support vector machines (SVMs) are a set of related supervised learning methods that analyze data and recognize patterns, and it is used for the classification and regression analyses. The original SVM algorithm was invented by Vapnik (1995) and the current standard incarnation (soft margin) was proposed by Cortes and Vapnik (1995).

The basic SVM deals with only the binary classification. The data are separated by a *hyper plane* defined by a number of *support vectors*. The standard SVM takes a set of input data, and predicts, for each given input, which of two possible classes the input is a member of, which makes the SVM a *non-probabilistic binary linear classifier*. Since the SVM is a classifier, then given a set of training examples, each marked as belonging to one of two categories, the SVM training algorithm builds a model that predicts whether a new example falls into one category or the other. Intuitively, the SVM model is a representation of the examples as points in space, mapped so that examples of the separate categories are divided

by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on (Burgees, 1998).

The classical learning approaches are designed to minimize the error on the training data set and it is called the *empirical risk minimization* (ERM). Whereas the SVM is based on the *structural risk minimization* (SRM) principle rooted in the statistical learning theory. It gives better generalization abilities and the SRM is achieved through a minimization of the upper bound of the generalization error. This will also benefit in faults classification, because the number of features to be the basis of fault diagnosis may not have to be limited. Also, SVM based classifier is claimed to have good generalization properties compared to conventional classifiers. It is because in training the SVM classifier the so-called structural misclassification risk is to be minimized, whereas traditional classifiers are usually trained so that the empirical risk is minimized.

In Figure 3.1, a series of points for two different classes of data are shown, i.e. black circles (class I) and opaque circles (class II). The SVM attempts to place a linear boundary (dotted line) between the two different classes, and orient it in such a way that the margin (represented by solid lines) is maximized. The SVM tries to orient the boundary such that the distance between the boundary and the nearest data point in each class is maximal. The boundary is then placed in the middle of this margin between the two points. The nearest data points are used to define the margins and are known as *support vectors* (SVs, represented by small black circle for class I and small opaque circle for class II). Once the support vectors are selected, the rest of the feature set can be discarded, since the SVs contain all the necessary information for the classifier. The boundary can be expressed as follows

$$(w_{\bullet}x)+b=0, \quad w\in R^{n}, \quad b\in R \tag{3.1}$$

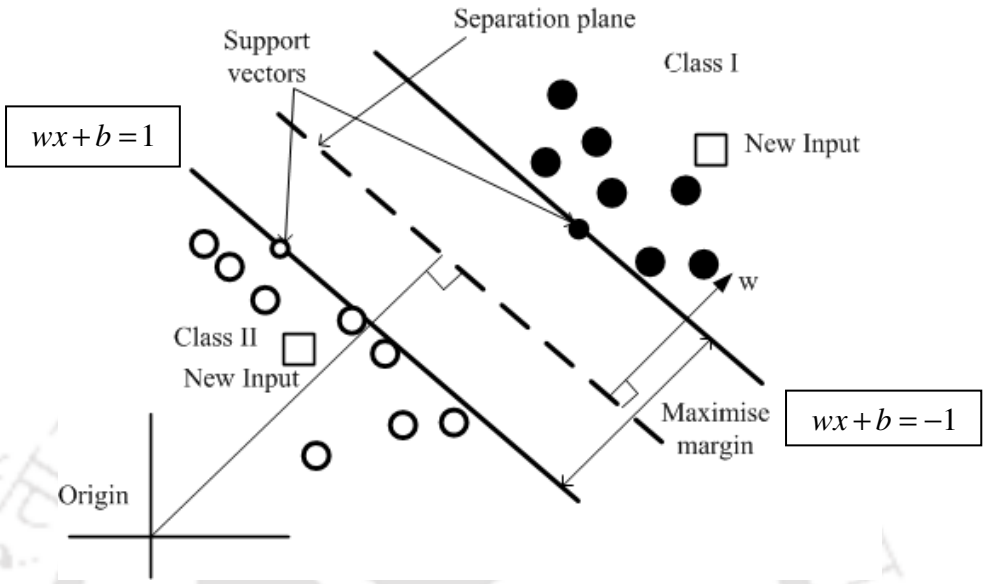


Figure 3.1 Classification of data by linear SVM

where the vector *w* defines the boundary, *x* is the input vector of dimension *n*, and *b* is a scalar threshold dot '•' represents the dot product between two vectors. At the margins, where the SVs are located, equations for class I and class II, respectively, are as follows

$$(w_{\bullet}x)+b=1 \quad \text{and} \quad (w_{\bullet}x)+b=-1 \tag{3.2}$$

As SVs correspond to extremities of the data for a given class, the following decision function can be used to classify any data point in either class A or B

$$f(x)=sign\{(w_{\bullet}x)+b\} \tag{3.3}$$

The optimal hyperplane separating the data can be obtained as a solution to the following optimization problem

$$\min \quad \tau(w) = \frac{1}{2}\|w\|^2 \tag{3.4}$$

Subjected to $y_i\{(w_\bullet x_i) + b\} \geq 1, \quad$ with $\quad i = 1,\ldots,l$

where l is the number of training sets. The solution of the constrained optimization problem can be obtained as follows

$$w = \sum v_i x_i \tag{3.5}$$

where x_i are SVs obtained from training and v_i is used as weighting factors. Putting Eqn. (3.5) into Eqn. (3.3), the decision function is obtained as follows

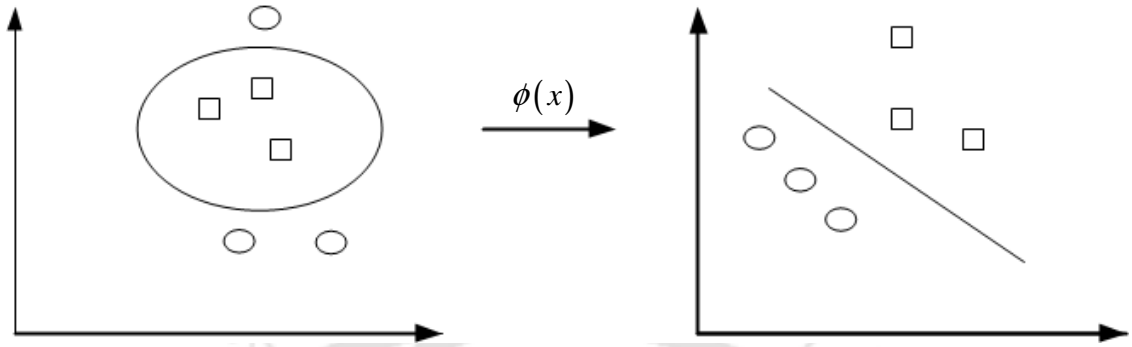
$$f(x) = sign\left\{\sum_{i=1}^{l} v_i (x_\bullet x_i) + b\right\} \tag{3.6}$$

In cases where the linear boundary in input spaces will not be enough to separate two classes properly, it is possible to create a hyper plane that allows a linear separation in the higher dimension (corresponding to a curved surface in the lower dimensional input space). In SVMs, this is achieved through the use of a transformation $\phi(x)$ that converts the data from an N-dimensional input space to Q-dimensional feature space

$$s = \phi(x) \tag{3.7}$$

where $x \in R^n$ and $s \in R^Q$.

Figure 3.2 shows the transformation from the *input space* to the *feature space*, where the non-linear boundary is transformed into a linear boundary in the feature space.



(a) Input space (b) Feature space

Figure 3.2: Non-linear separation of the input and feature spaces

Substituting the transformation Eqn. (3.7) into Eqn. (3.6), it gives the decision function as

$$f(x) = sign\left[\sum_{i=1}^{l} v_i \{\phi(x) \bullet \phi(x_i)\} + b\right] \tag{3.8}$$

The transformation into the higher-dimensional feature space is relatively computation-intensive. A kernel can be used to perform this transformation and the dot product in a single step provided the transformation can be replaced by an equivalent kernel function. This helps in reducing the computational load and at the same time retaining the effect of higher-dimensional transformation. The kernel function $k(x, y)$ is defined as follows

$$k(x, y) = \phi(x) . \phi(y) \tag{3.9}$$

The basic form of the SVM is accordingly obtained after substituting Eqn. (3.9) in the decision function Eqn. (3.8) as follows

$$f(x) = sign\left\{\sum_{i=1}^{l} v_i k(x, x_i) + b\right\} \tag{3.10}$$

Parameters v_i are used as weighting factors to determine which of input vectors are actually support vectors $(0 < v_i < \infty)$. There are different kernel functions like the polynomial, sigmoid and radial basis function (RBF) used in SVMs.

In the present discussion, we consider two SVM variations the *C*-support vector and the ν-support vector for classifications. The formulation of optimization technique for the optimum classification will be also discussed.

***C-Support Vector Classification* (C-SVC):** The *C*-SVC primal binary optimization (Boser *et al.*, 1992; Cortes and Vapnik, 1995) problem can be formulated for a training data $x_i \in R^n, i = 1,\ldots,l,$ and an indicator vector $y \in R^l$ such that $y_i \in \{1,-1\}$ as

$$\min_{w,b,\xi} \quad \tfrac{1}{2} w^T w + C\sum_{i=1}^{l} \xi_i \tag{3.11}$$

Subject to $y_i\left\{w^T \phi(x_i) + b\right\} \geq 1 - \xi_i$ with $\xi_i \geq 0, i = 1,\ldots,l,$

where $\phi(x_i)$ function maps x_i into a higher-dimensional space, w is the weight vector, b is the bias, ξ is the slack variable, and $C>0$ is the regularization parameter. The dual of the above Eqn. (3.11) can be formulated as

$$\min_{\alpha} \quad \left(\tfrac{1}{2}\alpha^T Q\alpha - e^T\alpha\right) \tag{3.12}$$

$$\text{Subject to} \qquad y^T\alpha = 0, \qquad \text{with} \quad 0 \le \alpha_i \le C, \qquad i = 1,\ldots,l,$$

where $e = \{1,\ldots,1\}^T$ is the vector of all ones, Q is an l by l positive semi-definite matrix, $Q_{ii} \equiv y_i y_j K(x_i, x_j)$, and $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$ is the kernel function. In the present discussion, we basically used the radial basis function (RBF): $k(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right), \gamma > 0$. Here γ is a kernel parameter. For solving the primal-dual relationship, the optimal, w, could be written as

$$w = \sum y_i \alpha_i \phi(x_i) \tag{3.13}$$

The decision function is

$$\text{sgn}\left\{w^T\phi(x) + b\right\} = \text{sgn}\left\{\sum_{i=1}^{l} y_i \alpha_i k(x_i, x) + b\right\} \tag{3.14}$$

Here $y_i\alpha_i \vee i, b$ are called as the *label name* (+1 or -1) and the *support vector*, respectively.

ν-Support Vector Classification: In the ν-support vector classification (Scholkopf *et al.*, 2000) a new parameter $\nu \in (0,1)$ is introduced. It has been proved that ν an upper bound on the fraction of support vectors. The binary primal problem can be written for a training data $x_i \in R^n, i = 1,\ldots,l$ and an indicator vector $y \in R^l$ such that $y_i \in \{1,-1\}$ as

$$\min_{w,b,\xi,\rho} \left(\tfrac{1}{2} w^T w - \nu\rho + \tfrac{1}{l} \sum_{i=1}^{l} \xi_i \right) \tag{3.15}$$

$$\text{Subject to} \quad y_i \left\{ w^T \phi(x_i) + b \right\} \geq \rho - \xi_i \quad \text{with} \quad \xi_i \geq 0, \quad i = 1,\ldots,l, \quad \rho \geq 0$$

The dual problem of the above Eqn. (3.15) can be written as

$$\min_{\alpha} \quad \tfrac{1}{2} \alpha^T Q \alpha \tag{3.16}$$

$$\text{Subject to} \quad 0 \leq \alpha_i \leq 1/l, \quad i = 1,\ldots,l, \quad \text{with} \quad e^T \alpha \geq \nu, \quad y^T \alpha = 0,$$

where $Q_{ij} = y_i y_j k(x_i, x_j)$. Chang and Lin (2001) showed that above equation is feasible if and only if

$$\nu \leq \frac{2\min(\text{number of } y_i = +1, \text{number of } y_i = -1)}{l} \leq 1,$$

So the usable range of ν is in between 0 and 1. The decision function can be written as

$$\operatorname{sgn}\left\{ \sum_{i=1}^{l} y_i \alpha_i k(x_i, x) + b \right\} \tag{3.17}$$

Crisp and Burges (2000) and Chang and Lin (2001) showed that $e^T\alpha \geq \nu$ can be replaced by $e^T\alpha = \nu$. Numerically α_i may be very small due to a constraint $\alpha_i \leq 1/l$. A scaled version of Eqn. (3.16) can be written as

$$\min_{\alpha} \quad \left(\tfrac{1}{2}\bar{\alpha}^T Q\bar{\alpha}\right) \tag{3.18}$$

$$\text{Subject to} \quad 0 \leq \bar{\alpha}_i \leq 1/l, \quad i = 1,\ldots,l, \quad \text{with} \quad e^T\bar{\alpha} \geq \nu l, \quad y^T\bar{\alpha} = 0$$

If α is an optimal for the dual equation and ρ is an optimal for the primal equation, Chang and Lin (2001) showed that α/ρ is an optimal solution of *C*-SVM with $C = 1/(\rho l)$. Thus, the output in the model is $(\alpha/\rho, b/\rho)$.

***Multi-class Classification*:** In the real-world problem, however, we find more than two classes of faults. For example, in the fault diagnosis of rotating machineries there are several fault classes such as gear faults, mechanical unbalances, misalignments, bearing faults, etc. and in the gear fault also several faults like the wear of teeth, the missing tooth, the chipped tooth, etc. Methods like one-against-all, one-against-one, direct acyclic graph, etc. are addressed for the multi-class problem.

Here we implement the 'one-against-one' approach (Knerr *et. al.,* 1990; Kressel, 1998) for the multi-class classification. If k is the number of classes, then $k(k-1)/2$ classifiers are constructed and each one trains data from two classes. For example for training in the *C*-SVC by the i^{th} and j^{th} classes of data, the classification function can be defined as

$$\min_{w^{ij},b^{ij},\xi^{ij}} \quad \left\{ \frac{1}{2}\left(w^{ij}\right)^T w^{ij} + C\sum_t \left(\xi^{ij}\right)_t \right\} \tag{3.19}$$

Subject to $\left(w^{ij}\right)^T \phi\left(x_t\right) + b^{ij} \geq 1 - \xi_t^{ij}$, if x_t in the i^{th} class

$\left(w^{ij}\right)^T \phi\left(x_t\right) + b^{ij} \geq -1 + \xi_t^{ij}$, if x_t in the j^{th} class

$\xi_{ij} \geq 0$

In the classification we use a voting strategy in which each binary classification is considered to be a voting, where votes could be cast for all data points, *x*, at the end a point is designated to be in a class with the maximum number of votes. In case those two classes have identical votes, though it may not be a good strategy, it chooses the class appearing first in the array of storing class names. Many methods are available for the multi-class SVM classification and Hsu and Lin (2002a) gave a detailed comparison and concludes that 'one-against-one' is a competitive approach.

***Performance Measure*:** Predict labels of the testing data could be determined after solving the SVM formulation. Let $x_1,\ldots,x_l$ be the training/ testing data and $f\left(x_1\right),\ldots,f\left(x_l\right)$ be the class (classified by the classifier). If a true class of training/ testing data are known and denoted as $y_i,\ldots,y_l$, the predicted results could be evaluated by the classification accuracy.

$$\text{Accuracy} = \left(\text{number of correctly predicted data} / \text{total number of testing data}\right) \times 100\% \tag{3.20}$$

The LIBSVM (2011) freely available software package is used for the classification. This package has been used by several users from its introduction and this has shown a good classification capability for the multi-class classification.

3.2.1 Parameter Selection

The selection of the regularization parameter (C), the upper bound on the fraction of support vector (ν) and the RBF kernel parameter (γ) are required before training of the SVM. The optimized selection of parameters can be possible by using optimizing tool such as the grid-search technique (GST), the genetic algorithm (GA), the artificial bee colony (ABC) algorithm, etc. These are described briefly as follows

***The Grid Search Technique*:** In the grid search technique (Hsu *et al.*, 2010), the data set was further randomly divided into training sets and independent test sets via a k-fold cross validation. Each of the k subsets acted as an independent test set for the training with the remaining $(k-1)$ subsets. Advantages of the cross validation are that all of the test sets were independent and the reliability of results could be improved. The data set is divided into k subsets for the cross validation. Here 10-fold cross validation (CV) is chosen, which means that all of the data will be divided into 10 parts, each of which will take turns at being the testing data set. The other nine data parts serve as the training data set for adjusting the model prediction parameters. The grid search method (GSM) is the common method for searching best SVM parameters.

The combining GSM with the SVM is illustrated in Figure 3.3. The total features (data set) are divided into two sub-sets parameter estimation data and the final testing data. The parameter estimation data set is trained in 10 fold cross validation within grid points. Grid

points are selected in the logarithmic scale for the CV calculation. After finding the better value in the grid, a finer grid search on that region is done. At the end, the highest CV accuracy parameters are selected. The training of the database is performed with that selected parameters. After that the testing of data sets are done for the calculation of accuracy. The LIBSVM (2011) tool is used for the classification.

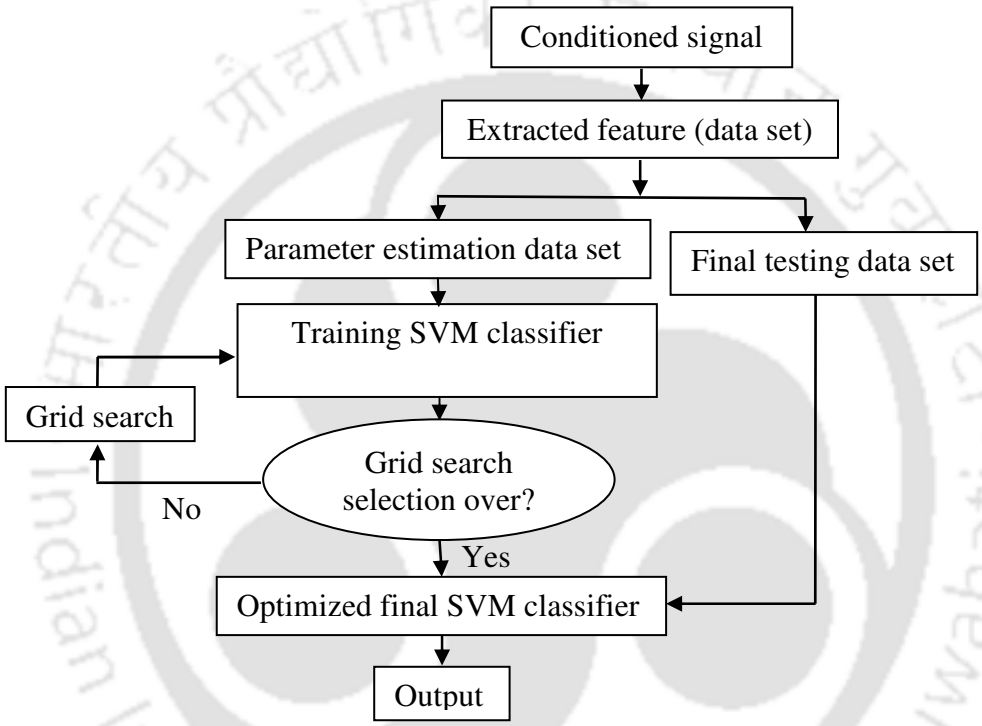


Figure 3.3 A flow chart for the optimization of SVM parameters by the GSM

***Genetic Algorithm*:** Genetic algorithms (GA), a general adaptive optimization search methodology based on a direct analogy to the Darwinian natural selection and genetics in biological systems, is a promising alternative to conventional heuristic methods. The GA work with a set of candidate solutions called a population. Based on the Darwinian principle of 'survival of the fittest', it obtains the optimal solution after a series of iterative computations. It generates successive populations of alternate solutions that are represented by a chromosome, i.e. a solution to the problem, until acceptable results are obtained.

Associated with characteristics of the exploitation and exploration search, the GA can deal with large search spaces efficiently, and hence has less chance to get a local optimal solution than other algorithms.

The GA has some advantages in solving the problem with respect to the traditional optimization tools like the penalty function method (a method for solving a constrained optimization problem (Deb, 2005)). At first, we do not have to care much about the non-linearity of cost functions and constraints. The next advantage is that the GA does not need the initial value due to stochastic nature in itself. The GA can also be useful for the discrete optimization problem. The GA shows it's excellencies in finding global solution of constrained optimization problems. Unlike many other optimization techniques, GAs can find solutions in the broader search space, and the better result can be found than the ones by traditional methods (Michalewicz, 1994).

In spite of above advantages, GAs are inherently suitable for unconstrained problems and take longer time to solve problems due to more frequent function calls. To overcome these drawbacks, a penalty method (Deb, 2005) is introduced to handle constraints with the GA. This method is formulated to combine cost function values and violation values of constraints, generate a new cost function, and then minimize it.

Since the GA algorithm works under the survival of the fittest principle of the nature. Therefore, GAs are naturally suitable for solving maximization problems. Minimization problems are usually transformed into maximization problems by a suitable transformation. In general, a fitness function $f'(x)$ is first derived from the objective function and used in successive genetic operations (Deb, 2005).

The genetic algorithm (GA) technique is well known and hence the discussion is limited to the five fundamental issues and is described below (Deb, 2003 and 2005). Figure 3.4 is illustrated the general configuration GA. The ranking selection, arithmetic crossover, and uniform probability distribution are used for the selection operator, crossover operator and mutation operator, respectively.

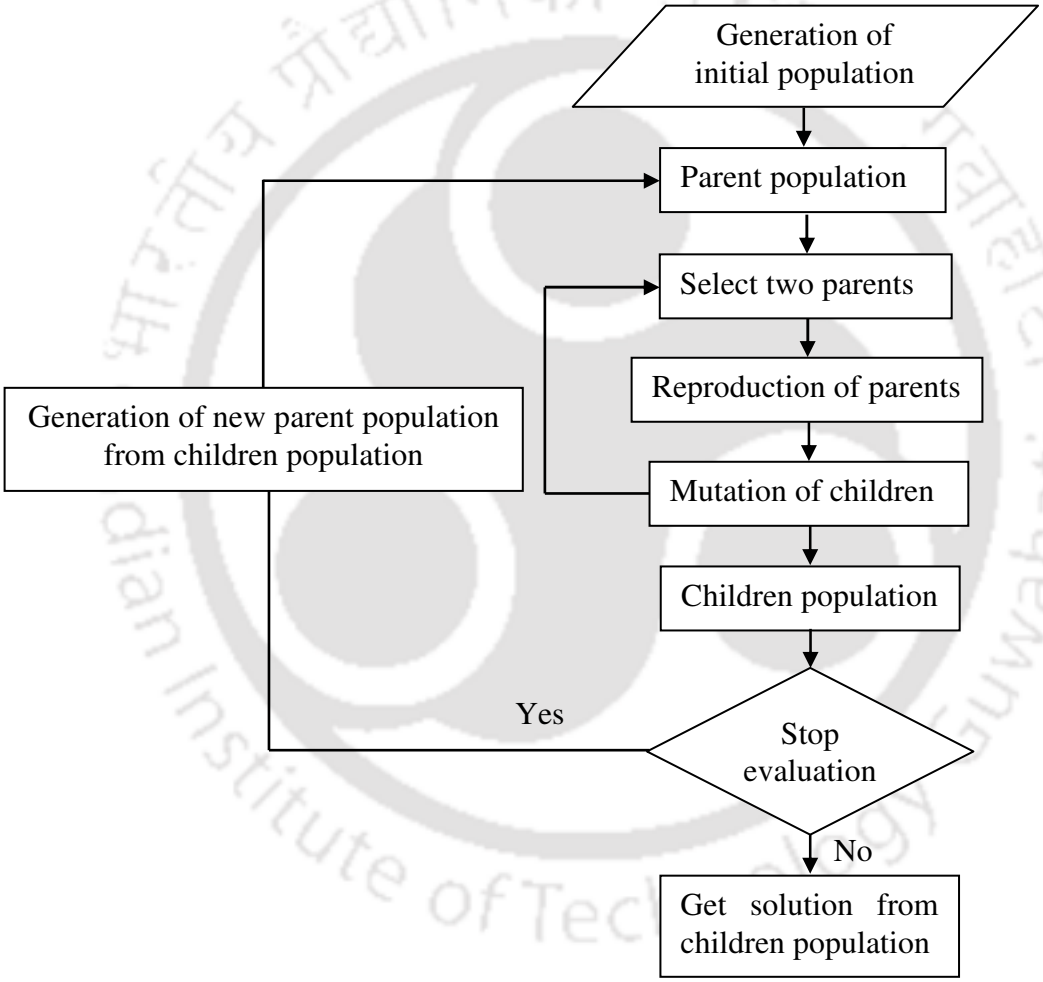


Figure 3.4 Flow chart of GA

The five fundamental issues for the use of a GA are

- Chromosome representation
- Selection function

- Genetic operators making up the reproduction function
- The creation of the initial population
- Termination criteria

Chromosome Representation: The chromosome representation is needed to describe each individual in the population of interest. The representation scheme determines how the problem is structured in the GA and also determines the genetic operators that are used. Each chromosome is made up of genes from a certain alphabet. An alphabet could consist of binary digits (0 and 1), floating point numbers, integers, symbols (i.e. A, B, C, D), matrices, etc. Michalewicz (1994) did extensive experimentation comparing real valued and binary GAs and showed that the real valued GA is an order of magnitude more efficient in terms of the CPU time. He also showed that a real valued representation moves the problem closer to the problem representation, which offers higher precision with more consistent results across replications (Micalewicz, 1994).

Selection Function: The selection of individual to produce successive generations plays an extremely important role in a GA. A probabilistic selection is performed based upon the individual's fitness such that better individuals have an increased chance of being selected. An individual in the population can be selected more than once with all individuals in the population having a chance of being selected to reproduce into the next generation. There are several schemes for the selection process: the roulette wheel selection and its extensions, the scaling technique, tournament, elitist models and ranking methods (Goldberg 1989, Michalwicz, 1994).

A common selection approach assigns a probability of selection, P_j, to each individual, *j*, based on its fitness value. A series of *N* random numbers is generated and compared against the cumulative probability, $C_j = \sum_{j=1}^{i} P_j$, of the population. An appropriate individual, *i*, is selected and copied into the new population if $C_{i-1} < U(0,1) \leq C_i$. Various methods exit to assign probabilities to individuals: the roulette wheel, the linear ranking and the geometric ranking.

The ranking method only requires the evaluation function to map the solutions to a partially ordered set, thus allowing for the minimization and the negativity. Ranking methods assign P_i based on the rank of solution *i* when all solutions are stored. The normalizing geometric ranking (Joines and Houck, 1994) defines P_i for each individual by

$$P\left(\text{Selecting} \quad \text{the} \quad i^{th} \quad \text{individual}\right) = q'(1-q)^{r-1} \tag{3.21}$$

where q is the probability of selecting the best individual, r is the rank of individual with value of 1 is the best individual, P is the population size and $q' = \frac{q}{1-(1-q)^P}$. In the present discussion we used 0.08 as the probability of selecting the best.

Genetic Operators Making up the Reproduction Function: Genetic operators provide the basic solutions based on existing solutions in the population. There are two basic types of operators: the crossover and the mutation. The crossover takes two individuals and produces two new individuals while the mutation alters one individual to produce a single new solution. The

application of these two basic types of operators and their derivatives depends on the chromosome representation used.

Operators for real-valued representations, i.e. an alphabet of floats, were developed by Michalewicz (1994). For real $\tilde{X}$ and $\tilde{Y}$, following operators are defined: the uniform mutation, non-uniform mutation, boundary mutation, simple crossover, arithmetic crossover and heuristic crossover. Let a_i and b_i be the lower and upper bounds, respectively, for each variable *i*.

Uniform mutation randomly selects one variable, *j*, and sets it equal to a uniform random number $U(a_i, b_i)$:

$$x_i' = \begin{cases} U(a_i, b_i), & \text{if} \quad i = j \\ x_i, & \text{otherwise} \end{cases} \tag{3.22}$$

The arithmetic crossover produces two complimentary linear combinations of parents, with $r = U(0,1)$, as

$$\tilde{X}' = r\tilde{X} + (1-r)\tilde{Y} \tag{3.23}$$

$$\tilde{Y}' = (1-r)\tilde{X} + r\tilde{Y} \tag{3.24}$$

Initialization and termination criteria: The GA must be provided initial populations. The most common method is to randomly generate solutions of the entire population. However,

since GAs iteratively improve existing solutions (i.e., solutions from other heuristics and/or current practices); hence the initial population can be randomly generated.

The GA moves from generation to generation selecting and reproducing parents/children until a termination criterion is met. The most frequently used stopping criterion is a specified maximum number of generations.

The freely available GA optimization toolboxes developed by Kay (GAOT, 1996) is used here. The initial random designed variables are fed to the LIBSVM (2011) software for finding the classification accuracy. On the basis of the optimized classification accuracy the new variables are generated until the convergence or stopping criteria is reached. Parameters for the GA are tabulated in Table 3.1.

***Artificial Bee Colony Algorithm (ABCA)*:** Artificial bee colony algorithm (ABCA) is one of the most recently defined algorithm by Devis Karaboga in 2005, motivated by the intelligent behaviour of honey bees. It is as simple as the Particle Swarm Optimization (PSO) and Differential Evolution (DE) algorithms, and uses only common control parameters such as colony size and maximum cycle number. The ABCA as an optimization tool provides a population-based search procedure in which individuals called foods positions are modified by the artificial bees with time and the bee's aim is to discover the places of food sources with high nectar amount and finally reached the place of the highest nectar amount. Figure 3.5 illustrates the ABCA (Karaboga and Basturk, 2007 and 2008). Each cycle of the search consists of three steps after the initialization stage:

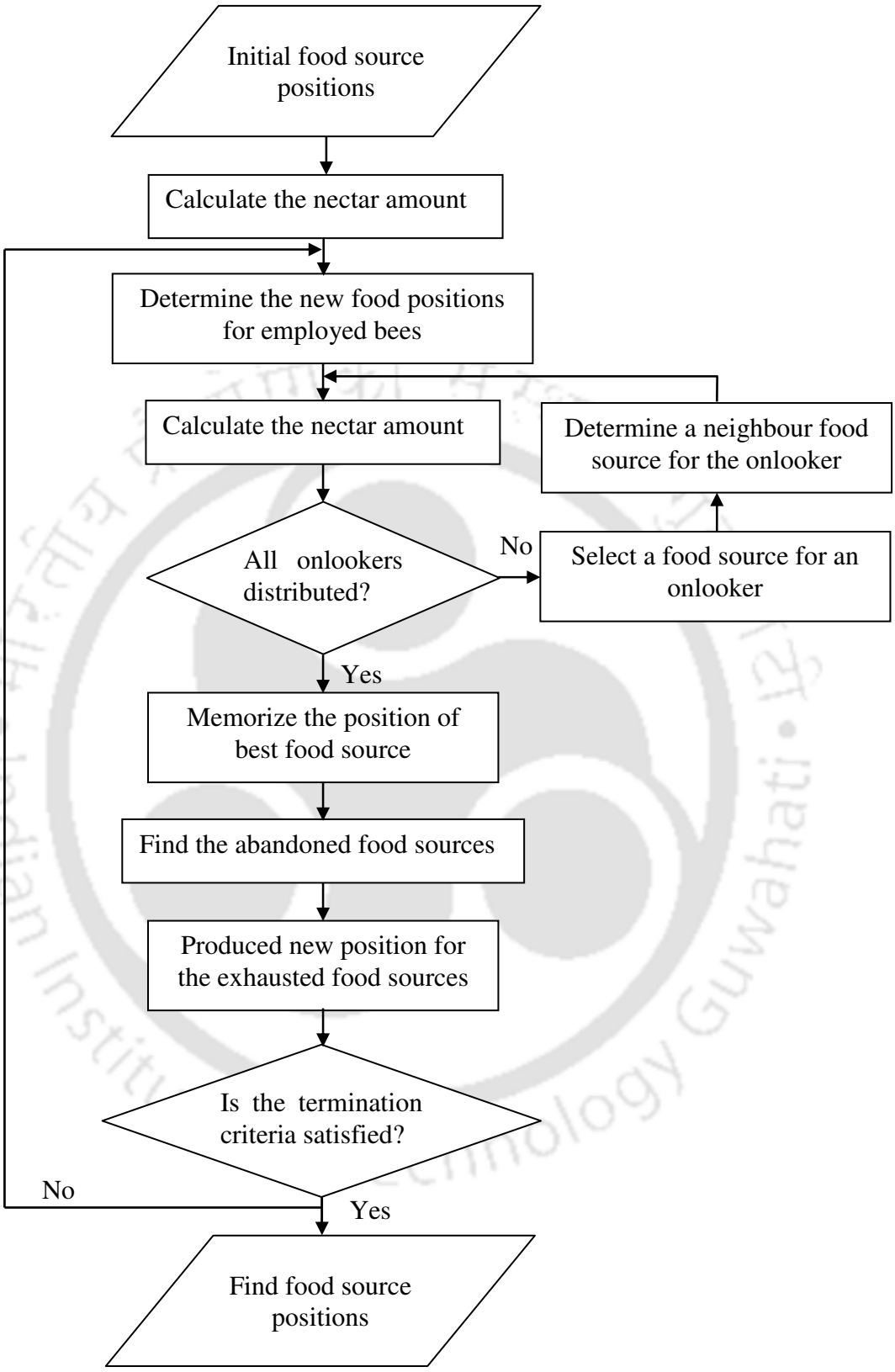


Figure 3.5 Flow chart of ABCA

- placing the onlookers onto the food sources and calculate their nectar amount
- placing the onlookers onto the food sources and calculating the nectar amounts and
- determining the scout bees and placing them onto the randomly determined food sources

In the ABCA, a food source position represents a possible solution to the problem to be optimized. At initialization, a set of food source positions are randomly produced and values of the algorithm control parameters are assigned. The nectar amount retrievable from the food source corresponds to the quality of the solution represented by that food source. So the nectar amount of food sources existing at initial positions is determined. In other words, the quality (values) of initial solutions is calculated. Each employed bee is moved onto her food source area to determine a new food source within the neighbourhood of the present location, and then its nectar amount is evaluated. If the new nectar amount is higher, then she forgets the previous amount and remembers the new one. After employed bees complete their search, they come back into the hive and their information about the nectar amount of their sources with the onlookers waiting on the dance area. All onlookers successively determine a food source area with a probability based on their nectar amounts. Each onlooker determines a neighbour food source within the neighbourhood of the one to which she has been assigned and then its nectar amount is evaluated.

A honey bee colony has scouts that are the colony's explorers who do not need any guidance while looking for the food. In ABCA, if a solution representing a food source has been exhausted by bees and employed bees of this food source becomes a scout. The position of the abandoned food source is replaced with a randomly produced food position. The number of trials for releasing a food source is equal to the value of 'limit', which is an important control

parameter of the ABCA. These three steps are repeated until the termination criteria are satisfied.

The food source represents a possible solution of design variables (C,ν,γ) and the nectar amount of a food source corresponds the value of objective function (i.e., the accuracy of classification). At the beginning, the ABCA randomly generates the population $P_{initial}$ of size N and from that generated solutions x_i $(i=1,2,\ldots,N)$ in S-dimensional vector, where S is the number of optimization parameters. After initialization, the population of solutions is subject to repeated cycles, $C=1,2,\ldots G$, of search processes of the employed, onlooker and scout bees. An employed bee memorizes the new position if the objective function value of a new solution is higher than that of the previous, else she keeps the old position in her memory. The employee bees share the nectar information on the dance area with onlooker bees. An onlooker bee evaluates the fitness information taken from all employed bees and chooses a food source with a probability related to its fitness value. An onlooker bee also produces a new solution and it memorizes the new position if its fitness value is better than the previous position. An artificial onlooker bee chooses a food source depending on the probability value associated with that food source, P_i , calculated by $P_i = F_i \Big/ \sum_{n=1}^{N} F_n$, where F_i is the fitness value of the solution i, which is proportional to the nectar amount of the food source in the position i and N is the number of food sources, which is equal to the number of employed bees. In order to produce a candidate new food position (v_{ij}) from the old one (x_{ij}) in the memory, the ABCA uses the following expression (Karaboga and Basturk, 2008)

$$v_{ij} = x_{ij} + R_{ij}\left(x_{ij} - x_{kj}\right) \tag{3.25}$$

where $k \in \{1,2,\ldots,N\}$ and $j \in \{1,2,\ldots,S\}$ are randomly chosen indices. Although k is determined randomly, it has to be different from i. R_{ij} is a random number between (-1, 1). It controls the production of neighbour food sources around x_{ij} and represents a comparison of two food positions visually by a bee. As can be seen from equation (3.25) as the difference between parameters of x_{ij} and x_{kj} decreases, the perturbation on the position x_{ij} also decreases. Thus, as the search approaches to the optimum solution in the search space, the step length is adaptively reduced. The food source is abandoned if the position of the food source cannot be improved after some predetermined number of cycles and it is replaced by a new food source by the scout. In the ABCA, this is simulated by producing a position randomly and replacing it with the abandoned one. The value of predetermined number of cycles is an important control parameter of the ABCA, which is called the '*limit*' for the abandonment. The value of limit, L, is generally taken as $L = N \times S$ (Karaboga and Basturk, 2008). Assume that the abandoned source is x_i and $j \in \{1,2,\ldots,D\}$, then the scout discovers a new food source to be replaced with x_i. This operation can be defined as in (Karaboga and Basturk, 2007)

$$x_i^j = x_{\min}^j + \text{rand}(0,1)\left(x_{\max}^j - x_{\min}^j\right) \tag{3.26}$$

It is clear from the above explanation that there are three control parameters used in the ABCA. The number of food sources, which is equal to the number of employed or onlooker bees (N), the value of limit (L), and the maximum cycle number (G). The ABCA (MATLAB, 2009) is applied to the present optimization problem. Parameters for the ABCA are tabulated in Table 3.1. Figure 3.6 illustrates the selection of the parameter by the ABCA and the GA. The total features (data set) are divided into two sub-sets parameter estimation data and final

testing data. Again the parameter estimation data is subdivided into the training and testing data set for the selection of optimized parameters by the GA or ABCA. After getting the optimized SVM classifier, the final testing data set is used for finding the final classification. The fitness function used is the accuracy as discussed in Eqn. (3.20). Bounds for parameters are selected so that the computational time is minimized. Table 3.2 shows the fitness function and design parameters with bounds for the optimization technique for two SVM techniques with the RBF kernel.

Population sizes (50 for GA and 100 for ABCA) are chosen by trial and error method, and now relevant analysis figures are added on Appendix C. The limit for ABCA is the multiplication of number of employee bee (50) and number of optimization parameters (2), which is equal to 100 (Karaboga and Basturk, 2007).

The maximum number of generation for the GA is chosen by trial and error method, and its value is fixed as 50. Other parameters for GA (uniform mutation: 4; arithmetic crossover: 4; normalized geometric section: 0.08) are taken as suggested by the GA software (GAOT) developer (Houck *et al.*, 1996). The classification of gear faults from time domain data is discussed in the next chapter.

Table 3.1 Parameter chosen for optimization

Parameter	Value
For GA	
Population size	50
Generation number	50
For ABCA	
Number of food sources	100
Limit	100

Table 3.2 Optimization fitness function and design parameters

Fitness function	Design parameters	Bounds
***C*-SVC with RBF kernel**		
Maximize $f(x)$ = (number of correctly predicted data/total number of testing data)×100%	$X = \lfloor \gamma \quad C \rfloor^T$	For γ: 0-1 For C: 0-1.5
ν-SVC with RBF kernel		
Maximize $f(x)$ = (number of correctly predicted data/total number of testing data)×100%	$X = \lfloor \gamma \quad \nu \rfloor^T$	For γ: 0-1 For ν: 0-1

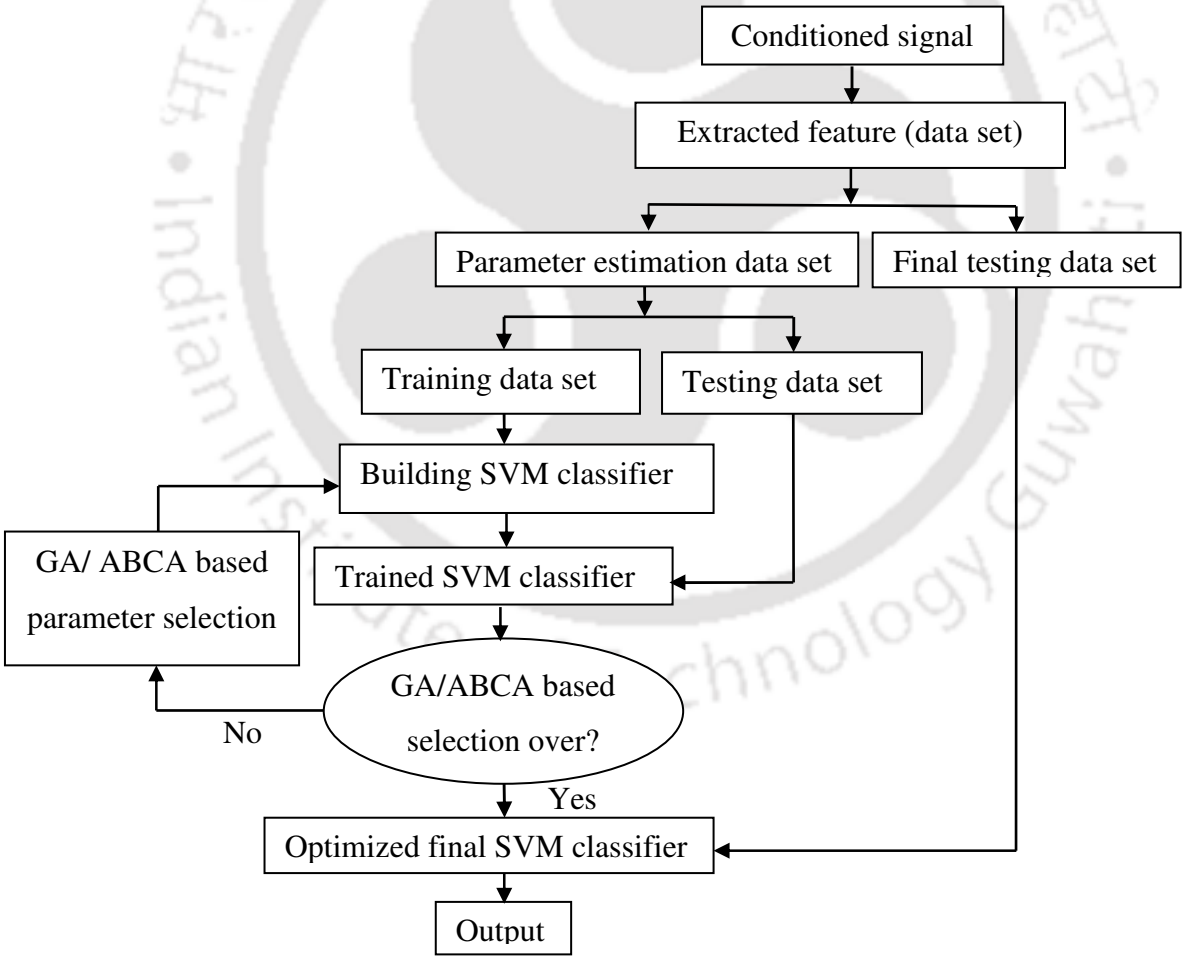


Figure 3.6 Flow chart for the optimization of SVM parameters by GA and ABCA

3.3 Summary

The fault classification methodology (SVM) has been discussed in the present chapter. Two variation of the SVM (C-SVC and ν-SVC) is presented for the fault classification. The procedure for optimization of parameters of classifier for optimum prediction accuracies is discussed in detail. Here the GSM, GA and ABCA methods have been used for optimization tool. The procedure for feeding of data (features) to the classifier for training and classification is established. In Chapter 4, the procedure for features calculation from time domain data and the classification of faults by the SVM is discussed.

CHAPTER 4

Multiclass Classification Using Time Domain Data

4.1 Introduction

Time domain methods try to analyze the amplitude and phase information of the vibration time signal to detect the fault of gear in a gear-rotor-bearing system. Time domain is a perceptive that feels natural, and provides physical insight into the vibration. It is particularly useful in analyzing impulsive signals from gear defects with non-steady and short-transient impulses. The time domain analysis is focused to determine the statistical characteristics of the original function by manipulating the series of discrete numbers. Therefore, some statistical methods are employed to describe the spiky nature of time signals.

In previous chapters, collections of vibration signals from the faulty and healthy gear-box along with the classification methodology by the SVM classifier and its optimization techniques have been discussed. Statistical features that are calculated from the collected time domain data set will be discussed here. The classification of faults at same rotational speed as measured data, at speed in between two measured data (i.e., intermediate) and at speed outside the measured data (i.e., extrapolated) is carried out and prediction accuracies are discussed.

4.2 Statistical Feature Extraction

The procedure for capturing the vibration signature is discussed in Chapter 2. The information concerning the health of gear is contained in such vibration signatures. Hence, faulty vibration signatures could be compared with healthy signatures to determine whether the component is behaving normally or exhibiting signs of failure. In practice, such comparisons are not effective. Due to the large variations, the direct comparison of signatures is difficult. Instead, a more useful technique that involves the extraction of features from the vibration signature

data could be used. Ideally, these features are more stable and well behaved than the raw signature data itself. Features also provide a reduced data set for the application of pattern recognition and tracking techniques.

The peak level, root mean square (RMS) level, and the crest factor are often used to quantify the time signal. This can be called as the indices. The definition of the peak, RMS and crest factor is defined as under.

Peak: The peak level (T_1) of the signal is defined simply as half the difference between the maximum and minimum vibration levels. This can be written as

$$T_1 = \max|x_i| \tag{4.1}$$

RMS: The RMS (root mean square, T_2) value of the signal is the normalized second statistical moment of signal (standard deviation). This can be written as

$$T_2 = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \tag{4.2}$$

Crest Factor: The crest factor (T_3) is defined as the ratio of the peak value to the RMS of the signal. This can be written as

$$T_3 = \frac{T_1}{T_2} \tag{4.3}$$

where, x_i is the amplitude of the i^{th} digitized point in time domain, and N is the number of point in time domain. The peak level is not a statistical quantity and hence may not be reliable in detecting damage in continuously operating systems. The RMS of the signal is commonly used to describe the 'steady-state' or 'continuous' amplitude of a time varying signal. This feature gives good results in tracking the overall noise level in the signal but will not give any information about the component that is going to fail. The crest factor, defined as the ratio of the peak value to RMS level, has been proposed as a trending parameter as it includes both parameters. If any peaks appear in time domain signal will result in an increase of the crest factor. The best result using this feature is obtained in the case of impulsive vibration sources such as tooth breakage on gears.

The statistical analysis of the vibration signal is one of the criteria for the selection of features. Toyota *et al.* (1999) proposed a method to detect and diagnose the rotating machinery condition based on the hypothesis that if the machine is in good condition, its density function of the vibration signal is distributed normally. Based on this hypothesis, they expanded the density function of vibration signal into orthogonal feature space and separated the good and deteriorated components in the density function by Gram-Charlier (GC) expansion theory. For vibration $f(x)$, let the mean $\mu_1 = \bar{x}$ and the variance σ^2, r^{th} moment about the mean μ_r can be expressed as

$$\bar{x} = \mu_1 = E\{x\}$$

$$\sigma^2 = \mu_2 = E\left\{(x-\mu_1)^2\right\} \tag{4.4}$$

$$\mu_r = E\left\{(x-\mu_1)^r\right\}$$

Using r^{th} mean centered moment μ_r, expansion coefficient can be calculated

$$c_0 = 1; \qquad c_1 = 0; \qquad c_2 = \frac{1}{2}(\mu_2 - 1); \qquad c_3 = \frac{1}{6}\mu_3$$

$$c_4 = \frac{1}{24}(\mu_4 - 6\mu_2 + 3); \qquad c_5 = \frac{1}{120}(\mu_5 - 10\mu_3); \text{ and so on.} \tag{4.5}$$

The expansion coefficient c_i is very good feature parameters to detect and identify the failure types, which is selected by engineer's intuition, and by try and error method. As seen in Eqn. (4.5), the coefficient c_3 is equivalent to the skewness and coefficient c_4 is equivalent to the kurtosis.

Jack and Nandi (2002) used a number of different statistical features estimated based on the moments and cumulants of vibration data for the classification of bearing faults. They used up to fourth moments and cumulants of two perpendicular direction signals. Moments and cumulants features were used by Samanta (2004) for the gear fault classification. Different moments are the mean (first moment), variance (second moment about mean), skewness (third moment about mean), and kurtosis (fourth moment about mean). The variance can also represent as a square of the standard deviation.

In the present study we extracted standard deviation, skewness and kurtosis from the time domain signal. Three statistical features are explained in the following paragraph.

Standard Deviation (S_d): It shows how much variation or "dispersion" is there from the "average" (i.e., the mean, or expected/budgeted value). A low standard deviation indicates that data points tend to be very close to the mean, whereas a high standard deviation indicates

that the data are spread out over a large range of values. The standard deviation is the second standardized moment of data. It is defined as follows for the sampled signal $\{ x_1, x_2, \ldots, x_n \}$,

$$S_d = \sqrt{\frac{n\sum x_i^2 - \left(\sum x_i\right)^2}{n(n-1)}} \tag{4.6}$$

where the summation is implied that it is overall index values, i.e. $i = 1, 2, \ldots, n$.

Skewness (S_k)*:* The skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable. It can take positive or negative value, or even undefined. Qualitatively, a negative skew indicates that the tail on the left side of the probability density function is longer than the right side and the bulk of values (including the median) lie to the right of the mean. A positive skew indicates that the tail on the right side is longer than the left side and the bulk of values lie to the left of mean. A zero value indicates that values are relatively evenly distributed on both sides of the mean, typically but not necessarily implying a symmetric distribution. The skewness is the third standardized moment of the data. It is defined as follows with $\overline{x}$ as the mean

$$S_k = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \overline{x}}{S_d} \right)^3 \quad \text{with} \quad \overline{x} = \sum x_i / n \tag{4.7}$$

Kurtosis (K_u): It is defined as the fourth moment of the distribution and measures the relative peakedness or flatness of a distribution as compared to a normal distribution. As a gear wears and breaks, this feature should indicate an error due to the increased level of vibration. The equation of kurtosis can be written as

$$K_u = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_i - \bar{x}}{S_d} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)} \tag{4.8}$$

The three statistical features are calculated from 2,000 data points. Finally, 300 data points are collected in three directions for each three features (i.e., total 9 combinations), which means 9×300 data sets are available for each faults. For four different faults (i.e., CT, MT, WT and ND) there are total $9 \times 4 \times 300$ data sets available.

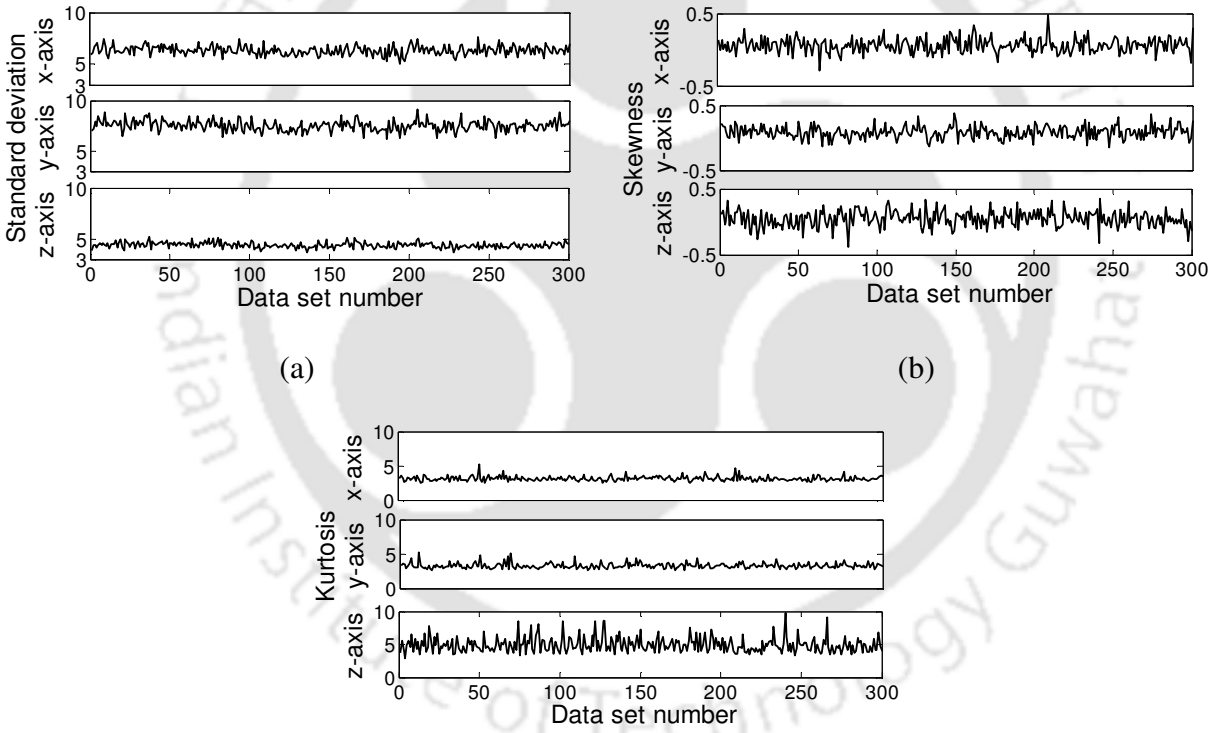


Figure 4.1 Time domain features of acquired signals for CT at 30 Hz rotational speed

These total data points are divided and used for the training, testing of optimizing parameters and for the final testing in the simulation. Figure 4.1 illustrates the statistical features

extracted from time domain vibration signals in three orthogonal directions for the CT gear fault at 30 Hz rotational speed.

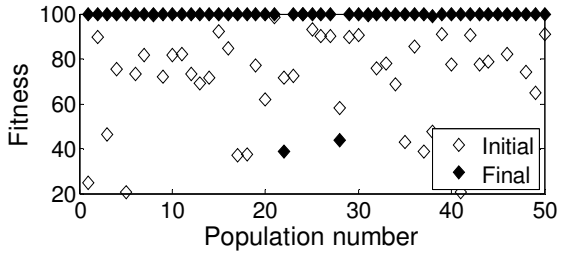
4.3 Simulation Results

The SVM software first needs to be trained with vibration data comprising of four types of gear fault conditions (i.e., CT, MT, WT and ND). For testing and training following three conditions are considered.

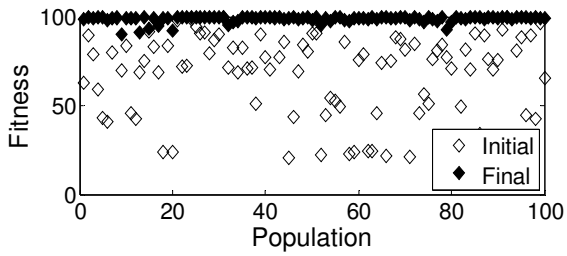
4.3.1 Training and Testing at the Same Rotational Speed

In this case the SVM is trained and tested at the same rotational speed. Rotational speeds selected are in a range of 10-30 Hz in a step of 5 Hz.

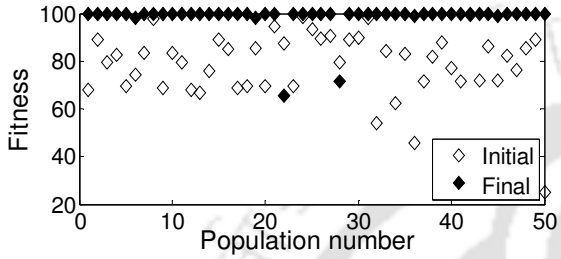
Optimization of Parameters: During the optimization of SVM parameters by the GA and the ABCA, $9\times4\times180$ data sets are used for the training of the fault classification, and $9\times4\times90$ data sets are used for the testing. The variation of initial and final fitness values (i.e., the percentage accuracy) with the population for the *C*-SVC using the GA and the ABCA, respectively, for 30 Hz rotational speeds are shown in Figure 4.2(a) and Figure 4.2(c). Similarly, the variation of the initial and final fitness values with the population for the ν-SVC using the GA and the ABCA, respectively, for 30 Hz rotational speed are shown in Figure 4.2(b) and Figure 4.2(d). It includes all the population, including solutions violating constraints. From the choice of initial solutions, it could be seen that it has a quite good divergence of the possible solution domain. In the last generation also all solutions have been shown including those violating constraints (however, very few), and it could be observed that the most of solutions have converged.



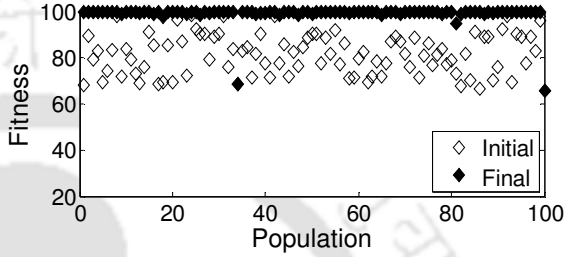
(a) 30 Hz (*C*-SVC using GA optimization)



(b) 30 Hz (*C*-SVC using ABCA optimization)

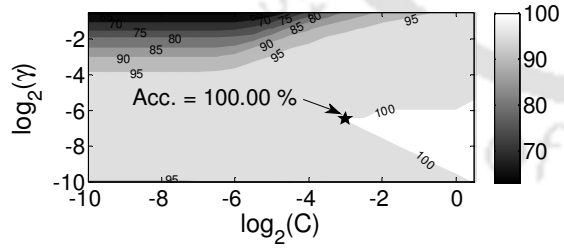


(c) 30 Hz (ν-SVC using GA optimization)

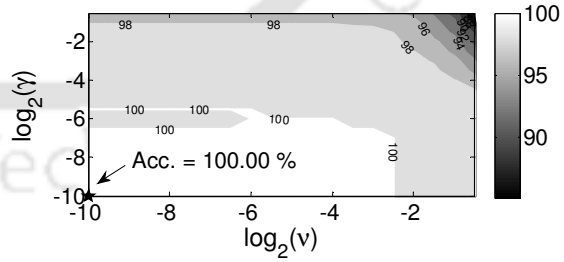


(d) 30 Hz (ν-SVC using ABCA optimization)

Figure 4.2 Variation of the initial and final fitness (percentage accuracy) with the population (a)-(b) for the *C*-SVC and (c)-(d) for the ν-SVC using the GA and ABCA for 30 Hz training and testing speed



(a) 30 Hz (*C*-SVC)



(b) 30 Hz (ν-SVC)

Figure 4.3 Cross validation accuracy (a) for the *C*-SVC (b) for the ν-SVC for 30 Hz training and testing speed

In the case of GSM $9\times4\times270$ data sets are used for the parameter estimation. The cross validation accuracy in the grid search technique for 30 Hz rotational speed for the *C*-SVC and the ν-SVC are shown in Figure 4.3(a-b), respectively. In these the contour line for the percentage accuracy is plotted and the best CV accuracy is marked. All CV values are stored with the progress of calculation in a stack. SVC parameters are found from the best CV accuracy, and the testing accuracy is tabulated. If more than one best (similar) CV is found, then any one of corresponding parameters (i.e., logarithmic values of SVC and kernel parameters) are selected. The optimized percentage accuracy (all classes) of different SVM formulation is shown in Table 4.1. Here the accuracy found during training and testing time is named as training and testing accuracy, respectively. It is noticed that in few cases, testing accuracy is higher than the training because the testing data may be more easily separable with reference to training.

Table 4.1 Classification of gear faults (all classes) with the training and the testing for the same speed (time domain features)

Training speed (Hz)	Testing speed (Hz)	SVM formulation	Training accuracy* (%)			Testing accuracy (%)		
			CV	GA	ABCA	CV	GA	ABCA
10	10	*C*-SVC	97.41	97.50	98.06	97.50	96.67	97.50
		ν-SVC	98.43	99.17	99.44	**99.17**	**99.17**	98.33
15	15	*C*-SVC	93.43	93.33	93.33	**90.83**	**90.83**	**90.83**
		ν-SVC	93.61	95.28	95.56	88.33	90.00	90.00
20	20	*C*-SVC	96.11	97.78	98.06	97.50	97.50	**98.33**
		ν-SVC	97.04	97.78	98.33	88.33	95.83	97.50
25	25	*C*-SVC	99.72	99.44	99.44	100.00	**100.00**	**100.00**
		ν-SVC	99.72	99.44	99.44	100.00	99.17	**100.00**
30	30	*C*-SVC	100.00	100.00	100.00	**100.00**	**100.00**	**100.00**
		ν-SVC	100.00	100.00	100.00	**100.00**	**100.00**	**100.00**

*CV means the *cross validation* accuracy of grid search technique. Bold values represent prediction accuracies of the maximum accuracy.

Fault Prediction Ability: After the optimization of parameters total $9\times4\times30$ data sets are used for the final testing of the fault classification for the GA and the ABCA, and $9\times4\times30$ data sets are used for the final testing of the fault classification for the GSM. Table 4.3 illustrates data points used in the parameter estimation and the final testing. In many cases the accuracy is more in the GA and the ABCA as compared with the GSM, which reflect the soundness of the GA and the ABCA. If we look upon the all classes testing accuracy, the lowest one is equal to 88.33% and this occurs at 15 Hz and 20 Hz rotational speeds on the ν-SVC case. It is also observed that in the higher rotational speed the prediction accuracy is near perfect (i.e., from 25 Hz to 30 Hz).Table 4.5 illustrates the percentage prediction (individual classes) accuracy in various rotational speeds against the best prediction accuracy. The prediction accuracy of 76.67% is the individual lowest against the MT gear fault case at 15 Hz rotational speed.

4.3.2 Training at two Different Rotational Speeds and Testing at an Intermediate Rotational Speed

In this case the SVM is trained at two rotational speeds and tested at an intermediate rotational speed for which the SVM was not trained. Values of rotational speed range and the intermediate testing rotational speed are tabulated in Table 4.2. In this case beyond 10 Hz (i.e., 600 rpm) rotational speed range are not considered, because the range is wide and hence the percentage accuracy becomes low.

Optimization of Parameters: Here $9\times8\times90$ data sets are used for the training and $9\times8\times45$ data sets are used for the testing of optimized parameters for the GA and the ABCA. For example, in the case of optimization of parameters for the first range, $9\times4\times90$ data sets are taken from 10 Hz rotational speed and $9\times4\times90$ data sets are taken from 15 Hz rotational speed so that all together $9\times8\times90$ data sets are taken for the training. Whereas, $9\times4\times45$

data sets are taken from 10 Hz rotational speed and $9\times4\times45$ data sets are taken from 15 Hz rotational speed so that all together $9\times8\times45$ data sets are taken for the testing. The optimized percentage accuracy (all classes) of different SVM formulations is provided in Table 4.4.

Fault Prediction Ability: After the optimization, $9\times4\times30$ data sets from intermediate rotational speeds are used for the final testing or classification. Similarly $9\times8\times125$ data sets are used for the parameter estimation and $9\times4\times30$ data sets are used for the final testing for the GSM. Data sets used for the training and the testing for the optimization of parameters and the final testing in the interpolation and the extrapolation are illustrated in Table 4.3. For the intermediate rotational speed range of 5 Hz, for the lowest rotational speed the classification accuracy (all classes) is nearly 84.17% and it increases at the high rotational speed to 100%. In the case of intermediate rotational speed range of 10 Hz, for the lowest rotational speed the classification accuracy (all classes) is nearly 35% and it increases at the high rotational speed to 94.17%. It is observed that the GA and ABCA techniques have shown the same prediction accuracy as with the conventional GSM, but in a particular speed range 10 Hz to 20 Hz (intermediate 10 Hz range), the GSM has shown a good result (12.5% more accuracy level then the GA and ABCA techniques). This may be due to reason that the former method is less sensitive to noise as compared to latter ones. The percentage prediction (individual classes) accuracy at various rotational speeds based on the best optimized condition is illustrated in Table 4.5. In the case of rotational speed range of 5 Hz, the classification for different faults is more than 76.67%. But in the case of rotational speed range of 10 Hz and the prediction accuracy at 15 Hz rotational speed the individual classification for fault conditions shows very low prediction accuracy and that is down to 33.33%.

4.3.3 Training at two Different Rotational Speeds and Testing at Extrapolated Rotational Speed

In this case the SVM is trained at two rotational speeds and tested at an extrapolated rotational speed. Training rotational speeds and the extrapolated testing rotational speed are tabulated in Table 4.2.

Table 4.2 The training and testing speeds for the fault prediction using the interpolation and the extrapolation

Training speeds (Hz)	Testing speed (Hz)	Training speed range (Hz)
For interpolation speed		
10, 15	12.5	5
15, 20	17.5	5
20, 25	22.5	5
25, 30	27.5	5
10, 20	15.0	10
15, 25	20.0	10
20, 30	25.0	10
For extrapolation speed		
10, 12.5	15	5
20, 22.5	25	5
25, 27.5	30	5
10, 15	20	5
15, 20	25	10
20, 25	30	10

Optimization of Parameters: For selecting data points the same procedure is considered as used for the intermediate interpolation, but only difference is that for testing the extrapolation speed is considered. The optimized percentage accuracy (all classes) of the different SVM formulation is shown in Table 4.4.

Table 4.3 Data sets used for parameter estimation and validation

Cases	Same speed	Interpolation and extrapolation speed
For GSM		
Parameters estimation	9×4×270	$\left.\begin{matrix}9\times4\times125\\9\times4\times125\end{matrix}\right\}\Rightarrow 9\times8\times125$
Final testing or validation	9×4×30	9×4×30
For GA and ABCA		
Training for optimizing the parameters	9×4×180	$\left.\begin{matrix}9\times4\times90\\9\times4\times90\end{matrix}\right\}\Rightarrow 9\times8\times90$
Testing of optimizing the parameters	9×4×90	$\left.\begin{matrix}9\times4\times45\\9\times4\times45\end{matrix}\right\}\Rightarrow 9\times8\times45$
Final testing or validation	9×4×30	9×4×30

Fault Prediction Ability: After optimization the same strategy for data points selection are considered for the testing as described in the interpolation case, whereas here the extrapolated speed is considered. In the extrapolated speed range of 5 Hz, the prediction accuracy (all classes) is 70.83% to 100%; and in case of 10 Hz range, it is 38.33% to 77.50%. It is observed that at 15 Hz and 30 Hz testing speeds for the case of 5 Hz range the GSM shows marginally good results, which is 1.16% and 0.83% more than the GA and ABCA techniques, respectively. On the other hand for 25 Hz testing speed the ABCA dominates in prediction accuracies (all classes). At 10 Hz range case, the GA and the ABCA shows reasonable results. For 5 Hz range it gives the best classification accuracy, and it increases gradually from 70.83% to 100% from the lowest to highest rotational speeds. In 10 Hz range also the best accuracy (all classes) increases from 62.50% to 77.50% from the lower to higher rotational speeds.

Table 4.4 Classification of gear faults (all classes) with the training and the testing for interpolation and extrapolation speed (time domain features)

Training speed (Hz)	Testing speed (Hz)	SVM formulation	Training accuracy* (%)			Testing accuracy (%)		
			CV	GA	ABCA	CV	GA	ABCA
For interpolation speed (range 5 Hz)								
10, 15	12.5	*C*-SVC	92.96	90.00	90.00	88.33	87.50	87.50
		ν-SVC	93.43	91.39	91.94	**89.17**	**89.17**	84.17
15, 20	17.5	*C*-SVC	93.43	91.39	91.39	**94.17**	93.33	93.33
		ν-SVC	93.80	91.67	92.22	**94.17**	**94.17**	93.33
20, 25	22.5	*C*-SVC	97.22	97.22	96.94	95.83	95.83	95.83
		ν-SVC	97.78	96.94	97.22	**96.67**	**96.67**	**96.67**
25, 30	27.5	*C*-SVC	98.98	98.61	98.61	**100.00**	**100.00**	**100.00**
		ν-SVC	99.54	99.72	99.72	**100.00**	**100.00**	**100.00**
For interpolation speed (range 10 Hz)								
10, 20	15	*C*-SVC	95.00	95.83	96.11	48.33	46.67	46.67
		ν-SVC	96.11	96.11	96.67	**59.17**	36.67	35.00
15, 25	20	*C*-SVC	95.46	94.17	94.44	65.00	63.33	65.83
		ν-SVC	95.93	93.61	94.44	79.17	63.33	**85.83**
20, 30	25	*C*-SVC	96.11	96.11	96.11	90.00	93.33	93.33
		ν-SVC	96.85	96.39	96.39	92.50	**94.17**	91.67
For extrapolation speed (range 5 Hz)								
10, 12.5	15	*C*-SVC	95.00	95.00	94.72	71.67	70.83	74.17
		ν-SVC	95.65	95.83	95.83	**75.83**	73.33	73.33
20, 22.5	25	*C*-SVC	96.39	95.83	96.11	80.83	80.00	80.00
		ν-SVC	97.04	94.72	97.22	83.33	70.00	**90.83**
25, 27.5	30	*C*-SVC	99.54	99.72	99.72	98.33	99.17	96.67
		ν-SVC	99.63	99.72	100.00	**100.00**	99.17	95.00
For extrapolation speed (range 10 Hz)								
10, 15	20	*C*-SVC	92.96	90.00	90.00	45.00	38.33	38.33
		ν-SVC	93.43	91.39	91.94	**62.50**	**62.50**	50.00
15, 20	25	*C*-SVC	93.43	91.39	91.39	55.00	49.17	49.17
		ν-SVC	93.80	91.67	92.22	62.50	62.50	**69.17**
20, 25	30	*C*-SVC	97.22	97.22	96.94	71.67	72.50	71.67
		ν-SVC	97.78	96.94	97.22	73.33	73.33	**77.50**

* CV means the *cross validation* accuracy of the GSM. Bold values represent prediction accuracies of the maximum accuracy.

Table 4.5 Percentage prediction (individual classes) accuracies at various speeds with the training and the testing (time domain features)

Speed (Hz)	CT (%)	MT (%)	WT (%)	ND (%)
For same speed				
10	100.00	100.00	100.00	96.67
15	93.33	76.67	100.00	93.33
20	96.67	93.33	100.00	100.00
25	100.00	100.00	100.00	100.00
30	100.00	100.00	100.00	100.00
For interpolation speed (range 5 Hz)				
12.5	86.67	96.67	96.67	76.67
17.5	90.00	90.00	100.00	96.67
22.5	96.67	90.00	100.00	100.00
27.5	100.00	100.00	100.00	100.00
For interpolation speed (range 10 Hz)				
15	70.00	93.33	40.00	33.33
20	86.67	80.00	86.67	90.00
25	100.00	96.67	93.33	86.67
For extrapolation speed (range 5 Hz)				
15	43.33	73.33	100.00	86.67
25	90.00	100.00	100.00	73.33
30	100.00	100.00	100.00	100.00
For extrapolation speed (range 10 Hz)				
20	43.33	86.67	96.67	23.33
25	73.33	100.00	90.00	13.33
30	96.67	100.00	96.67	16.67

The best percentage prediction accuracy (individual classes) at various rotational speeds is illustrated in Table 4.5. In the case of speed range of 5 Hz the lowest prediction accuracy (individual classes) is at 15 Hz for the fault detection of CT (i.e., 43.33%). Similarly, in the case of speed range of 10 Hz, the lowest performance is for the ND condition at 25 Hz (i.e.,

13.33%). For more clarity the overall performances are tabulated in Table 4.6. The lowest prediction accuracy (all classes) and highest prediction accuracy (all classes) are the lowest and highest testing accuracy (all classes) found from Table 4.1 and Table 4.4, respectively. The lowest and highest fault predictions are found from Table 4.5.

Table 4.6 Summary of overall fault prediction performances (time domain features)

	Same rotational speed	Interpolation speed		Extrapolation speed	
		Range 5 Hz	Range 10 Hz	Range 5 Hz	Range 10 Hz
Lowest prediction accuracy (all classes)	88.33%	88.33%	35.00%	70.83%	38.33%
Highest prediction accuracy (all classes)	100.00%	100.00%	94.17%	100.00%	77.50%
Lowest individual faults prediction	76.67% (15 Hz, MT case)	76.67% (12.5 Hz, ND case)	33.33% (15 Hz, ND case)	43.33% (15 Hz, CT case)	13.33% (25 Hz, ND case)
Highest individual faults prediction	100.00%	100.00%	100.00%	100.00%	100.00%

Initially for each of the four classification cases (CT, MT, WT and ND), the training data was provided at running speeds from 10 Hz to 30 Hz in intervals of 2.5 Hz. Then the multiclass classification capability of two classes of SVM was noted for these running speeds. It is concluded that the SVM has the ability to make perfect classifications if the training data is available for that particular running speed. The classification capability of the SVM for the interpolation and the extrapolation beyond its training data speeds are performed and it is noted that the SVM can still be able to make accurate classification utilizing the training data

from the beginning and end of a range of running speeds, whereas the test data could be at any of speeds from within that range (interpolation) and out of range (extrapolation). In the interpolation and extrapolation fault predictions, it has been observed that in many cases the GA and the ABCA show similar results as the GSM. On the other hand, the GSM shows its marginal superiority over the GA and ABC results with order of 12.5 %, 1.16% and 0.83% more for three cases. It is also observed that the prediction accuracy gradually increases with the increase of the rotational speed. This is due to the high signal-to-noise level at higher rotation speeds due to better manifestation of faults in vibration signals. It is also observed that interpolation accuracy is better than the extrapolated one. It is know that the classifier is tested for classification ability within the range and outside the range of speeds. The classifier distinguishes the data on the basis of the classification boundary created by support vectors during the training. In the intermediate speed, the fault frequency and its sidebands are within two trained data fault frequency and its sidebands. It may better influence fault features to be tested than the extrapolated one, in which fault frequencies and its sidebands are outside the trained data fault frequencies and its sidebands.

4.4 Summary

It has been established that the SVM have a capability to classify gear faults at the same rotational speed, but it is rare to see the classification ability to classify gear faults at the intermediate and extrapolated speeds, at which the data is not available. Taking into consideration this point, this chapter examines the SVM prediction performance for the classification of gear faults by training the classifier at the beginning and at end of rotational speeds of a speed range, and classified faults at the intermediate and outside rotational speeds of that range. In the case of training and testing (by time domain features) at same rotational speed a near perfect prediction accuracies are found. On the other hand the training at beginning and at end of the rotational speed of a speed range, and testing at the intermediate

and extrapolated rotational speeds, the prediction accuracy is still encouraging. The potential of two SVM classifiers (C-SVC and ν-SVC) with respect of its tuning of parameters by using two optimizing tools, i.e. the GA and the ABCA, and their performance are studied and compared with the classical optimization technique called the GSM. The GA and ABCA based optimization show its effectiveness at several cases. Experimental fault predictions that have been shown in this study are based on the RBF kernel and time domain vibration data. In the following chapter the multiclass ability using the frequency domain data will be studied.

CHAPTER 5

Multiclass Classification using Frequency Domain data

5.1 Introduction

There are three main advantages in spectral or frequency analysis as compared to the time domain analysis. The first is that it simplifies the understanding of the waveform. Secondly, the physical property of the signal, for example its propagation through a medium often depends on the frequency, and thirdly it is a mathematical tool for solving equations (Cohen, 1995). The first two reasons make spectral analysis a very significant tool for the condition monitoring signal analysis, which is of interest to this study. The third reason is that convolution's presenting a computational convenience multiplication in time-domain corresponds to convolution in frequency domain and vice versa. Any time series equation can be solved in frequency domain conveniently, and then the result can be back to time domain by utilizing forward and inverse Fourier transform. An overall vibration time signal is usually dominated by a few major frequency components such as machine rotation speed, gear tooth meshing frequency, natural frequencies of structures, etc. These dominant frequencies are important to monitor and traditional it is being used for long.

On taking on this perception the classification ability of the SVM classifier using frequency domain data are now discussed in this chapter. The acquisition of frequency domain data set from the healthy and faulty gear box has been discussed in Chapter 2. In this chapter also statistical features are extracted from frequency domain data set instead of picking up only few predominate peaks. These statistical features are expected to give overall features of the whole frequency domain data instead of limited frequency spikes. Extracted frequency domain features are used for the classification in the SVM as discussed in Chapter 3. The classification of faults at same rotational speed as well as at the intermediate and out of range is documented

and results are discussed. Fault classification accuracies documented here shows better prediction accuracies than found from time domain data set in Chapter 4.

5.2 Statistical Feature Extraction

Three statistical features (i.e., the standard deviation, skewness and kurtosis) are extracted from frequency domain vibration signals, which are already acquired and stored as discussed in Chapter 2. The typical plot for frequency domain responses at 30 Hz rotational speed for CT gear faults is illustrated in Figure 2.11(b), in Chapter 2. Three statistical features are calculated from the magnitude of sampled signal $\{ s_1, s_2, \ldots, s_k \}$ in frequency domain as follows,

$$S_d = \sqrt{\frac{n\sum s_i^2 - \left(\sum s_i\right)^2}{k(k-1)}} \tag{5.1}$$

$$S_k = \frac{k}{(k-1)(k-2)} \sum \left(\frac{s_i - \bar{s}}{S_d} \right)^3 \qquad \text{with} \quad \bar{s} = \sum s_i / k \tag{5.2}$$

$$K_u = \left\{ \frac{k(k+1)}{(k-1)(k-2)(k-3)} \sum \left(\frac{s_i - \bar{s}}{S_d} \right)^4 \right\} - \frac{3(k-1)^2}{(k-2)(k-3)} \tag{5.3}$$

where the summation is implied that it is overall index values, i.e. i = 1, 2, …, k. Three statistical features are calculated from 1,000 samples in three orthogonal directions. The calculation is performed to 300 sets of data (which is collected for different rotational speeds as explained in Chapter 2) in three orthogonal directions. In other words, for three statistical features in each three directions (i.e., total 9 combinations), which means 9×300 data sets are available for each faults. For four different fault conditions (i.e., CT, MT, WT and ND) there

are total $9\times4\times300$ data sets available. These total data sets are divided and used for the training, testing of optimizing the parameters and for the final testing in the simulation. Figure 5.1 illustrates statistical features extracted from frequency domain vibration signals in three axes for the CT gear fault at 30 Hz rotational speed.

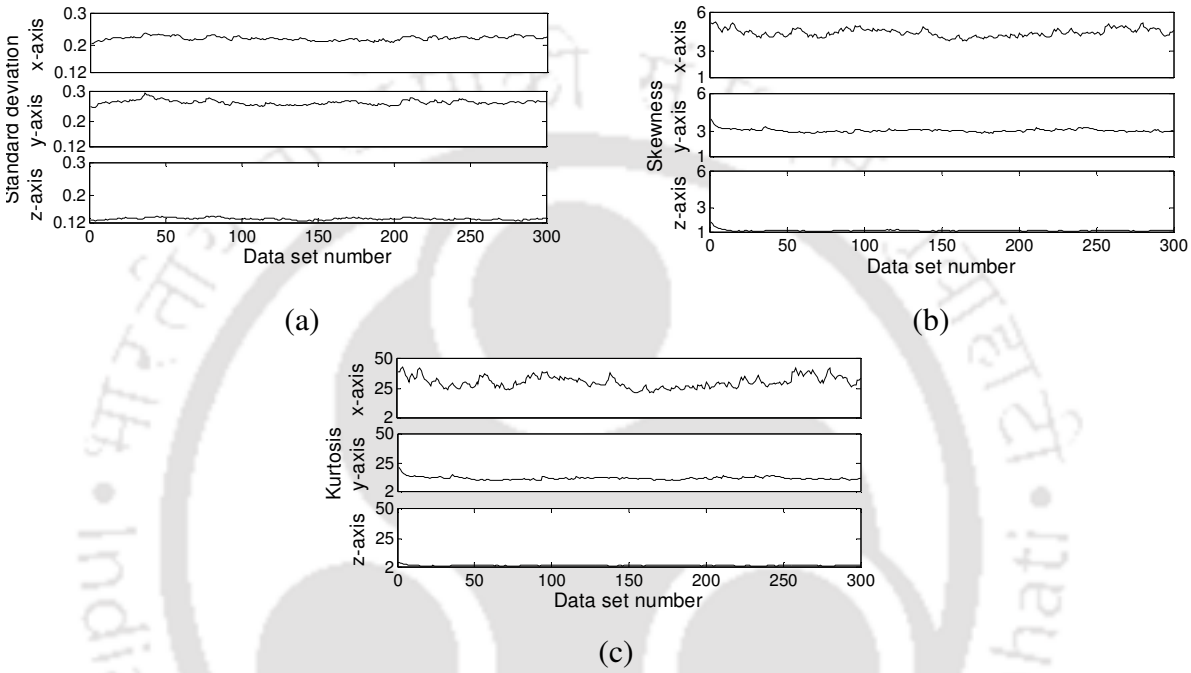


Figure 5.1 Frequency domain features of acquired signals for CT at 30 Hz rotating speed

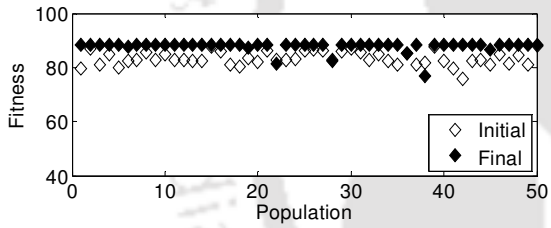
5.3 Simulation Results

The SVM software first needs to be trained with features drawn from the vibration data comprising of four types of gears (i.e., CT, MT, WT and ND) and it has been discussed in Chapter 3 in detail. For the testing and the training of the SVM software the following three conditions are considered.

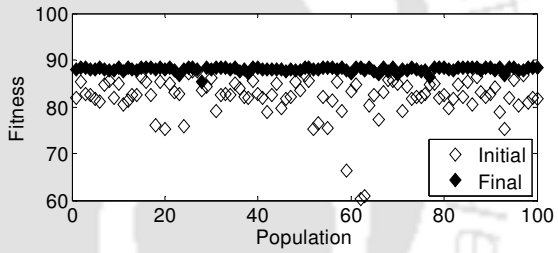
5.3.1 Training and Testing at the Same Rotational Speed

In this case the SVM is trained and tested with frequency domain statistical features at the same rotating frequency. Rotational speeds selected are 10-30 Hz in a step of 5 Hz.

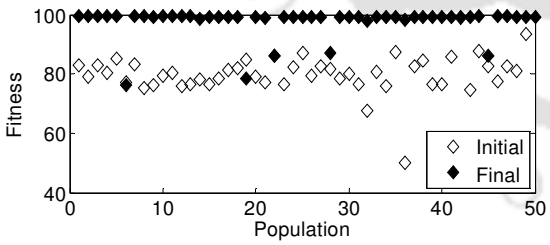
Optimization of Parameters: During the optimization of SVM parameters by the GA and the ABCA, $9\times4\times180$ data set is used for the training of the fault classification, and $9\times4\times90$ data sets are used for the testing. Then $9\times4\times30$ data sets are used for the final testing of the fault classification. In the case of GSM $9\times4\times270$ data sets are used for the parameter estimation and $9\times4\times30$ data sets are used for the final testing of the fault classification. The variation of the initial and final fitness values (i.e., the percentage accuracy) with the population for the *C*-SVC using the GA and the ABCA for 30 Hz rotational speeds are shown in Figure 5.2(a) and Figure 5.2(b), respectively. Similarly, the variation of the initial and final fitness with the population for the ν-SVC using the GA and the ABCA for 30 Hz rotational speed are shown in Figure 5.2(c) and Figure 5.2(d), respectively.



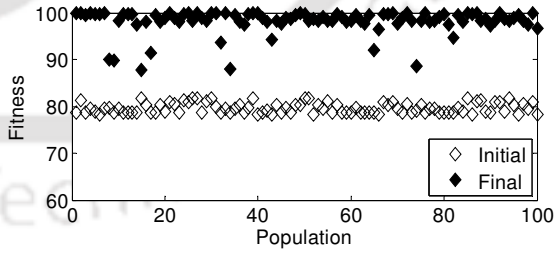
(a) 30 Hz (*C*-SVC using GA optimization)



(b) 30 Hz (*C*-SVC using ABCA optimization)

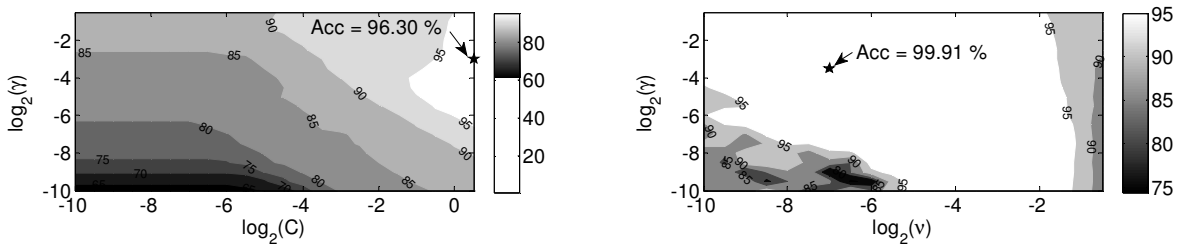


(c) 30 Hz (ν-SVC using GA optimization)



(d) 30 Hz (ν-SVC using ABCA optimization)

Figure 5.2 Variation of the initial and final fitness (percentage accuracy) with the population (a)-(b) for the *C*-SVC and (c)-(d) for the ν-SVC using the GA and ABCA for 30 Hz training and testing speed



(c) 30 Hz (*C*-SVC) (d) 30 Hz (ν-SVC)

Figure 5.3 Cross validation accuracy (a) for the *C*-SVC (b) for the ν-SVC for 30 Hz training and testing speed

It includes all the population, including solutions violating constraints. From the choice of initial solutions, it could be seen that it has quite good divergence of the possible solution domain. In the last generation also all solutions have been shown including those violating constraints (however, very few), and it could be observed that the most of solutions have converged.

The cross validation accuracy in the GSM for 30 Hz rotational speed for the *C*-SVC and the ν-SVC is shown in Figure 5.3(a–b), respectively. In these the contour line for the percentage accuracy is plotted and the best CV accuracy is marked. All CV values are stored with the progress of calculation in a stack. If more than one best (similar) CV is found, then any one of corresponding parameters (i.e., logarithmic values of SVC and kernel parameters) are selected. Correspondingly the best SVC parameters are found from the best CV accuracy and the testing accuracy is tabulated. The optimized percentage (all classes) accuracy of different SVM formulation is shown in Table 5.1.

Fault Prediction Ability: After optimization of parameters $9\times4\times30$ data sets are used for the final testing of the fault classification for the GA and the ABCA, and $9\times4\times30$ data sets are

used for the final testing of the fault classification for the GSM. Number of data points used for optimization and final testing is tabulated in Table 4.3. In many times the accuracy is more in the GA and the ABCA as compared with the GSM, which reflect the soundness of the GA and the ABCA. If we look upon the testing accuracy (all classes), the lowest one is the 88.33% that occurs at 30 Hz rotational speeds on the *C*-SVC case. It is also observed that at many rotational speeds the prediction accuracy (all classes) is a near perfect (i.e., at 10 Hz, 20 Hz and 30 Hz). At 25 Hz rotational speed the accuracy (all classes) is 97.50%, on the other hand at 15 Hz rotational speed, the CV accuracy (i.e. 97.58%) slightly more than the GA and ABCA accuracy (i.e. 97.50%), which is only 0.08% more. Table 5.3 illustrates the percentage prediction (individual classes) at various rotational speeds against the best fault prediction. The prediction accuracy of 90.00% is the individual lowest for the CT gear fault case at 25 Hz rotational speed.

Table 5.1 Classification of gear faults (all classes) with the training and the testing at same rotational speed (frequency domain features)

Training speed (Hz)	Testing speed (Hz)	SVM formulation	Training accuracy* (%)			Testing accuracy (%)		
			CV	GA	ABCA	CV	GA	ABCA
10	10	*C*-SVC	99.63	100.00	100.00	**100.00**	**100.00**	**100.00**
		ν -SVC	99.72	100.00	100.00	**100.00**	**100.00**	98.33
15	15	*C*-SVC	99.26	98.33	98.33	96.77	97.50	97.50
		ν -SVC	99.41	98.33	98.89	**97.58**	96.67	96.67
20	20	*C*-SVC	97.13	93.33	93.33	94.35	98.33	98.33
		ν -SVC	100.00	99.17	100.00	**100.00**	**100.00**	**100.00**
25	25	*C*-SVC	98.24	96.67	96.67	96.77	95.00	92.50
		ν -SVC	99.91	100.00	100.00	**97.50**	**97.50**	96.77
30	30	*C*-SVC	96.30	88.33	88.33	87.90	88.33	88.33
		ν -SVC	99.91	99.44	100.00	96.77	**100.00**	**100.00**

*CV means the *cross validation* accuracy of grid search technique. Bold values represent prediction accuracies of the maximum accuracy.

5.3.2 Training at two Different Rotational Speeds and Testing at an Intermediate Rotational Speed

In this case, the SVM is trained at two rotational speeds and tested at an intermediate rotational speed for which the SVM was not trained. Values of rotational speed ranges and intermediate testing rotational speeds are tabulated in Table 4.2. In this case also, beyond 10 Hz (i.e., 600 rpm) rotational speed range is not considered, because the speed range is wide and hence the percentage accuracy becomes low.

Optimization of Parameters: Here $9\times8\times90$ data sets are used for the training and $9\times8\times45$ data sets are used for the testing of optimized parameters for both the GA and the ABCA. For example, in the case of optimization of parameters for the first range, $9\times4\times90$ data sets are taken from 10 Hz rotational speed and $9\times4\times90$ data set is taken from 15 Hz rotational speed so that all together $9\times8\times90$ data sets are taken for the training. Whereas, $9\times4\times45$ data sets are taken from 10 Hz rotational speed and $9\times4\times45$ data sets are taken from 15 Hz rotational speed so that all together $9\times8\times45$ data sets are taken for the testing. After the optimization, $9\times4\times30$ data sets from intermediate rotational speeds are used for the final testing or classification. Similarly, $9\times8\times125$ data sets are used for the parameter estimation and $9\times4\times30$ data sets are used for the final testing for the GSM. Data sets used for optimizing and final testing is tabulated in Table 4.3. The optimized percentage accuracy (all classes) of different SVM formulations is shown in Table 5.2.

Fault Prediction Ability: At the intermediate rotational speed range of 5 Hz, for the lowest rotational speed the classification accuracy (all classes) is nearly 63.33% and it increases at higher rotational speeds to 90%. In the case of intermediate rotational speed range of 10 Hz, for the lowest rotational speed the classification accuracy (all classes) is nearly 52.50% and it increases at high rotational speed to 80.00%. It is observed that the GA and ABCA techniques

have shown the good prediction accuracy (all classes) at 20–25, 25–30 and 20–30 Hz range with the conventional GSM, but at a particular speed range 10–15, 15–20, 10–20, 15–25 Hz, the grid search technique shows good result (which has 3.33%, 1.66%, 0.83%, 5.83% more accuracy level then the GA and ABCA techniques). This may be due to the former method is less sensitive to noise as compared to latter ones. The percentage prediction (individual classes) at various rotational speeds based on the best optimized condition is illustrated in Table 5.3. In the case of rotational speed range of 5 Hz, the classification for different faults for individual classes is more than 63.33%. But in the case of the rotational speed range of 10 Hz prediction accuracy at 20 Hz rotational speed, the individual classification for fault conditions shows very low fault prediction and that is down to 16.67%. Though, fault prediction accuracies (individual classes) are not perfect (varying from 68.33% to 90%), still there is a reasonable prediction of faults at the intermediate speed, with the lower and upper speed training data. Individual faults classification is good in a narrow range of rotational speed.

5.3.3 Training at two Different Rotational Speeds and Testing at an Extrapolated Rotational Speed

In this case, the SVM is trained at two rotational speeds and tested at an extrapolated rotational speed. Training rotational speeds and the extrapolated testing rotational speed are tabulated in Table 4.2.

Optimization of Parameters: For selecting data points the same procedure is considered as used for the intermediate interpolation, but only difference is that for the testing an extrapolation speed is considered. The optimized percentage accuracy (all classes) of different SVM formulations is shown in Table 5.2.

Table 5.2 Classification of gear faults (all classes) with the training and the testing at the interpolation and extrapolation speeds (frequency domain features)

Training speed (Hz)	Testing speed (Hz)	SVM formulation	Training accuracy* (%)			Testing accuracy (%)		
			CV	GA	ABCA	CV	GA	ABCA
For interpolation speed (range 5 Hz)								
10, 15	12.5	*C*-SVC	98.89	100.00	100.00	**90.00**	84.17	86.67
		ν-SVC	99.35	100.00	100.00	80.00	85.00	85.83
15, 20	17.5	*C*-SVC	97.31	91.58	91.94	84.17	73.33	71.67
		ν-SVC	99.72	98.64	99.44	**88.33**	86.67	85.83
20, 25	22.5	*C*-SVC	96.20	91.85	92.22	67.50	63.33	64.17
		ν-SVC	100.00	97.83	98.89	75.83	79.17	**84.17**
25, 30	27.5	*C*-SVC	94.17	86.96	88.61	73.33	65.83	**82.50**
		ν-SVC	99.72	100.00	100.00	68.33	76.67	75.00
For interpolation speed (range 10 Hz)								
10, 20	15	*C*-SVC	98.15	93.06	93.33	66.66	52.50	54.17
		ν-SVC	99.72	100.00	100.00	**68.33**	67.50	61.67
15, 25	20	*C*-SVC	96.85	88.33	88.33	66.67	64.17	64.17
		ν-SVC	99.35	88.33	88.61	**70.83**	65.00	64.17
20, 30	25	*C*-SVC	95.56	86.39	86.39	74.17	74.17	74.17
		ν-SVC	99.81	97.50	97.50	79.17	74.17	**80.00**
For extrapolation speed (range 5 Hz)								
10, 12.5	15	*C*-SVC	99.17	100.00	100.00	68.33	74.17	**76.67**
		ν-SVC	99.72	100.00	100.00	68.33	75.83	73.33
20, 22.5	25	*C*-SVC	96.76	91.94	91.94	69.17	71.67	70.00
		ν-SVC	99.91	99.17	100.00	**75.00**	71.67	72.50
25, 27.5	30	*C*-SVC	96.76	96.67	96.67	**90.83**	80.83	79.17
		ν-SVC	99.72	97.22	97.78	88.33	80.00	81.67
For extrapolation speed (range 10 Hz)								
10, 15	20	*C*-SVC	98.89	100.00	100.00	49.17	54.17	54.17
		ν-SVC	99.35	100.00	100.00	52.50	56.67	**58.33**
15, 20	25	*C*-SVC	97.31	91.67	91.94	65.00	67.50	67.50
		ν-SVC	99.72	98.61	99.44	75.00	75.00	**76.67**
20, 25	30	*C*-SVC	96.20	92.22	92.22	68.33	70.83	70.83
		ν-SVC	100.00	98.06	98.89	69.17	**74.17**	72.50

* CV means the *cross validation* accuracy of the GSM. Bold values represent prediction accuracies of the maximum accuracy.

Table 5.3 Percentage fault predictions (individual classes) at various speeds with the training and the testing (frequency domain features)

Speed (Hz)	CT (%)	MT (%)	WT (%)	ND (%)
For same speed				
10	100.00	100.00	100.00	100.00
15	100.00	90.32	100.00	100.00
20	100.00	100.00	100.00	100.00
25	90.00	100.00	100.00	100.00
30	100.00	100.00	100.00	100.00
For interpolation speed (range 5 Hz)				
12.5	100.00	100.00	83.33	76.67
17.5	100.00	70.00	86.67	96.67
22.5	73.33	100.00	63.33	100.00
27.5	50.00	100.00	86.67	93.33
For interpolation speed (range 10 Hz)				
15	60.00	76.67	73.33	63.33
20	93.33	16.67	76.67	96.67
25	100.00	76.67	46.67	96.67
For extrapolation speed (range 5 Hz)				
15	100.00	50.00	76.67	80.00
25	73.33	53.33	83.33	90.00
30	76.67	100.00	86.67	100.00
For extrapolation speed (range 10 Hz)				
20	96.67	0.00	86.67	50.00
25	100.00	16.67	90.00	100.00
30	93.33	63.33	56.67	83.33

Fault Prediction Ability: At the extrapolated speed range of 5 Hz, the prediction accuracy (all classes) is from 68.33% to 90.83%, and in case of 10 Hz range it is from 49.17% to 76.67%. It is observed that at 25 Hz and 30 Hz testing speeds for the case of 5 Hz range the GSM shows marginally good results, which is 2.50% and 9.16% more than the GA and ABCA techniques, respectively. On the other hand for the 15 Hz testing speed the ABCA dominates. For the case

of 10 Hz range the GA and the ABCA show good performance. The best percentage fault predictions (individual classes) at various rotational speeds are illustrated in Table 5.3. In the case of speed range of 5 Hz, the lowest prediction accuracy (individual classes) is at 15 Hz for the fault detection of MT (i.e., 50%). Similarly, in the case of speed range of 10 Hz, the lowest performance is for the MT condition at 20 Hz (i.e., 0%). Prediction accuracies (all classes) at the extrapolated speed beyond its lower and upper training speeds at the low range are good but at the high range it is average. Individual classifications of faults also follow the intermediate range predictions. For more clarity the overall performances are tabulated in Table 5.4. The lowest prediction accuracy (all classes) and highest prediction accuracy (all classes) are the lowest and highest testing accuracy (all classes), respectively, as per Table 5.1 and Table 5.2. The lowest and highest fault predictions are found from Table 5.3.

Table 5.4 Summary of overall fault prediction performances (frequency domain features)

	Same rotational speed	Interpolation speed		Extrapolation speed	
		Range 5 Hz	Range 10 Hz	Range 5 Hz	Range 10 Hz
Lowest prediction accuracy (all classes)	88.33%	63.33%	52.55%	68.33%	49.17%
Highest prediction accuracy (all classes)	100.00%	90.00%	80.00%	90.83%	76.67%
Lowest individual faults prediction	90.00% (25 Hz, CT case)	50.00% (27.5 Hz, CT case)	16.67% (20 Hz, MT case)	50.00% (15 Hz, MT case)	0.00% (20 Hz, MT case)
Highest individual faults prediction	100.00%	100.00%	100.00%	100.00%	100.00%

Table 5.5 Comparison of prediction percentage accuracies (all classes) in time and frequency domain data

Speed (Hz)	Testing accuracy (all classes) from time domain data (%)	Testing accuracy (all classes) from frequency domain data (%)
For same speed		
10	99.17	**100.00**
15	90.83	**97.50**
20	98.33	**100.00**
25	**100.00**	97.50
30	**100.00**	**100.00**
For interpolation speed (range 5 Hz)		
12.5	89.17	**90.00**
17.5	**94.17**	88.33
22.5	**96.67**	84.17
27.5	**100.00**	82.50
For interpolation speed (range 10 Hz)		
15	59.17	**68.33**
20	**85.83**	70.83
25	**94.17**	80.00
For extrapolation speed (range 5 Hz)		
15	75.83	**76.67**
25	**90.83**	75.00
30	**100.00**	90.83
For extrapolation speed (range 10 Hz)		
20	**62.50**	58.33
25	69.17	**76.67**
30	**77.50**	74.17

Initially for each of the four classification cases (CT, MT, WT and ND), the training data was provided at running speeds from 10 Hz to 30 Hz at the interval of 2.5 Hz. Then the multiclass

classification capability of two classes of the SVM was noted for these running speeds. It is concluded that the SVM has the ability to make perfect classifications if the training data is available for that particular running speed. The prediction of faults at the intermediate and extrapolated rotational speeds by the training data at two extremes of a speed range is demonstrated. It is found that the SVM can be used in these two cases along with at the same rotational speed. Altogether fault predictions are studied for 5 different speeds for the same speed case, 7 different speed ranges for the interpolation and 6 different speed ranges for the extrapolation. In fault predictions (all classes), it has been observed that 4 times out of 5 cases at the same speed, 3 times out of 7 cases in interpolation case and 4 times out of 6 cases in extrapolation case; the GA and ABCA shows similar or better fault predictions with respect to the GSM. On the other hand, the GSM shows its marginal superiority greater than 0.08% for the same speed fault prediction (all classes) at 15 Hz speed. In the case of interpolation testing speed at 12.5 Hz (range of 5 Hz), 17.5 Hz (range of 5 Hz), 15 Hz (range of 10 Hz) and 20 Hz (range of 10 Hz), the GSM prediction accuracies (all classes) are more than 3.33%, 1.66%, 0.83% and 5.83%, respectively, as compared to the GA or ABCA. Similarly, in the case of extrapolation testing speed at 25 Hz (range of 5 Hz) and 30 Hz (range of 5 Hz), the GSM prediction accuracies (all classes) are more than 2.5% and 9.16%, respectively, as compared to the GA or ABCA.

Table 5.5 illustrates the comparison of the best prediction accuracies (all classes) found using the time and frequency domain data set. The bold form of number in the table indicates the best at that particular speed from either time domain or frequency domain data. It has been notice in the table that at the same speed case for all cases frequency domain fault prediction (all classes) is better except at 25 Hz speed; in that case time domain data shows 2.5% more than frequency domain case. In the case of interpolation speed with time domain data 5 cases (out of 7 cases) show the superiority as compared to frequency domain data. In this case, the accuracy (all

classes) is 0.83% and 9.16% more than in frequency domain with respect to time domain at 12.5 Hz and 15 Hz speed range, respectively. For the case of extrapolation speed with time domain data 4 cases (out of 6 cases) show better results compared to frequency domain. In this case, the accuracy is 0.84% and 7.5% more in 15 Hz and 25 Hz speed range, respectively.

5.4 Summary

In this chapter an attempt is made to study the fault diagnosis method based on the SVM for the classification of gear faults by training the classifier using the frequency domain data. In the case of training and testing at the same rotational speed a near perfect fault predictions are found.

On the other hand the training at the beginning and at the end of the rotational speed of a range of rotational speed, and testing at the intermediate and extrapolated rotating speeds, the fault prediction is reasonable. The potential of two SVM classifiers (C-SVC and ν-SVC) with respect of its tuning of parameters by using two optimizing tool the GA and the ABCA, and their performance have been studied and compared with the classical GSM technique. The GA and ABCA based optimization methods show its effectiveness compared with the conventional GSM at many occasions. Frequency domain data shows better fault prediction results at the same rotational speed for several cases, and its fault predictions at the interpolation and extrapolation speed cases are encouraging. In the next chapter, now the multiclass ability using time-frequency domain analysis will be studied.

CHAPTER 6

Multiclass Classification using Time-Frequency Domain data

6.1 Introduction

It has been quite difficult to satisfactorily handle signals carrying non-stationary or transient components, such as a small or short lasting vibrations emerging as chirping sounds, using conceptualizations based on the stationary. The production of particular frequencies depends on physical parameters, which may change in time possibly due to an incipient failure originating from various causes. Therefore, examining local behavior of the vibration signal with reasonably precise frequency information may be useful to interpret the signal in a better way. The aim of time-frequency signal analysis is to describe how the frequency or spectral content of a signal evolves and to develop the physical and mathematical ideas needed to understand what a time varying spectra is (Cohen, 1995).

A relatively new method has been developed in the signal processing recently called the wavelet transform. Although the wavelet theory has been studies for many years by mathematicians (before the word wavelet was even coined), Haar (1911), Franklin (1928), Calderon and Zygmund (1952), an explosion of applications have been seen in engineering since the mid-eighties. This resurgence was initiated by Morlet *et al.* (1982), in geo-exploration studies as a technique for the analysis of seismic signals, and was followed by a detailed mathematical analysis by Grossmann and Morlet (1984) and Rioul and Vetterli (1991).

Faults in gears produce transient responses during impacts. Staszewski and Tomlinson (1994) applied wavelet transform to fault detection in a spur gear using similarity analysis of patterns featured by the wavelet transform coefficients. It was shown that Mahalanobis distance between contour plots of the modulus of the wavelet transform increases monotonically with gear fault advancement, which can be used as a fault detection symptom, but there is not any

regularity for the modulus itself with the increase of gear fault advancement. To remove this deficiency people used the de-noising methods based on Morlet and Maxican hat in CWT and WPT (Lin and Qu (2000), Dalpiaz et al. (2000), Wang et al. (2001), Saravana et al. (2008)) is used. Statistical parameters are used to have overall properties in classification instead of selecting few of them based on some criteria.

In previous two chapters, classifications of gear faults using the time and frequency domain have been discussed. It has been noticed that there is a scope to investigate the classification of gear faults by using the time-frequency data set. In Chapters 2 and 3, the collection of time domain data from the healthy and faulty gears with the classification methodology is discussed. In this chapter, time domain data is used for extracting time-frequency signal components. The extraction of time-frequency data set (wavelet) from the time domain data is discussed here. A group of statistical features like the kurtosis, the standard deviation and the skewness have been extracted from the wavelet transform data. The optimization of SVM parameters has been performed by the GSM, GA and ABCA techniques before final training and testing of the classifier. Initially, the training and testing data have been selected at the same rotational speed for a range of rotational speeds; subsequently, the training was performed at two rotational speeds and the testing at other than the trained ones, i.e. at the intermediate and extrapolated rotational speeds. This has a practical advantage in that it is not feasible to have measured data at all speeds of interest.

6.2 Feature Extraction

The continuous wavelet transform (CWT) and wavelet packet transform (WPT) are used for obtaining the wavelet coefficients from time domain vibration signal. These have been described here briefly.

6.2.1 CWT Based Feature Extraction

The CWT is a crucial tool to extract features that indicates both the time and frequency nature of the waveform. Vibration signals measured in time domain from the accelerometer were in three orthogonal directions. The CWT is applied on each of these signals with *scales* selected from 1 to 128, and the wavelet to be wavelet families. The CWT is defined (Mallat, 1998) as the sum of overall time of the signal multiplied by scaled and shifted versions of the wavelet function, as

$$W_s(t) = \int_{-\infty}^{+\infty} f(t)\psi_{s,j}(t)\,dt \tag{6.1}$$

where

$$\psi_{s,j}(t) = \frac{1}{\sqrt{|s|}}\psi\left(\frac{t-T}{s}\right)$$

is a *window function* called the *mother wavelet*, *s* is a *scale* and *T* is a *translation* means the location or position of windows. The term translation is associated to the location of the window, as the window is shifted through the signal. This corresponds to the time information in transform domain. But instead of a frequency parameter, we have a scale. Scaling, as a mathematical operation, either dilates or compresses a signal. A smaller scale represents to the high frequency of signals and a large scale represents to the low frequency signals. Figure 6.1 illustrates the plot of CWT coefficients for a typical vibration signal on the scale and time (or space) axes. In the plot, the abscissa represents position along the signal (time or space), the ordinate represents the scale, and the colour at any point represents the magnitude of the CWT coefficients. The higher scales correspond to the most "stretched" wavelets. The more stretched the wavelet, the longer the portion of the signal with which it is being compared, and thus the coarser the signal features being measured by CWT coefficients.

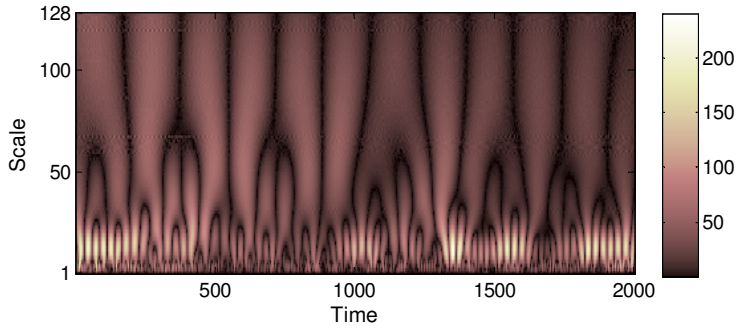


Figure 6.1 Plot of CWT coefficients for a typical vibration signal

There is a correspondence between CWT scales and frequency as revealed by wavelet analysis as is summarized in Table 6.1. The wavelet series is simply a sampled version of the CWT, and the information it provides is highly redundant as far as the reconstruction of the signal is concerned. This redundancy, on the other hand, requires a significant amount of computation time and resources. Calculating CWT coefficients at all potential scale is a huge amount of effort, and it generates a tremendous quantity of data. A skilled way to implement this method using filters was developed by Mallat (1989).

Table 6.1 Correspondence between CWT scales and frequency by wavelet analysis

Low scale *s*	Compressed wavelet	Rapidly changing details	High frequency
High scale *s*	Stretched wavelet	Slowly changing, coarse features	Low frequency

The continuous wavelet one-dimensional function, available in MATLAB™ is chosen with the Mexican hat and Morlet wavelets. The Mexican hat and Morlet are defined, respectively, as

$$\psi(t) = \left(\frac{2}{\sqrt{3}}\pi^{-1/4}\right)\left(1-t^2\right)e^{-t^2/2} \tag{6.2}$$

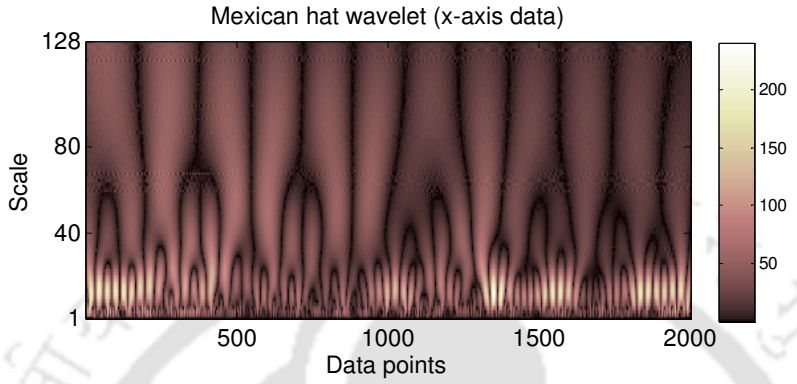
and

$$\psi(t) = Ce^{-s^2/2}\cos(kt) \tag{6.3}$$

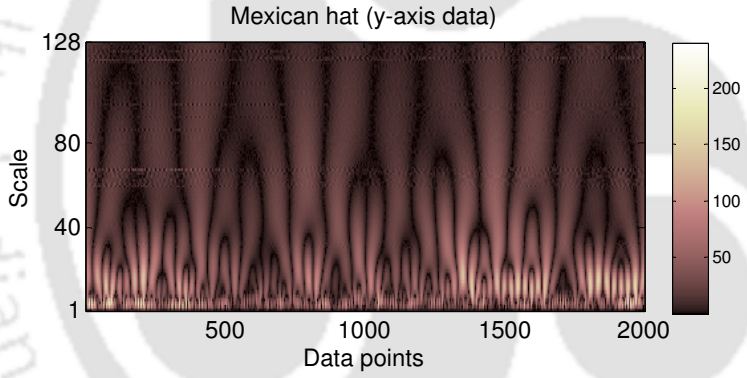
where C is a constant used for normalization in view of reconstruction and k is the wave number. It returns CWT coefficients of signal x at scale s (Soman *et al*., 2005). Typical patterns of CWT coefficients produced for the CT gear fault case at 30 Hz rotational speed is shown in Figure 6.2(a)-(c) for the Mexican hat and in Figure 6.3(a)-(c) for the Morlet wavelet family in three orthogonal directions, respectively.

Since all data values consist of CWT coefficients for different scales and the amount of data to be handled will be exceptionally high. Hence, it will be beneficial if only CWT coefficients corresponding to the best scale is chosen for the further study. In Figure 6.2 and Figure 6.3, for a given data set (data points from 1 to 2000) and all scales (from 1 to 128) the CWT coefficients have been obtained. This forms a grid of 2000×128 CWT coefficients for that particular data set and has been shown in the form of patterns in Figure 6.2 and Figure 6.3. For a particular data point with 1 to 128 scales, 1×128 number of CWT coefficients is found. For data point 1 (in horizontal axis) corresponding to all scales (in vertical axis) the CWT coefficient with the maximum value (darkest colour) is selected and the corresponding chosen scale is assigned to the data 1 and it is denoted as s_1, here subscript refers to the data point. Similarly, for the data point 2 the scale s_2 is chosen. This process is repeated till for all data points (up to 2000) and corresponding scales (up to s_{2000}) are chosen. It should be noted that chosen scales, s_i, may repeat (since it can take values from 1 to 128 only for 2000 data points) and some scale from 1 to 128 may not be chosen at all. Figure 6.4 (a) and (b) shows variations of the chosen scale for 2000 data points (for data sets in x, y and z directions) for the Mexican hat and Morlet wavelet scales, respectively. For example, the best scale 4 (with CWT

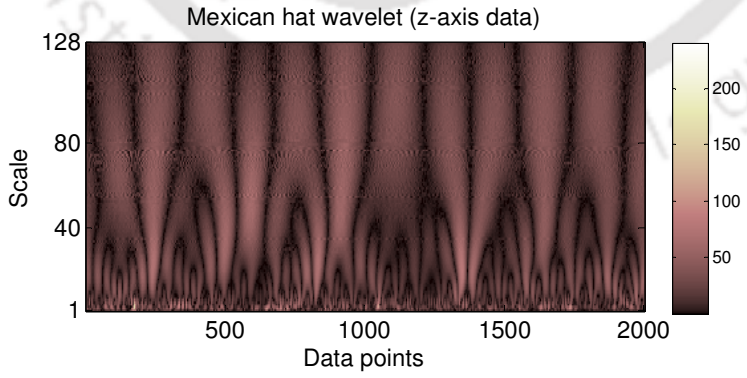
coefficient 9.5591) is chosen for the first data point in *x*-direction (as shown in Figure 6.4 (a)) for the Mexican hat wavelet family. Similarly, the best scale 13 (with CWT coefficient 17.553) is chosen for the first data point in *x*-direction (shown in Figure 6.4 (b)) for the Morlet wavelet family.



(a)

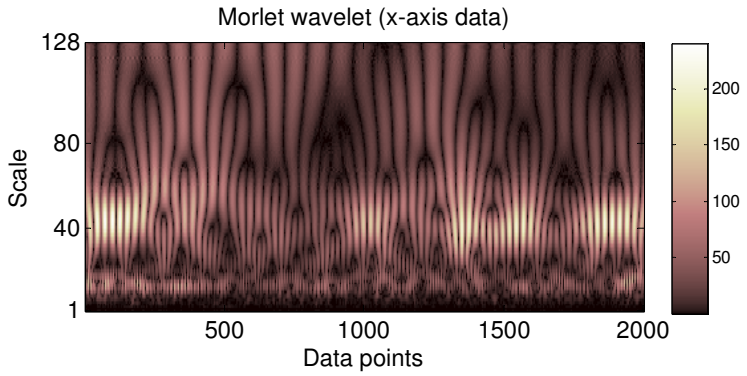


(b)

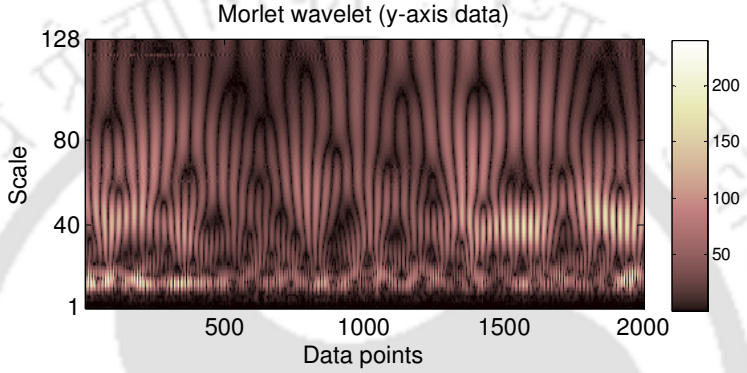


(c)

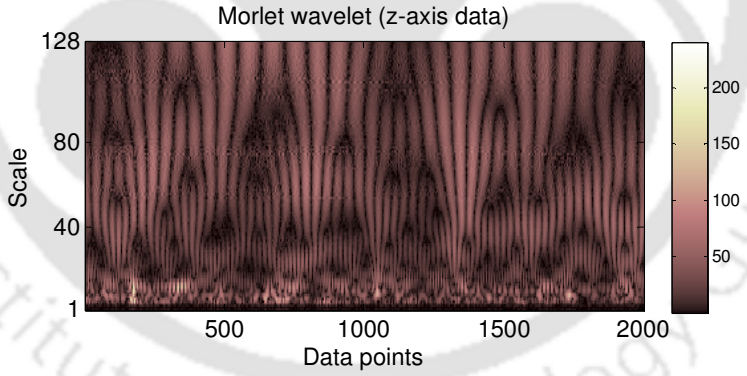
Figure 6.2 Patterns produced for CT at 30 Hz rotating speed by the Mexican hat in (a) *x*-direction (b) *y*-direction and (c) *z*-direction



(a)



(b)

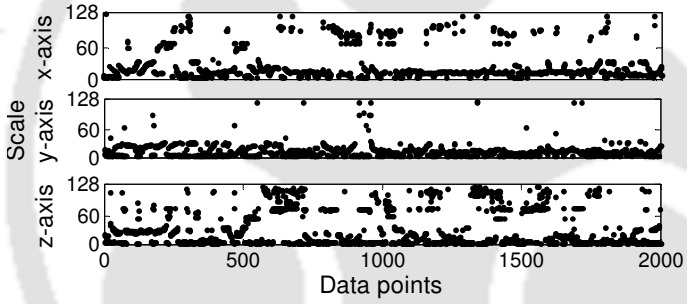


(c)

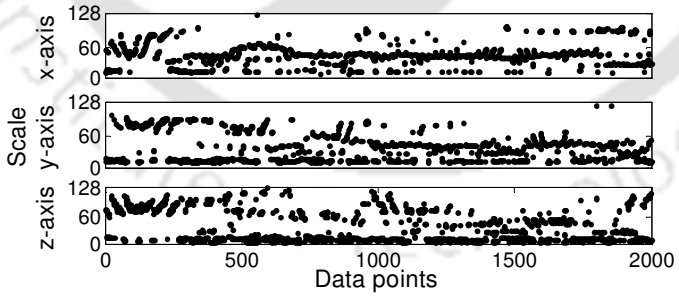
Figure 6.3 Patterns produced for CT at 30 Hz rotating speed by the Morlet wavelet family in (a) *x*-direction (b) *y*-direction and (c) *z*-direction

Now for each scale from 1 to 128, how many numbers of times it has been chosen, is counted. Based on this an *efficiency* is defined, which is calculated as the number of times a particular scale (i.e., one of them from 1 to 128) is chosen with respect to total number of chosen scales

(i.e., 2,000). The scale with the highest efficiency for that data set would be the one which is chosen the most number of times. This scale with the highest efficiency will be the best scale for a particular data set and it will be used for obtaining all CWT coefficients for that particular data set for further study. For example as shown in Figure 6.5 (a), for the first data set (*x*-direction) corresponding to the scale of 14 the best efficiency of 9.30% is chosen for the Mexican hat wavelet family for *x*-axis data. Similarly, for the same data set the best efficiency of 6.65% corresponding to scale of 42 is chosen (shown in Figure 6.5 (b)) for the Morlet wavelet family for *x*-axis data. Similarly, optimum scales can be obtained for all data sets from 1 to 300 and orthogonal axis, and corresponding CWT coefficients are obtained.



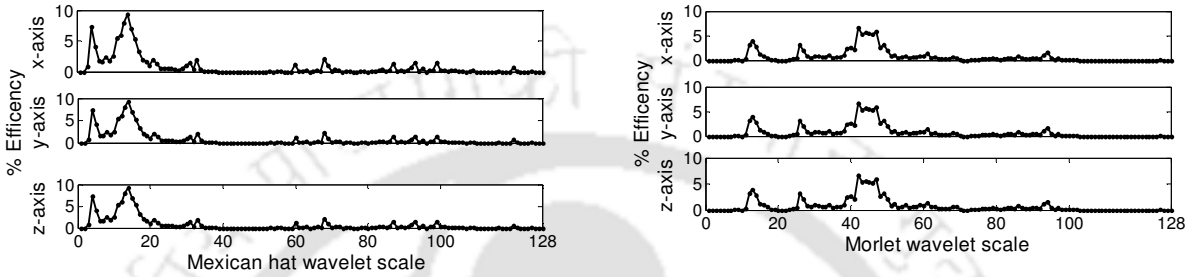
(a)



(b)

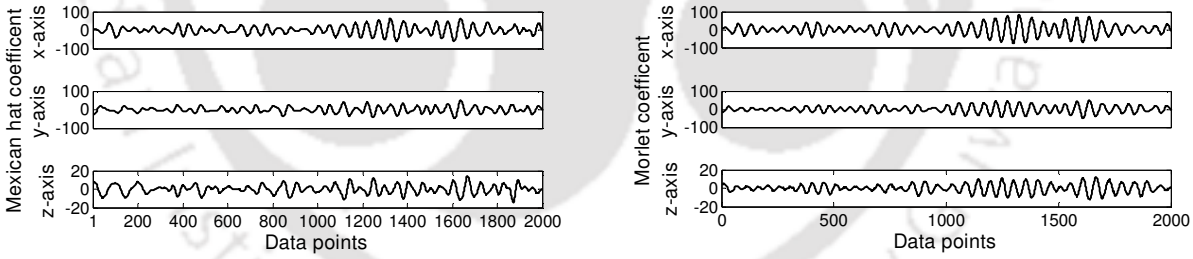
Figure 6.4 Variations of the chosen scale for 2000 data points (data sets in *x*, *y* and *z* directions) in three orthogonal directions for (a) Mexican hat wavelet scale and (b) Morlet wavelet scale for CT at 30 Hz rotating speed

The variation of CWT coefficient for the first set is illustrated in Figure 6.6 (a) for the Mexican hat and in Figure 6.6 (b) for the Morlet wavelet family. All CWT coefficients of that particular chosen scale are now selected for statistical features (i.e., the standard deviation, skewness and kurtosis).



(a) (b)

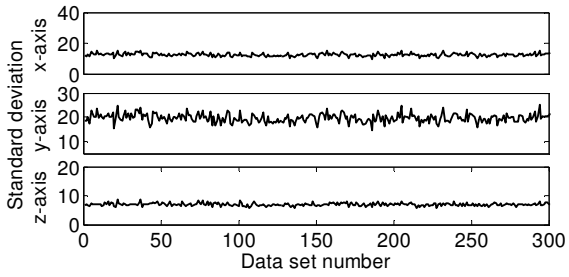
Figure 6.5 Variations of the efficiency for different scales from data sets in *x*, *y* and *z* directions for (a) Mexican hat wavelet scale and (b) Morlet wavelet scale for CT at 30 Hz rotating speed



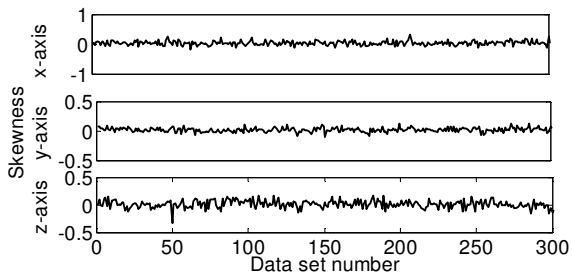
(a) (b)

Figure 6.6 Variation of the CWT coefficients for first 2,000 data points in three orthogonal directions for (a) Mexican hat wavelet scale and (b) Morlet wavelet scale for CT at 30 Hz rotating speed

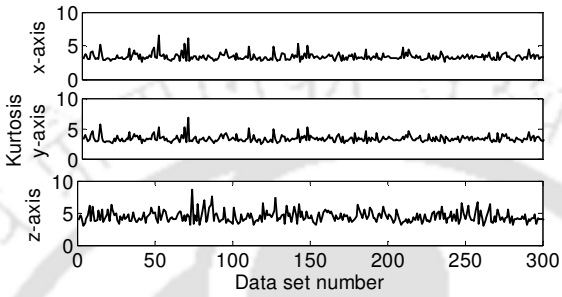
The CWT based features for the CT case at 30 Hz is shown in Figure 6.7 (a-c) for the Mexican hat family and in Figure 6.8 (a-c) for the Morlet wavelet family.



(a)

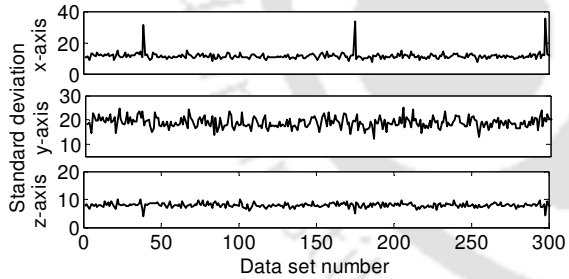


(b)

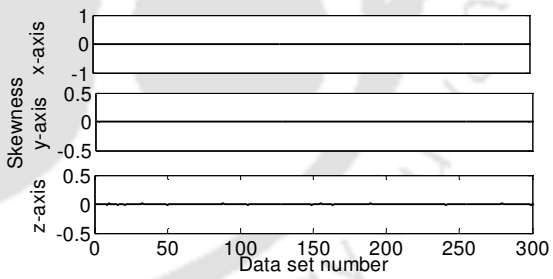


(c)

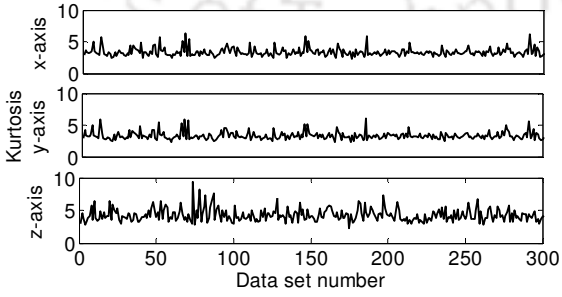
Figure 6.7 Features from the Mexican hat wavelet family of acquired signals for CT at 30 Hz rotating speed



(a)



(b)



(c)

Figure 6.8 Features of the Morlet wavelet family of acquired signals for the CT at 30 Hz rotational speed

6.2.2 WPT Based Feature Extraction

WPT is very effectual tools of the noise elimination in the signal to an optimum degree. Wavelet packets were introduced by Coifman *et al.* (1992) by generalizing the link between multi resolution approximations and wavelets. The *wavelet packet* method is a generalization of wavelet decomposition that offers a richer range of possibilities for the signal analysis. Figure 6.9 shows the wavelet packet tree decomposition for any general vibration signal.

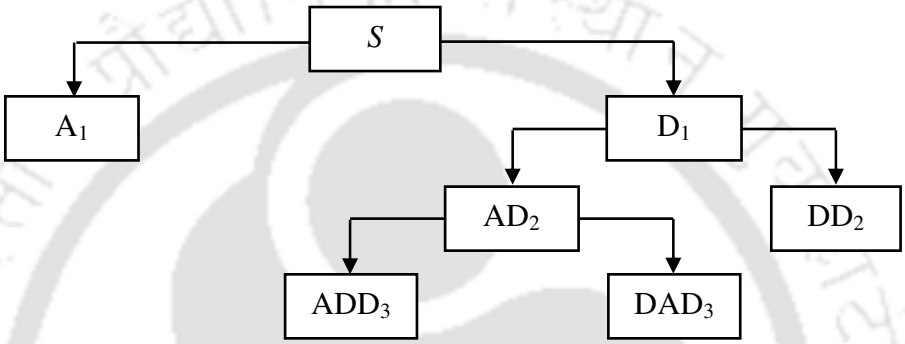


Figure 6.9 Wavelet packet tree decomposition for any general vibration signal

In the wavelet analysis, a signal is split into an *approximation* and a *detail*. The approximation is then itself divided into a second-level approximation and a detail signal, and the process is repeated. The wavelet decomposition tree is a part of this complete binary tree. For instance, the wavelet packet analysis allows the signal *S* to be represented as $A_1 + AAD_3 + DAD_3 + DD_2$.

Choosing one out of all these possible encodings presents an interesting problem. *Entropy* is a common concept in many fields, and also in the signal processing. Suppose that *s* is the signal and $(s_i)_i$ coefficients of *s* in an orthonormal basis. The entropy *E,* an additive cost function such that $E(0) = 0$, can be expressed as

$$E(s) = \sum_i E(s_i) \tag{6.4}$$

Here an *entropy-based criterion* is used to choose the most appropriate decomposition of a specified signal. This means, we look at each node of the decomposition tree and compute the information to be gained by executing each split. The straightforward and competent algorithms exist for both the wavelet packet decomposition and optimal decomposition selection. The MATLAB™ toolbox chosen for this work based on work by Coifman *et al.* (1992), with direct applications in the optimal signal coding and the data compression. Such algorithms allow the wavelet packet 1-D tools to include "Best Level" and "Best Tree" features that compute the optimal complete sub-tree of an initial tree with respect to an entropy type criterion.

In present case, the maximum height of the tree is selected to be one and the wavelet to be 'Db4' and then the wavelet packet decomposition is performed at the root of the tree high-pass filtering method. In this filtering method that is based on 'Db4' discrete wavelet transform selects the high-frequency low-scale components for the high-pass filtering, and the low-frequency high-scale components for the low-pass filtering. Next, the most appropriate node needs to be chosen from resulting nodes of the tree and this is done based on the maximum energy (or Shannon) criterion. The (non-normalized) Shannon entropy can be expressed from the general entropy expression as

$$E1(s_i) = s_i^2 \log(s_i^2) \tag{6.5}$$

So, the above can be written with the convention $0(\log 0) = 0$.

$$E1(s) = -\sum_i s_i^2 \log(s_i^2) \tag{6.6}$$

A node with the signal containing the maximum energy is selected and hence, the most optimally refined signal is obtained. The percentage of energy at each selected node is calculated (for each 2,000 data of 300 sets). The maximum energy is calculated at each 2,000 data of 300 sets. Figure 6.10 shows the maximum percentage energy in *x*-direction data set.

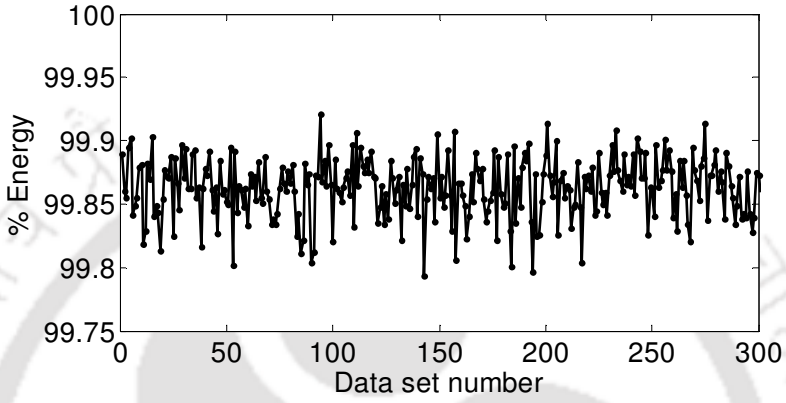


Figure 6.10 Variation of the maximum energy in *x*-direction data set for the WPT based feature for CT at 30 Hz rotating speed

Once the best energy level is obtained different statistical features (i.e., the standard deviation, skewness and kurtosis) could be calculated for the WPT coefficients. The WPT based features for the CT case at 30 Hz is shown in Figure 6.11.

Finally, 300 data points are collected in three directions for each three features (i.e., total 9 combinations), which means 9×300 data sets are available for each faults. For four different faults (i.e., CT, MT, WT and ND) there are total $9\times4\times300$ data sets available. These total data sets are divided and used for the training, the testing of optimizing the parameters and for the final testing in the simulation.

6.3 Simulation Results

The SVM software first needs to be trained with vibration data comprising of four types of gears faults (i.e., CT, MT, WT and ND). For the testing and the training following three conditions are considered.

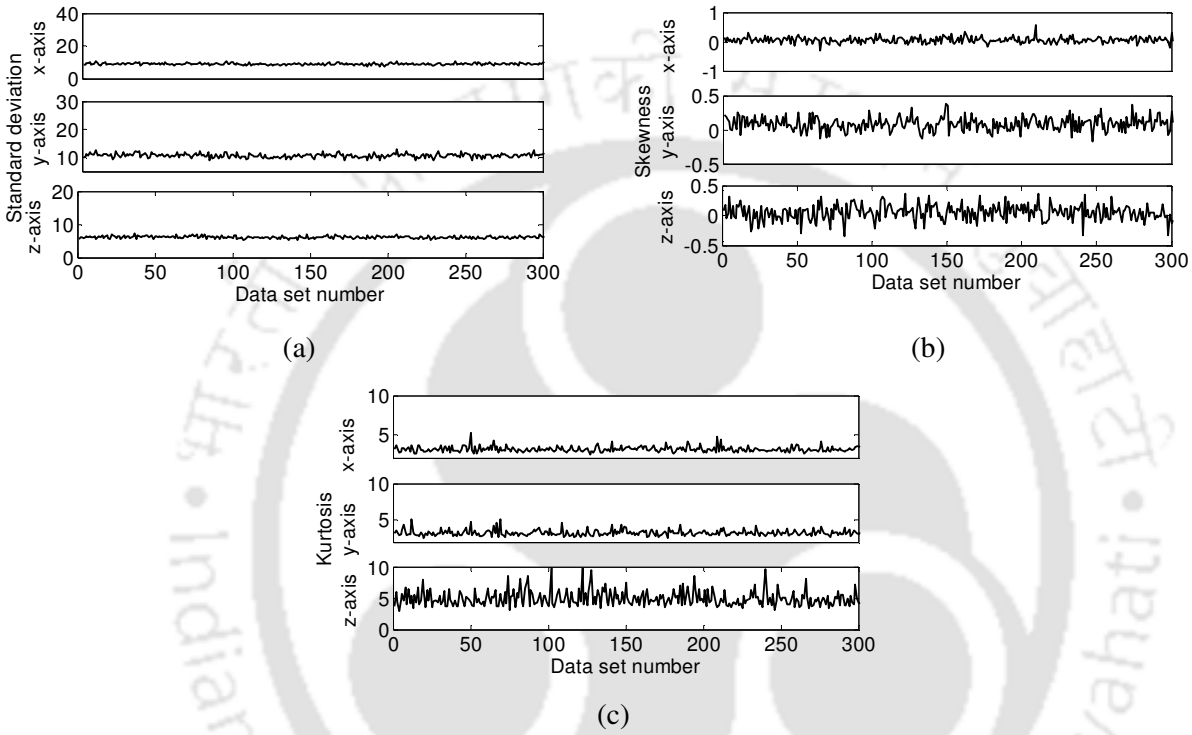


Figure 6.11 Features of the Wavelet packet family of acquired signals for the CT at 30 Hz rotational speed

6.3.1 Training and Testing at Same Rotational Speed

In this case the SVM was trained and tested at the same rotational speed. Rotational speeds selected were 10-30 Hz in a step of 5 Hz.

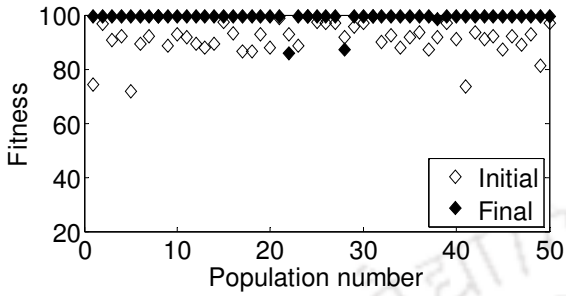
Optimization of Parameters: During the optimization of SVM parameters by the GA and the ABCA, $9\times4\times180$ data sets were used for the training of the fault classification, and $9\times4\times90$ data sets were used for the testing.

The variation of initial and final fitness values (i.e., the percentage accuracy) with the population for *C*-SVC at 30 Hz rotational speeds are shown in Figure 6.12(a), (c) and (e) with the GA and Figure 6.12(b), (d) and (f) with the ABCA for the CWT and WPT based features, respectively). Similarly, the variation of the initial and final fitness values with the population for the ν-SVC at 30 Hz rotational speed are shown in Figure 6.13(a), (c) and (e) using the GA and Figure 6.13(b), (d) and (f) using the ABCA with the CWT and WPT based features, respectively. It includes all the population, including solutions violating constraints. From the choice of initial solutions, it could be seen that it has quite good divergence of the possible solution domain. In the last generation also all solutions have been shown including those violating constraints (however, very few), and it could be observed that the most of solutions have converged.

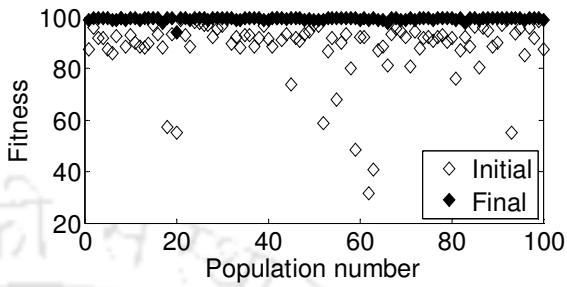
In the case of GSM $9\times4\times270$ data sets are used for the parameter estimation. Data points used for the optimization and the final testing is tabulated in Table 4.3. The cross validation accuracy in the case of GSM at 30 Hz rotational speed for the *C*-SVC is shown in Figure 6.14(a), (c) and (e); and for the ν-SVC is shown in Figure 6.14(b), (d) and (f) with the CWT and WPT based features, respectively. In these the contour line for the percentage accuracy is plotted and the best CV accuracy is marked. Correspondingly, the best SVC parameters are found from the best CV accuracy and the testing accuracy is tabulated. The optimized percentage accuracy (all classes) of different SVM formulation is shown in Table 6.2.

Fault Prediction Ability: After optimization of parameters $9\times4\times30$ data sets are used for the final testing of the fault classification for the GA and the ABCA, and $9\times4\times30$ data sets are used for the final testing of the fault classification for the GSM. In many times the accuracy (all classes) is more in the GA and the ABCA as compared with the GSM, which reflect the

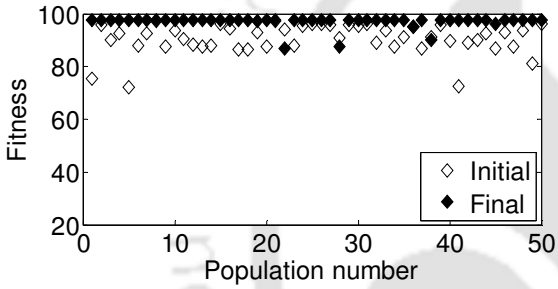
soundness of the GA and the ABCA. If we look upon the testing accuracy for all classes, the lowest one is equal to 88.33% and this occurs at 15 Hz rotational speeds on the *C*-SVC case.



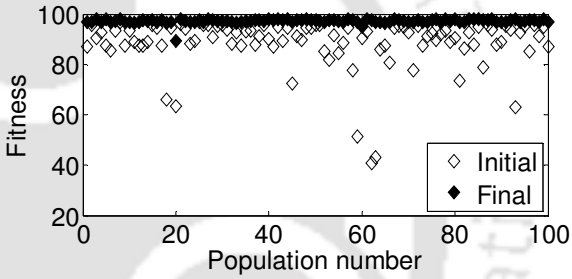
(a) 30 Hz (*C*-SVC with Morlet wavelet family using GA optimization)



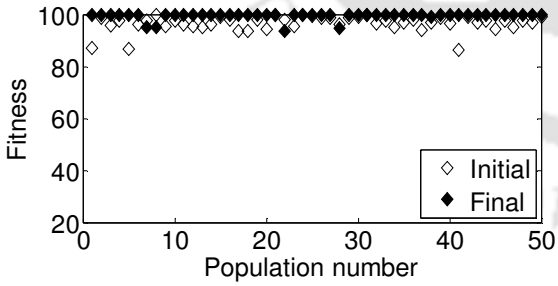
(b) 30 Hz (*C*-SVC with Morlet wavelet family using ABCA optimization)



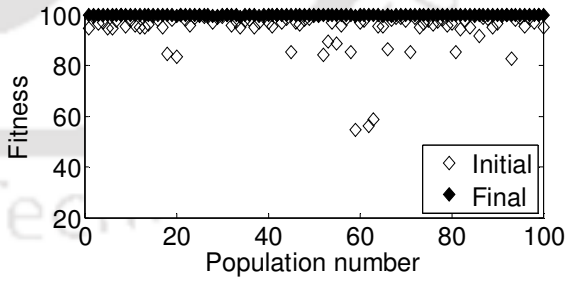
(c) 30 Hz (*C*-SVC with Mexican hat wavelet family using GA optimization)



(d) 30 Hz (*C*-SVC with Mexican hat wavelet family using ABCA optimization)

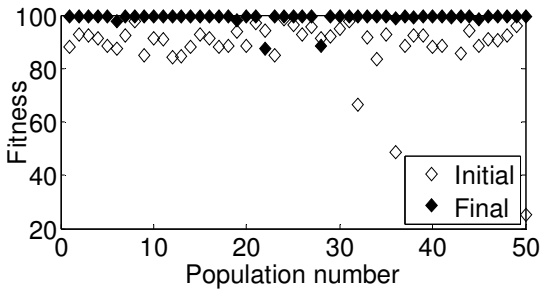


(e) 30 Hz (*C*-SVC with wavelet packet transform using GA optimization)

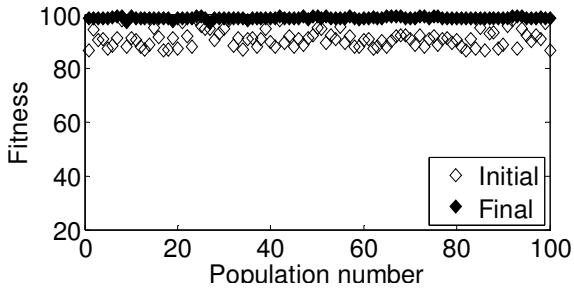


(f) 30 Hz (*C*-SVC with wavelet packet transform using ABCA optimization)

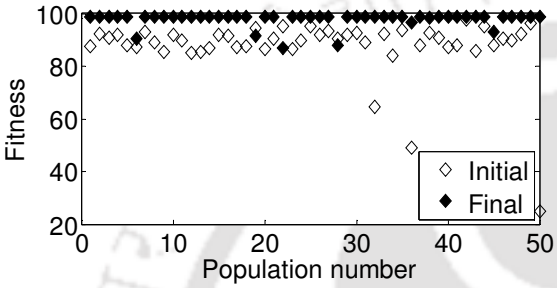
Figure 6.12 Variation of the initial and final fitness (percentage accuracy) with the population for the *C*-SVC (a), (c), (e) using the GA and (b), (d), (f) using the ABCA for 30 Hz training and testing speed



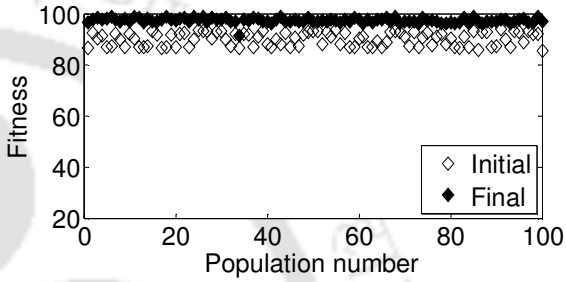
(a) 30 Hz (ν-SVC with Morlet wavelet family using GA optimization)



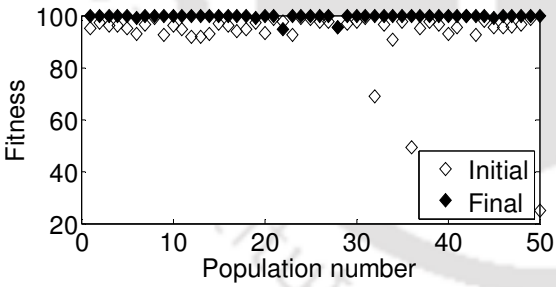
(b) 30 Hz (ν-SVC with Morlet wavelet family using ABCA optimization)



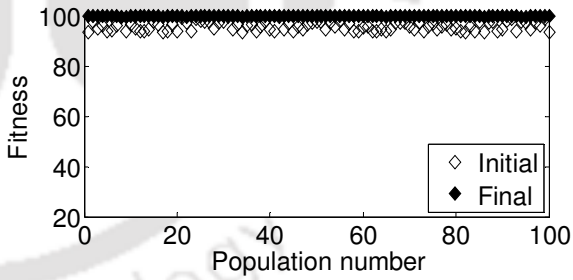
(c) 30 Hz (ν-SVC with Mexican hat wavelet family using GA optimization)



(d) 30 Hz (ν-SVC with Mexican hat wavelet family using ABCA optimization)

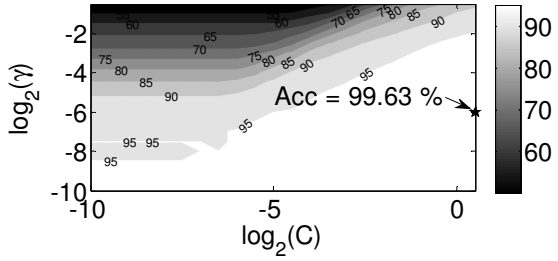


(f) 30 Hz (ν-SVC with wavelet packet transform using GA optimization)

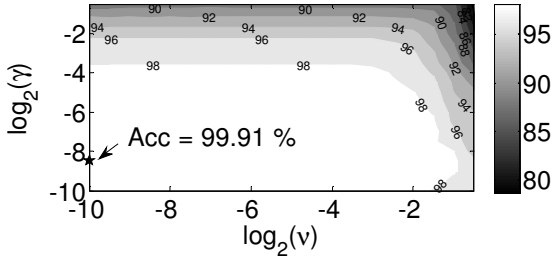


(g) 30 Hz (ν-SVC with wavelet packet transform using ABCA optimization)

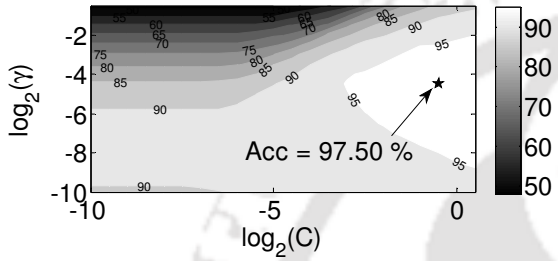
Figure 6.13 Variation of the initial and final fitness (percentage accuracy) with the population for the ν-SVC (a), (c), (e) using the GA and (b), (d), (f) using the ABCA for 30 Hz training and testing speed



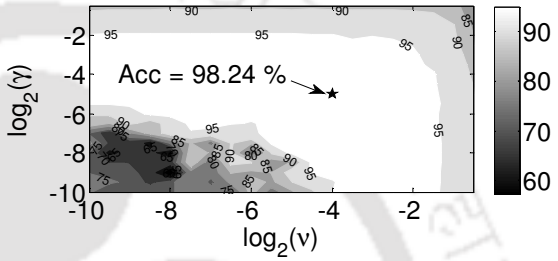
(e) 30 Hz (*C*-SVC with Morlet wavelet family)



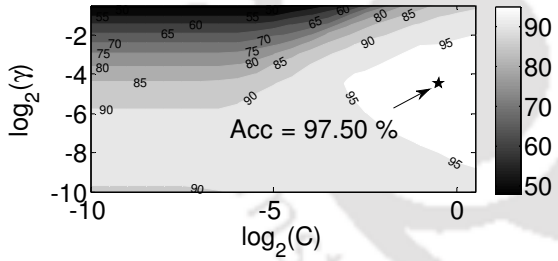
(f) 30 Hz (ν-SVC with Morlet wavelet family)



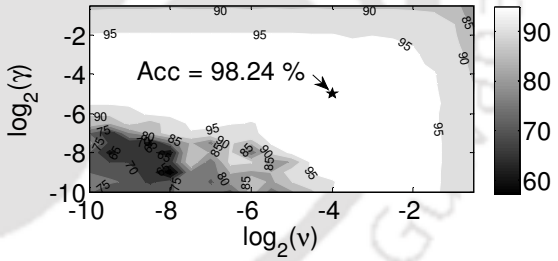
(g) 30 Hz (*C*-SVC with Mexican hat wavelet family)



(h) 30 Hz (ν-SVC with Mexican hat wavelet family)



(i) 30 Hz (*C*-SVC with WPT)



(j) 30 Hz (ν-SVC with WPT)

Figure 6.14 Cross validation accuracy for 30 Hz training and testing speed

It is also observed that other than 15 Hz rotational speed at all speeds the fault prediction is near perfect (i.e., at 10, 20, 25 and 30 Hz). Table 6.4 illustrates the percentage fault prediction (individual classes) in various rotational speeds against the best prediction accuracy (all classes). The prediction accuracy of 96.67% is the individual lowest for the WT case at 15 Hz rotational speed.

Table 6.2 Classification of gear faults (all classes) with the training and the testing from the continuous wavelet transforms (Morlet and Mexican hat family) and wavelet packet transforms features for same speed training and testing

Training speed (Hz)	Testing speed (Hz)	SVM formulation	Testing accuracy (%) for								
			Morlet wavelet family			Mexican hat wavelet family			Wavelet packet transform		
			CV*	GA	ABCA	CV*	GA	ABCA	CV*	GA	ABCA
For same speed											
10	10	*C*-SVC	95.00	95.83	96.67	90.83	91.67	92.50	90.83	97.50	98.33
		ν -SVC	95.00	96.67	95.83	94.17	93.33	93.33	94.17	95.83	**100.00**
15	15	*C*-SVC	**99.17**	**99.17**	**99.17**	94.17	93.33	93.33	94.17	88.33	90.00
		ν -SVC	92.50	**99.17**	**99.17**	95.83	96.67	95.83	95.83	89.17	90.83
20	20	*C*-SVC	**100.00**	**100.00**	99.17	95.83	95.83	95.00	95.83	97.50	98.33
		ν -SVC	**100.00**	**100.00**	**100.00**	97.50	97.50	95.83	97.50	98.33	99.17
25	25	*C*-SVC	99.17	99.17	99.17	95.00	95.00	95.00	95.00	**100.00**	**100.00**
		ν -SVC	**100.00**	**100.00**	**100.00**	95.83	95.00	95.83	95.83	**100.00**	**100.00**
30	30	*C*-SVC	99.17	99.17	98.33	91.67	90.83	90.83	91.67	**100.00**	**100.00**
		ν -SVC	99.17	98.33	98.33	95.00	95.83	95.83	95.00	**100.00**	**100.00**

*CV means the *cross validation* accuracy of the GSM. Bold values represent prediction accuracies of the maximum accuracy.

6.3.2 Training at two Different Rotational Speeds and Testing at an Intermediate Rotational Speed

In this case the SVM is trained at two rotational speeds and tested at an intermediate rotational speed for which the SVM was not trained. Values of the rotational speed range and the intermediate testing rotational speed are summarized in Table 4.2. In this case beyond 10 Hz (i.e., 600 rpm) the rotational speed range is not considered, because the range is wide and hence the percentage accuracy becomes low.

***Optimization of Parameters**:* Here $9\times8\times90$ data sets are used for the training and $9\times8\times45$ data sets are used for the testing of optimized parameters for the GA and the ABCA. For example, in the case of optimization of parameters for the first range, $9\times4\times90$ data sets are taken from 10 Hz rotational speed and $9\times4\times90$ data sets are taken from 15 Hz rotational

speed so that all together $9\times8\times90$ data sets are considered for the training. Whereas, $9\times4\times45$ data sets are taken from 10 Hz rotational speed and $9\times4\times45$ data sets are taken from 15 Hz rotational speed so that all together $9\times8\times45$ data sets are considered for the testing. Data sets used for optimizing and final testing is tabulated in Table 4.3. The optimized percentage accuracy (all classes) for the different SVM formulation is shown in Table 6.3.

Fault Prediction Ability: After the optimization, $9\times4\times30$ data sets from intermediate rotational speeds are used for the final testing or classification. Similarly $9\times8\times125$ data sets are used for the parameter estimation and $9\times4\times30$ data sets are used for the final testing for the GSM. For the intermediate rotational speed range of 5 Hz, at the lowest rotational speed the classification accuracy (all classes) is nearly 68.33% and it increases at the high rotational speed to 100%. In the case of intermediate rotational speed range of 10 Hz, for the lowest rotational speed the classification accuracy (all classes) is nearly 21.67% and it increases at the high rotational speed to 98.33%. It is observed that the GA and ABCA techniques have shown the same prediction accuracy (all classes) with conventional GSM, but in a particular speed range 10 Hz to 15 Hz (intermediate 5 Hz range), the GSM gave good results (0.86% more accuracy level then the GA and ABCA techniques). This may be due to the former method is less sensitive to noise as compared to latter ones. The percentage fault prediction (individual classes) at various rotational speeds based on the best optimized condition is illustrated in Table 6.4. In the case of rotational speed range of 5 Hz, the classification for different faults (individual classes) are more than 96.67% but in case of the rotational speed range of 10 Hz prediction accuracy at 15 Hz rotational speed the individual classification for fault conditions shows a very low prediction accuracy and that is down to 16.67%.

6.3.3 Training at two Different Rotational Speeds and Testing at an Extrapolated Rotational Speed

In this case the SVM is trained at two rotational speeds and tested at an extrapolated rotational speed. Training rotational speeds and the extrapolated testing rotational speed are tabulated in Table 4.2.

Optimization of Parameters: For selecting data points the same procedure is considered as used for the interpolation case, but only difference is that for the testing the extrapolation speed is considered. The optimized percentage accuracy (all classes) of the different SVM formulation is shown in Table 6.3.

Fault Prediction Ability: After optimization the same strategy for data points selection are considered for the testing as described in the interpolation case, whereas here the extrapolated speed is considered. At the extrapolated speed range of 5 Hz the prediction accuracy (all classes) is 65.83% to 97.50%, and in case of 10 Hz range it is 35.83% to 83.33%. At 10 Hz range case, the GA and the ABCA show good fault prediction results. For 5 Hz range, it gives the best classification accuracy, and it increases gradually from 76.67% to 97.50% from the lowest rotational speed to the highest rotational speed. In 10 Hz range also the best accuracy increases (all classes) from 66.67% to 83.33% from the lower rotational speed to the higher rotational speed. The best percentage fault predictions (individual classes) at various rotational speeds are illustrated in Table 6.4. In the case of speed range of 5 Hz the lowest prediction accuracy (individual classes) is at 15 Hz for the fault detection of CT (i.e., 43.33%). Similarly, in the case of speed range of 10 Hz, the lowest performance is for the ND condition at 25 Hz (i.e., 13.33%). For more clarity the overall performances are tabulated in Table 6.5.

Table 6.3 Classification of gear faults (all classes) with the training and the testing from the continuous wavelet transform (Morlet and Mexican hat family) and wavelet packet transform features for the interpolation and extrapolation speeds

Training speed (Hz)	Testing speed (Hz)	SVM formulation	Testing accuracy (%) for								
			Morlet wavelet family			Mexican hat wavelet family			Wavelet packet transform		
			CV*	GA	ABCA	CV*	GA	ABCA	CV*	GA	ABCA
For interpolation speed (range 5 Hz)											
10, 15	12.5	*C*-SVC	95.83	95.00	95.00	68.33	75.00	71.67	68.33	88.33	86.67
		ν -SVC	**97.50**	96.67	96.67	71.67	77.50	70.83	71.67	86.67	89.17
15, 20	17.5	*C*-SVC	99.17	99.17	99.17	91.67	93.33	91.67	91.67	94.17	95.00
		ν -SVC	**100.00**	99.17	98.33	92.50	94.17	92.50	92.50	94.17	95.00
20, 25	22.5	*C*-SVC	99.17	99.17	99.17	96.67	95.00	95.00	96.67	95.83	95.83
		ν -SVC	99.17	99.17	**100.00**	95.00	95.83	96.67	95.00	95.83	95.83
25, 30	27.5	*C*-SVC	99.17	99.17	99.17	91.67	89.17	92.50	91.67	99.17	99.17
		ν -SVC	**100.00**	**100.00**	99.17	90.83	93.33	91.67	90.83	**100.00**	**100.00**
For interpolation speed (range 10 Hz)											
10, 20	15	*C*-SVC	50.83	50.00	50.00	21.67	25.00	21.67	21.67	23.33	25.00
		ν -SVC	57.50	53.33	**65.83**	24.17	34.17	48.33	24.17	32.50	39.17
15, 25	20	*C*-SVC	**92.50**	**92.50**	**92.50**	58.33	39.17	56.67	58.33	75.83	74.17
		ν -SVC	89.17	79.17	87.50	70.00	45.83	69.17	70.00	63.33	88.33
20, 30	25	*C*-SVC	**98.33**	97.50	**98.33**	82.50	82.50	81.67	82.50	85.83	90.00
		ν -SVC	**98.33**	**98.33**	**98.33**	82.50	81.67	82.50	82.50	90.83	85.83
For extrapolation speed (range 5 Hz)											
10, 12.5	15	*C*-SVC	70.83	69.17	68.33	65.83	70.00	67.50	65.83	74.17	75.00
		ν -SVC	**76.67**	74.17	74.17	70.83	**76.67**	73.33	70.83	69.17	70.00
20, 22.5	25	*C*-SVC	90.00	**91.67**	**91.67**	79.17	79.17	77.50	79.17	82.50	82.50
		ν -SVC	88.33	**91.67**	88.33	85.00	75.83	81.67	85.00	79.17	90.83
25, 27.5	30	*C*-SVC	**97.50**	**97.50**	**97.50**	85.83	90.00	87.50	85.83	96.67	95.00
		ν -SVC	96.67	96.67	**97.50**	90.00	90.83	90.00	90.00	**97.50**	95.00
For extrapolation speed (range 10 Hz)											
10, 15	20	*C*-SVC	49.17	46.67	48.33	35.83	39.17	35.83	35.83	40.00	42.50
		ν -SVC	64.17	55.83	55.83	47.50	49.17	44.17	47.50	58.33	**66.67**
15, 20	25	*C*-SVC	60.83	**68.33**	63.33	52.50	57.50	52.50	52.50	46.67	55.00
		ν -SVC	54.17	66.67	65.00	54.17	53.33	53.33	54.17	44.17	66.67
20, 25	30	*C*-SVC	76.67	77.50	77.50	61.67	70.00	64.17	61.67	70.83	70.83
		ν -SVC	80.83	78.33	**83.33**	73.33	75.00	65.00	73.33	70.83	70.83

*CV means the *cross validation* accuracy of the GSM. Bold values represent prediction accuracies of the maximum accuracy.

The lowest prediction accuracy (all classes) and highest prediction accuracy (all classes) are the lowest and highest testing accuracy (all classes), respectively, as per Table 6.2 and Table 6.3. The lowest and highest individual fault predictions are found from Table 6.4.

Table 6.4 Best percentage prediction accuracies (individual classes) at various speeds (time frequency domain features)

Speed (Hz)	CT (%)	MT (%)	WT (%)	ND (%)
For same speed				
10	100.00	100.00	100.00	100.00
15	100.00	100.00	96.67	100.00
20	100.00	100.00	100.00	100.00
25	100.00	100.00	100.00	100.00
30	100.00	100.00	100.00	100.00
For interpolation speed (range 5 Hz)				
12.5	100.00	96.67	96.67	96.67
17.5	100.00	100.00	100.00	100.00
22.5	100.00	100.00	100.00	100.00
27.5	100.00	100.00	100.00	100.00
For interpolation speed (range 10 Hz)				
15	16.67	83.33	80.00	83.33
20	100.00	80.00	96.67	93.33
25	100.00	96.67	96.67	100.00
For extrapolation speed (range 5 Hz)				
15	43.33	76.67	90.00	96.67
25	73.33	100.00	96.67	96.67
30	100.00	100.00	100.00	90.00
For extrapolation speed (range 10 Hz)				
20	53.33	93.33	96.67	23.33
25	23.33	96.67	90.00	63.33
30	46.67	100.00	86.67	100.00

Initially, for each of the four classification cases (i.e., CT, MT, WT and ND), the training data was provided at running speeds from 10 Hz to 30 Hz at the interval of 2.5 Hz, and then the multiclass classification capability of two classes of SVM was noted for these running speeds. It is concluded that the SVM has the ability to make perfect classifications if the training data is available for that particular running speed. The capability of the SVM for the classification at the interpolated and extrapolated speeds beyond its training data speeds are performed and it is noted that the SVM can still be able to make accurate classification utilizing the training data from the beginning and the end of a range of running speeds whereas the test data could be at any of the speed from within that range (interpolation) and out of range (extrapolation).

Table 6.5 Summary of overall fault prediction performances from the continuous wavelet transform (Morlet and Mexican hat family) and wavelet packet transform features

	Same rotational speed	Interpolation speed		Extrapolation speed	
		Range 5 Hz	Range 10 Hz	Range 5 Hz	Range 10 Hz
Lowest prediction accuracy (all classes)	88.33%	68.33%	21.67%	65.83%	35.83%
Highest prediction accuracy (all classes)	100.00%	100.00%	98.33%	91.67%	83.33%
Lowest individual faults prediction	96.67% (15 Hz, WT case)	96.67% (12.5 Hz, MT, WT and ND case)	16.67% (15 Hz, CT case)	43.33% (15 Hz, CT case)	23.33% (25 Hz, ND case)
Highest individual faults prediction	100.00%	100.00%	100.00%	100.00%	100.00%

In the interpolation and extrapolation fault predictions, it has been observed that in many times the GA and the ABCA show the similar fault prediction results as the GSM. On the other hand, the GSM shows its marginal superiority (greater than 0.84% to 0.86%) in two cases. It is also

observed that the prediction accuracy gradually increases with the increase of the rotational speed. This is due to the high signal-to-noise level at the high rotation speed due to better manifestation of faults in vibration signals at these speeds. It is observed that at the same speed and the interpolation speed range the Morlet wavelet family has shown good prediction results. Again for the extrapolation speed range, in three cases the Morlet wavelet family and one case the WPT performed well.

6.4 Summary

The capability of SVM in multi-fault classifications of gears by using the wavelet based vibration data has been demonstrated in this chapter. In the case of training and testing at the same rotational speed a near perfect prediction accuracies are found. On the other hand, the training at the beginning and at the end of the rotational speed of a range of rotational speed, and testing at the intermediate and extrapolated rotational speeds, the prediction accuracy is still encouraging. The potential of two SVM classifiers (C-SVC and ν-SVC) with respect of its tuning of parameters by using the two optimizing tool the GA and the ABCA, and their performance are studied and compared with the GSM classical technique. The GA and ABCA based optimization methods show its effectiveness at several times. The Morlet wavelet family showed a better performance over the other wavelet in 3 out of 5 cases at the same rotational speed, 7 cases out of 7 cases at interpolation speed and 5 cases out of 6 cases at extrapolation speed.

CHAPTER 7

Multi-fault Classifications: A Comparative Analysis

7.1 Introduction

In Chapters 4 to 6, the classification of gear faults were performed in three different domains (time, frequency and time-frequency domain). Analyses of fault predictions from three optimization techniques were summarised for individual domains. The best prediction accuracies at the same rotational speed as well as at the intermediate and extrapolated speeds for a particular domain were analysed. In this chapter, prediction accuracies have been compared among three domains and conclusions are drawn. It further processes the three domain fault prediction results for getting unified prediction accuracies of gear faults. A fault for a particular data point is assigned according to the majority in voting of the fault defined by different methods and by using different domain data. On this basis new unified prediction accuracies are obtained and is expected to be better than averaged ones.

7.2 Comparison of Fault Predictions in Different Domains

The SVM based prediction of faults was discussed for the time, frequency and time-frequency domains in Chapters 4 to 6, respectively, in conjunction with different optimization techniques. In these chapters best fault predictions were indicated in classification tables irrespective to the classifier. In time-frequency domain also among three wavelets, the one with the best fault prediction was indicated. Hence, now the comparison has been performed in the present section for the best fault predictions of three domains. These prediction accuracies (all classes) have been compared and illustrated in Figure 7.1 to Figure 7.3, respectively, for three domains. Figure 7.1 illustrates the comparison of the wavelet based prediction accuracies with that of the time and frequency domains at the same rotational speed. In this bar chart wavelet prediction accuracies show better than other domains except at the high speed, where it is comparable. For example, at 15 Hz rotational speed the wavelet prediction accuracy (all classes) is 99.17%

compared to prediction accuracies in the time and frequency domains of 90.83% and 97.50%, respectively.

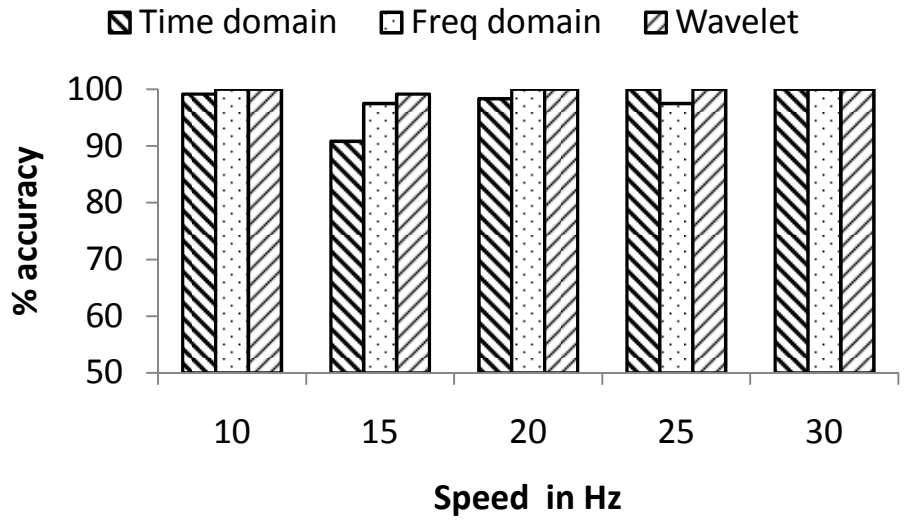
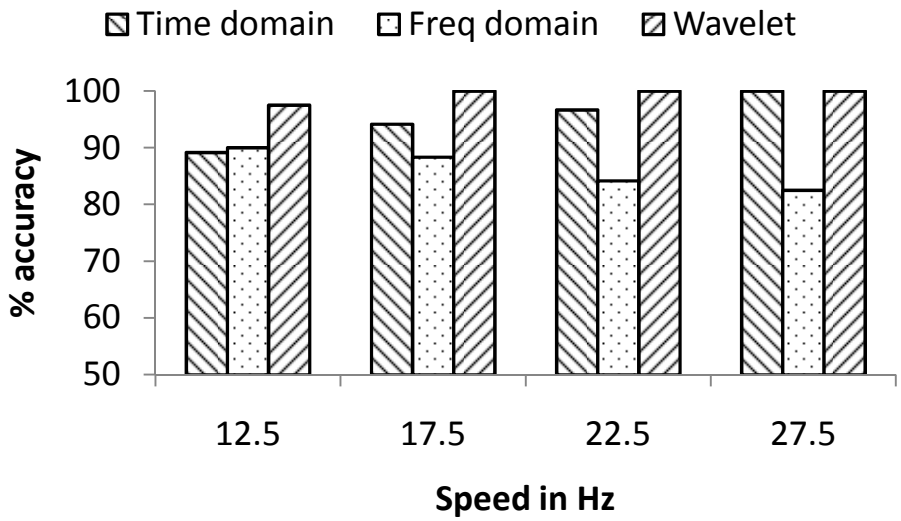


Figure 7.1 Comparison of the best testing accuracy (all classes) with the time, frequency and time-frequency domain data at the same rotational speed

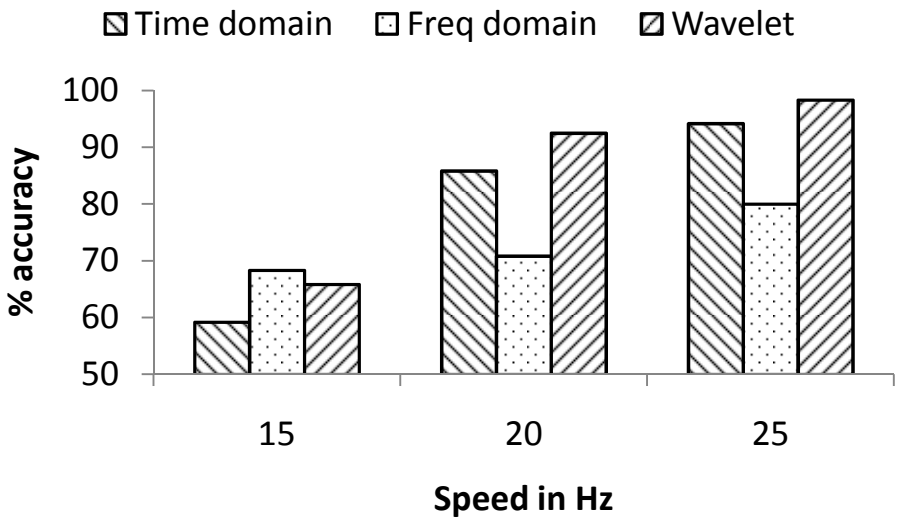
Similarly, Figure 7.2 and Figure 7.3 illustrate the fault prediction (all classes) comparison for the interpolation and extrapolation rotational speeds, respectively. It has been observed that in all cases the wavelet based accuracy revealed the supremacy except few exceptions. This happened for testing with the interpolation speed range of 10 Hz at 15 Hz of speed, for the extrapolation speed range of 5 Hz at 30 Hz of speed, and the extrapolation speed range of 10 Hz at 25 Hz of speed. Out of these three cases, in the first case frequency domain fault prediction accuracy is 3.79% more than the wavelet one and in last two cases time domain prediction accuracies are 2.56% and 1.23%, respectively, more than the wavelet ones; which are marginally higher.

From above comparisons, it is established that the wavelet data based fault predictions show better average results compared to other domain data. Among three wavelets considered, the Morlet wavelet family shows the best fault prediction. From the above analysis it is very

difficult to say that one particular domain results shows the best fault prediction. It will be a real challenge to choose a best fault prediction among the predicted results from various domains considered. To find out a more reliable fault prediction, in the following section a novel voting methodology has been proposed and applied.

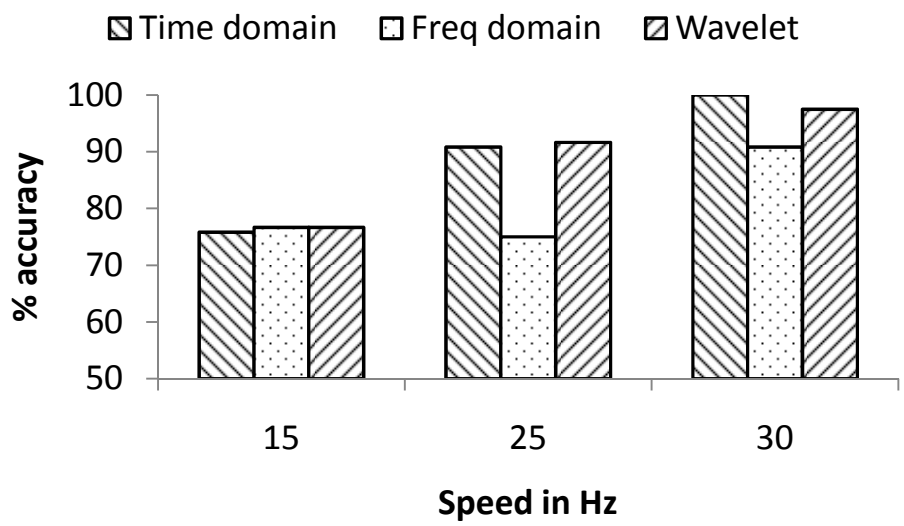


(a) Accuracy at the interpolation range of 5 Hz

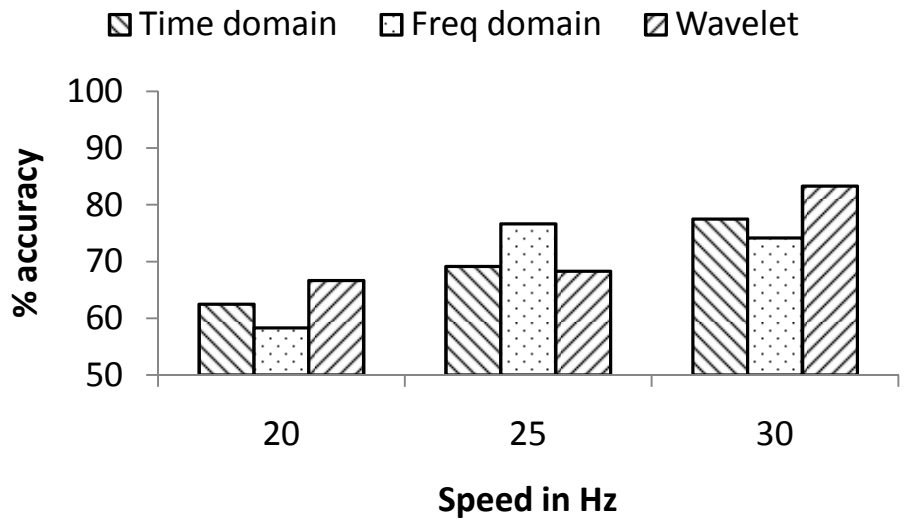


(b) Accuracy at the interpolation speed range of 10 Hz

Figure 7.2 Comparison of the best testing accuracy (all classes) with the time, frequency and time-frequency domain data at the interpolation speed



(a) Accuracy at the extrapolation speed range of 5 Hz



(b) Accuracy at the extrapolation speed range of 10 Hz

Figure 7.3 Comparison of the best testing accuracy (all classes) with the time, frequency and time-frequency domain data at the extrapolation speed

It is observed that the prediction (all classes) got from frequency domain parameters, in general, lower than those from other two domains. The symptom of gear faults (CT, MT, WT and ND) in frequency domain can be presented follows.

- The gear mesh frequency appears regardless of gear conditions.
- 1x component of the defective gear is the most common vibration.
- A chipped or missing tooth generates high amplitude at 1x. It also excites the gear natural frequency side bands at its running speed.
- The dominate peak corresponding to structural natural frequency is the indication of the excessive wear.

The indication spike of the frequency in the different rotational speed (10 Hz, 15 Hz and 20 Hz) is indicated in the Appendix B. It is very difficult to distinguish them with each other without using the proper filter. Here filter is not used for finding features. Because of that the frequency domain results may show smaller value compared to other.

7.3 Estimation of Unified Fault Prediction Accuracies Based on Voting Strategies

To summarise, there were $9\times4\times300$ statistical feature points available in three orthogonal directions for four faults conditions in the time, frequency and time-frequency domains. The choice of optimum parameters of different SVM schemes and classifications of multiple faults has already been done in previous Chapters 4-6 for different domain features. The best prediction accuracies (all classes) are found by choosing the top prediction accuracies among the three optimization techniques for the C-SVC or ν-SVC classifier. The data are divided into two components for the optimization and for the final testing according to Table 4.3. For the validation or final testing $9\times4\times30$ data sets are available. In the classification, a particular data point of a data set is assigned a class or fault (CT, MT, WT or ND) with reference to a particular domain. That means a particular data point of a data set may be assigned different or same fault (for which fault the vibration data are actually collected) by classifiers (C-SVC or ν-SVC). For example, Figure 7.4 illustrates the flow chart for the classification of a time domain

data set at the same speed conditions. In which three statistical features (standard deviation, kurtosis, skewness) are calculated from $9\times4\times30$ data sets comprising of three orthogonal directions (x, y and z) vibration data and four fault conditions. This is represented as S_d-*1* to S_d-*30*, K_u-*1* to K_u-*30* and S_k-*1* to S_k-*30* for the standard deviation, kurtosis and skewness, respectively. Altogether 120×9 feature points are available. This features points are classified by the SVM classifier and assigned a fault class, which is indicated as A. The assigned fault class is represented by a matrix form as $\left[A_{Ti,j}\right]$ with 120×9 number of elements using time domain data.

Now the post-processing of prediction accuracies will be performed to get unified prediction accuracies (all classes), which is explained in the following paragraph and is illustrated in Figure 7.5. The assigned fault class by using the time, frequency and time-frequency (wavelet packet, Mexican hat and Morlet) domain data are defined as $\left[A_{Ti,j}\right],\left[A_{Fi,j}\right],\left[A_{Wpi,j}\right],\left[A_{Wmi,j}\right]$ and $\left[A_{Wmoi,j}\right]$, respectively, with 120×9 number of elements each, which can be found by the procedure described above. These are taken together for finding more reliable fault classification by the voting strategy as compared to the averaging. In which a particular data set of different domains is assigned a particular fault class according to the majority in voting of that fault by different methods. For example, a particular class is assigned at one-by-one position $\left(A_{O1,1}\right)$ of the new fault class matrix, $\left[A_{Oi,j}\right]$, according to the majority of assigned class at one-by-one position ($A_{T1,1}$, $A_{F1,1}$, $A_{Wp1,1}$, $A_{Wm1,1}$ and $A_{Wmo1,1}$) of each matrix, $\left[A_{Ti,j}\right],\left[A_{Fi,j}\right],\left[A_{Wpi,j}\right],\left[A_{Wmi,j}\right]$ and $\left[A_{Wmoi,j}\right]$, respectively. For more clarity, suppose that the fault class assigned by the classifier at $A_{T1,1}$ is CT, at $A_{F1,1}$ is CT, at $A_{Wp1,1}$ is MT, at $A_{Wp1,1}$ is WT, at $A_{Wm1,1}$ is CT and at $A_{Wmo1,1}$ is ND then the majority of the fault class for the one-by-one

position of the different classifiers is CT and hence fault class of $A_{O1,1}$ is written as CT. In this way all the positions of the fault class matrix $\left[A_{Oi,j}\right]$ is filled. But the actual fault class matrix is the matrix which represents the respective fault generated from statistical features. For example, the first row matrices of the fault class matrix $\left[A_{Ai,j}\right]$ belong to the CT fault, the second row of the fault class matrix $\left[A_{Ai,j}\right]$ is the MT fault, the third row of the fault class matrix $\left[A_{Ai,j}\right]$ is the WT fault, and the fourth row of the fault class matrix $\left[A_{Ai,j}\right]$ is the ND fault. In general, the actual fault class matrix can be written by showing the denoted fault class as

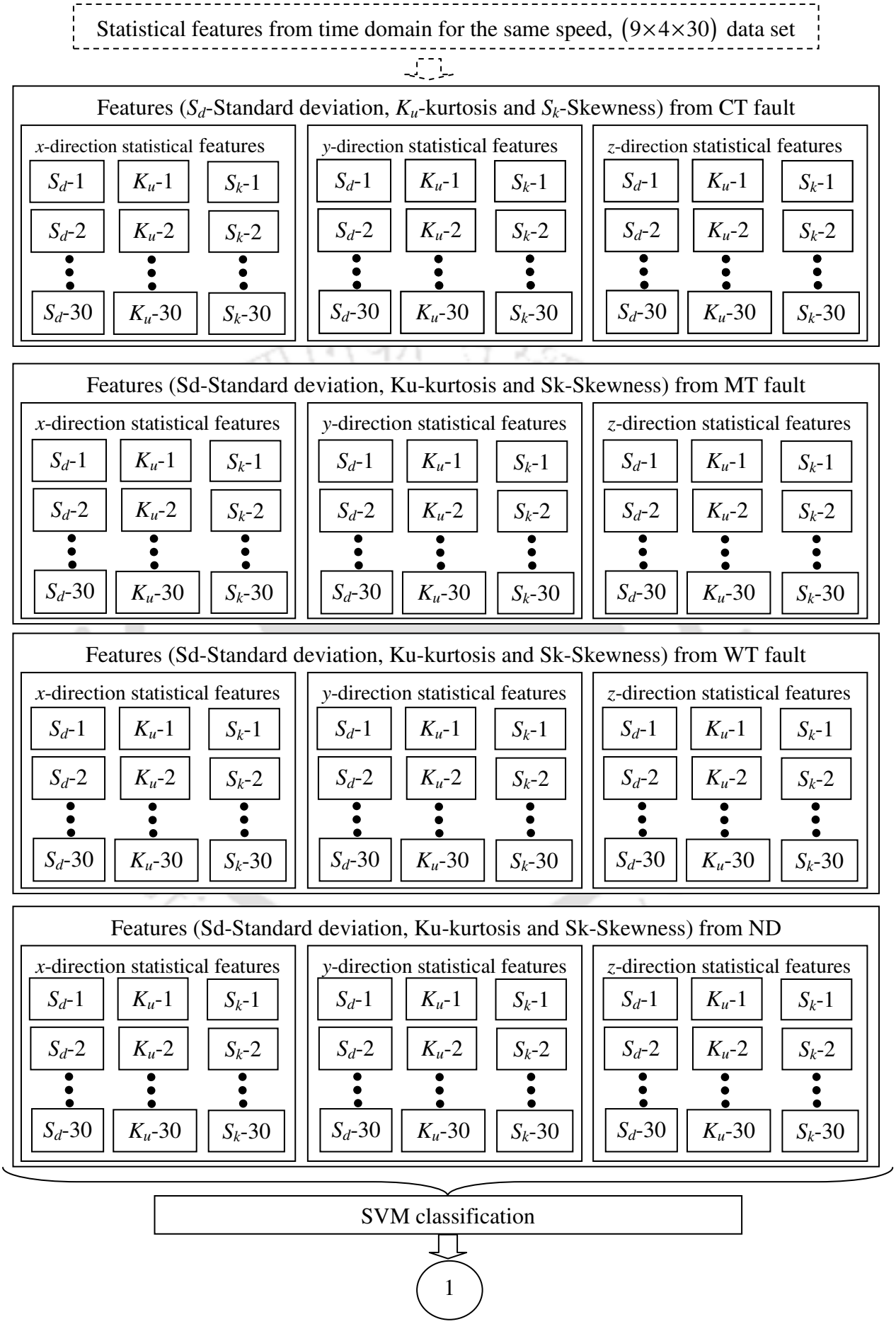
$$\left[A_{Ai,j}\right]=\begin{bmatrix}\begin{bmatrix}CT & \dots & CT\\ \vdots & \dots & \vdots\\ CT & \dots & CT\end{bmatrix} & \cdots & \begin{bmatrix}CT & \dots & CT\\ \vdots & \dots & \vdots\\ CT & \dots & CT\end{bmatrix}\\ \begin{bmatrix}MT & \dots & MT\\ \vdots & \dots & \vdots\\ MT & \dots & MT\end{bmatrix} & \cdots & \begin{bmatrix}MT & \dots & MT\\ \vdots & \dots & \vdots\\ MT & \dots & MT\end{bmatrix}\\ \begin{bmatrix}WT & \dots & WT\\ \vdots & \dots & \vdots\\ WT & \dots & WT\end{bmatrix} & \cdots & \begin{bmatrix}WT & \dots & WT\\ \vdots & \dots & \vdots\\ WT & \dots & WT\end{bmatrix}\\ \begin{bmatrix}ND & \dots & ND\\ \vdots & \dots & \vdots\\ ND & \dots & ND\end{bmatrix} & \cdots & \begin{bmatrix}ND & \dots & ND\\ \vdots & \dots & \vdots\\ ND & \dots & ND\end{bmatrix}\end{bmatrix} \tag{7.1}$$

Now the classification of the new matrix, $\left[A_{Oi,j}\right]$, are compared with actual fault class matrix, $\left[A_{Ai,j}\right]$, and new unified prediction accuracies (all classes) are found according to Eqn. (3.20). Here the algorithm is explained for one classifier and for a single speed case only. Similarly, this can be performed for the interpolation/extrapolation speed and for two classifiers also. Table 7.1 illustrates unified prediction accuracies for all classes (obtained from the above algorithm) for the same speed as well as for the interpolated and extrapolated rotational speeds, which are based on the *C*-SVC classifier. In the first row of the table describes the best

prediction accuracies (all classes) found in Chapter 4-6 by using the time, frequency and time-frequency domain data. This prediction accuracies (all classes) are found by using Eqn. (3.20) of the assigned fault class matrices $\left[A_{Ti,j}\right],\left[A_{Fi,j}\right],\left[A_{Wpi,j}\right],\left[A_{Wmi,j}\right]$ and $\left[A_{Wmoi,j}\right]$. On following the voting strategy the unified fault class matrix $\left[A_{Oi,j}\right]$ is found and the optimal accuracy (all classes) is calculated using Eqn. (3.20), which is 99.17%.

In the similar lines, the algorithm is applied in the first and second stages by choosing the two version of SVM classifier (*C*-SVC and ν-SVC) individually for finding the optimal classification. In the third stage, classifications of two SVM versions are considered together and unified prediction accuracies (all classes) are found.

***Unified Fault Prediction Accuracies using C-SVC Classifier*:** Table 7.1 illustrates unified prediction accuracies for all classes (obtained from the procedure described in preceding section) for the same speed as well as for the interpolated and extrapolated rotational speeds, which are based on the *C*-SVC classifier. The best prediction accuracies (all classes) for a rotational speed are tabulated for a particular domain in row wise, which are the best fault prediction found in Chapter 4-6 by three optimization techniques (GSM, GA, ABCA). The first row of Table 7.1 describes the best prediction accuracies (all classes) found in Chapter 4-6 by using the time, frequency and time-frequency domain data. This prediction accuracies are found by using Eqn. (3.20) of assigned fault class matrices $\left[A_{T_{i,j}}\right],\left[A_{F_{i,j}}\right],\left[A_{Wp_{i,j}}\right],\left[A_{Wm_{i,j}}\right]$ and $\left[A_{Wmo_{i,j}}\right]$. Complete elements of the each matrix (i.e., faults assigned in each position) are 120×9 and for better understanding in Table 7.2, Table 7.3, Table 7.4 and Table 7.5 only 120×1 elements are shown for various fault class matrices.



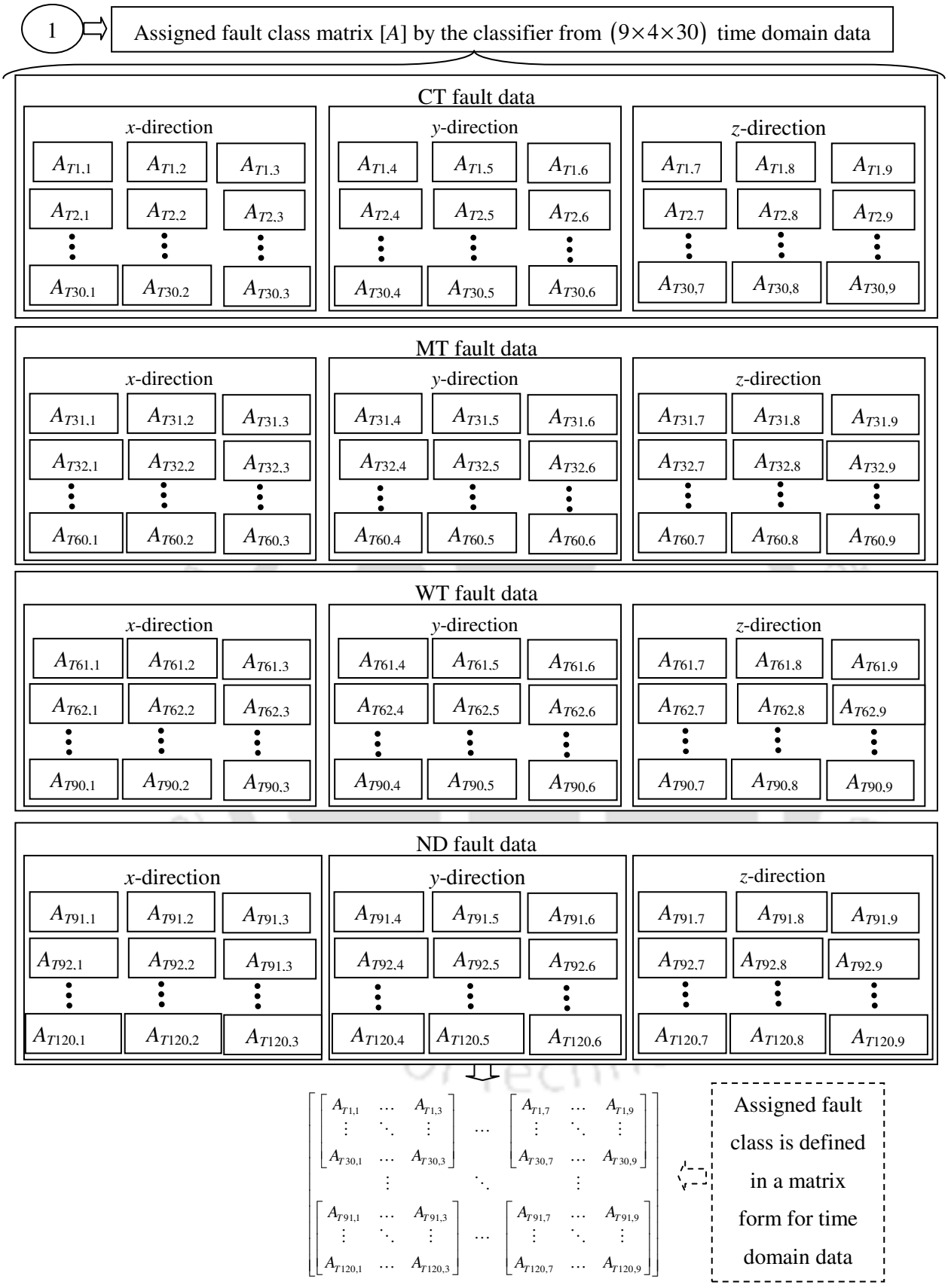


Figure 7.4 Flow chart for the classification of a time domain data set at the same speed conditions

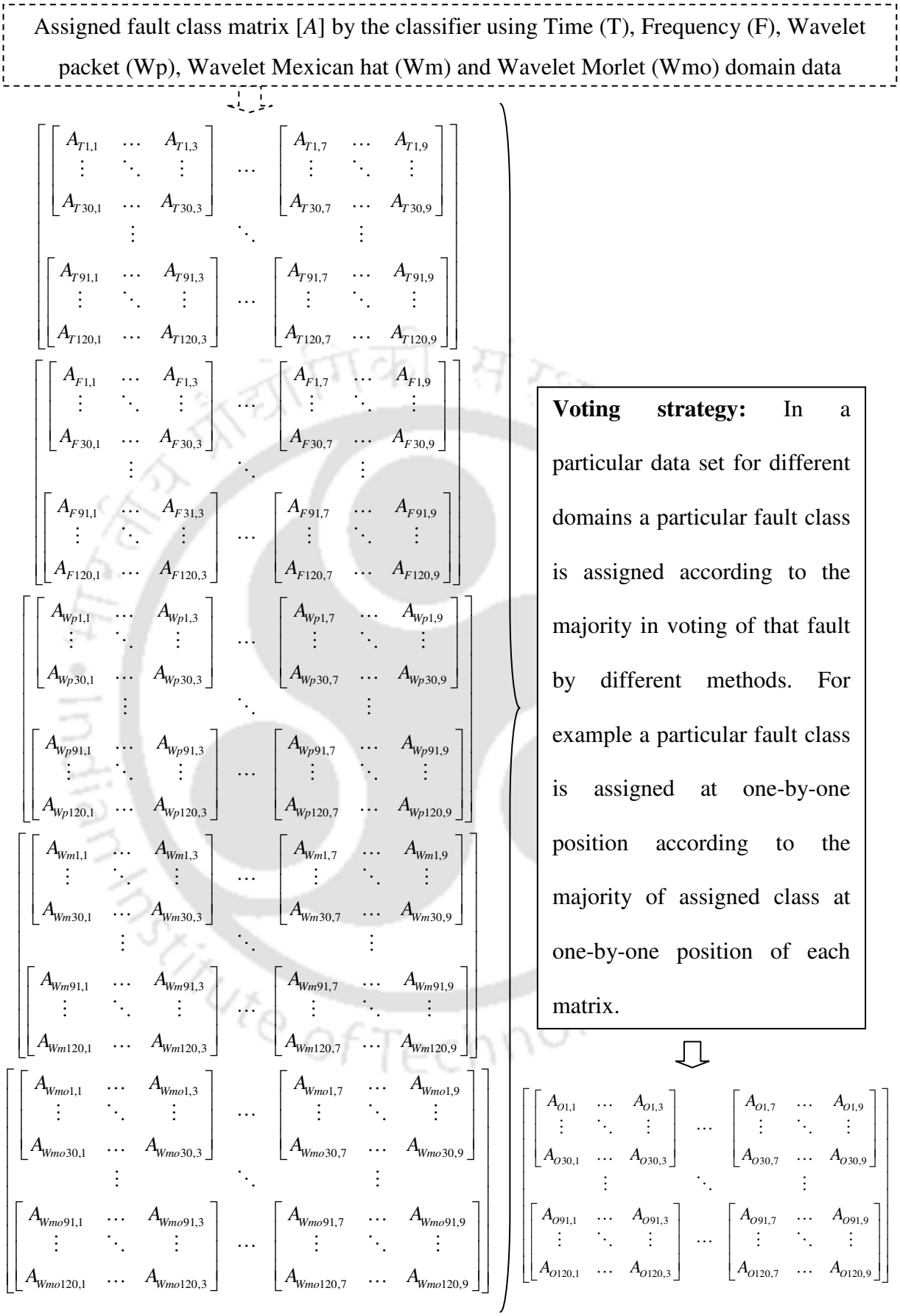


Figure 7.5 Flow chart for the improved classification at same speed condition and for one particular classifier

Table 7.1 Unified fault prediction accuracies (all classes) for the *C*-SVC

Speed (Hz)	Best accuracies (%) using					Unified accuracy (%)
	Time domain	Frequency domain	Wavelet packet	Wavelet Morlet	Wavelet Mexican hat	
For same speed						
10	97.50	100.00	98.33	96.67	92.50	**99.17**
15	90.83	97.50	90.00	99.17	94.17	**95.00**
20	98.33	98.33	98.33	100.00	95.83	**100.00**
25	100.00	95.00	100.00	99.17	95.00	**100.00**
30	100.00	88.33	100.00	99.17	91.67	**100.00**
For interpolation speed (range 5 Hz)						
12.5	87.50	90.00	88.33	95.83	75.00	**96.67**
17.5	94.17	84.17	95.00	99.17	93.33	**98.33**
22.5	95.83	67.50	96.67	99.17	96.67	**98.33**
27.5	100.00	82.50	99.17	99.17	92.50	**99.17**
For interpolation speed (range 10 Hz)						
15	48.33	66.66	25.00	50.83	25.00	**42.50**
20	65.83	66.67	75.83	92.50	56.67	**88.33**
25	93.33	74.17	90.00	98.33	82.50	**93.33**
For extrapolation speed (range 5 Hz)						
15	74.17	76.67	75.00	70.83	70.00	**80.83**
25	80.83	71.67	82.50	91.67	79.17	**89.17**
30	99.17	90.83	96.67	97.50	90.00	**100.00**
For extrapolation speed (range 10 Hz)						
20	45.00	54.17	42.50	49.17	39.17	**42.50**
25	55.00	67.50	55.00	68.33	57.50	**57.50**
30	72.50	70.83	70.83	77.50	70.00	**80.00**

Table 7.2, Table 7.3, Table 7.4 and Table 7.5 illustrate fault class matrix elements for CT, MT, WT and ND conditions, respectively. The second to sixth columns of each table represent the fault predictions in respective domains. However, the last column is obtained by the voting strategy.

Table 7.2 Elements fault class matrix for the CT fault data condition

Fault matrix / Elements	Time domain $\left[A_{T_{i,j}}\right]$	Frequency domain $\left[A_{F_{i,j}}\right]$	Wavelet packet $\left[A_{Wp_{i,j}}\right]$	Wavelet Mexican hat $\left[A_{Wm_{i,j}}\right]$	Wavelet morlet $\left[A_{Wmo_{i,j}}\right]$	Unified $\left[A_{Oi.j}\right]$
1	CT	CT	CT	CT	CT	CT
2	CT	CT	CT	CT	CT	CT
3	CT	CT	CT	CT	CT	CT
4	CT	CT	CT	CT	CT	CT
5	CT	CT	CT	CT	CT	CT
6	CT	CT	CT	CT	CT	CT
7	CT	CT	CT	CT	CT	CT
8	CT	CT	CT	**MT**	CT	CT
9	CT	CT	CT	CT	CT	CT
10	CT	CT	CT	CT	CT	CT
11	CT	CT	CT	CT	CT	CT
12	CT	CT	CT	CT	CT	CT
13	CT	CT	CT	CT	CT	CT
14	CT	CT	CT	CT	CT	CT
15	CT	CT	CT	CT	CT	CT
16	CT	CT	CT	CT	CT	CT
17	CT	CT	CT	CT	CT	CT
18	CT	CT	CT	CT	CT	CT
19	CT	CT	CT	CT	CT	CT
20	CT	CT	CT	CT	CT	CT
21	CT	CT	CT	CT	CT	CT
22	CT	CT	CT	**MT**	CT	CT
23	CT	CT	CT	CT	CT	CT
24	CT	CT	CT	CT	CT	CT
25	CT	CT	CT	CT	CT	CT
26	CT	CT	CT	CT	CT	CT
27	CT	CT	CT	CT	CT	CT
28	CT	CT	CT	CT	CT	CT
29	CT	CT	CT	CT	CT	CT
30	CT	CT	CT	CT	CT	CT

Table 7.3 Elements fault class matrix for the MT fault data condition

Fault matrix / Elements	Time domain $\left[A_{T_{i,j}}\right]$	Frequency domain $\left[A_{F_{i,j}}\right]$	Wavelet packet $\left[A_{Wp_{i,j}}\right]$	Wavelet Mexican hat $\left[A_{Wm_{i,j}}\right]$	Wavelet Morlet $\left[A_{Wmo_{i,j}}\right]$	Unified $\left[A_{Oi.j}\right]$
31	MT	MT	MT	MT	MT	MT
32	MT	MT	MT	MT	MT	MT
33	MT	MT	MT	MT	MT	MT
34	MT	MT	MT	MT	MT	MT
35	MT	MT	MT	**ND**	MT	MT
36	MT	MT	MT	MT	MT	MT
37	MT	MT	MT	MT	MT	MT
38	MT	MT	MT	MT	MT	MT
39	MT	MT	MT	MT	MT	MT
40	MT	MT	MT	MT	MT	MT
41	MT	MT	MT	MT	MT	MT
42	MT	MT	MT	MT	MT	MT
43	MT	MT	MT	MT	MT	MT
44	MT	MT	MT	MT	MT	MT
45	MT	MT	MT	MT	MT	MT
46	MT	MT	MT	MT	MT	MT
47	MT	MT	MT	**CT**	MT	MT
48	MT	MT	MT	MT	MT	MT
49	MT	MT	**ND**	**ND**	MT	MT
50	MT	MT	MT	MT	MT	MT
51	MT	MT	MT	MT	MT	MT
52	MT	MT	MT	MT	MT	MT
53	MT	MT	MT	**CT**	**CT**	MT
54	MT	MT	MT	MT	MT	MT
55	MT	MT	MT	MT	MT	MT
56	MT	MT	MT	MT	MT	MT
57	MT	MT	MT	MT	MT	MT
58	MT	MT	MT	MT	**ND**	MT
59	MT	MT	MT	MT	MT	MT
60	MT	MT	MT	**CT**	MT	MT

Table 7.4 Elements fault class matrix for the WT fault data condition

Fault matrix / Elements	Time domain $\left[A_{T_{i,j}}\right]$	Frequency domain $\left[A_{F_{i,j}}\right]$	Wavelet packet $\left[A_{Wp_{i,j}}\right]$	Wavelet Mexican hat $\left[A_{Wm_{i,j}}\right]$	Wavelet Morlet $\left[A_{Wmo_{i,j}}\right]$	Unified $\left[A_{Oi.j}\right]$
61	WT	WT	WT	WT	WT	WT
62	WT	WT	WT	WT	WT	WT
63	WT	WT	WT	WT	WT	WT
64	WT	WT	WT	WT	WT	WT
65	WT	WT	WT	WT	WT	WT
66	WT	WT	WT	WT	WT	WT
67	WT	WT	WT	**ND**	WT	WT
68	WT	WT	WT	WT	WT	WT
69	WT	WT	WT	WT	WT	WT
70	WT	WT	WT	WT	WT	WT
71	WT	WT	WT	WT	WT	WT
72	WT	WT	WT	WT	WT	WT
73	WT	WT	WT	WT	WT	WT
74	WT	WT	WT	WT	WT	WT
75	WT	WT	WT	WT	WT	WT
76	WT	WT	WT	WT	WT	WT
77	WT	WT	WT	WT	WT	WT
78	WT	WT	WT	WT	WT	WT
79	WT	WT	WT	WT	WT	WT
80	WT	WT	WT	WT	WT	WT
81	WT	WT	WT	WT	WT	WT
82	WT	WT	WT	WT	WT	WT
83	WT	WT	WT	WT	WT	WT
84	WT	WT	WT	WT	WT	WT
85	WT	WT	WT	WT	WT	WT
86	WT	WT	WT	WT	WT	WT
87	WT	WT	WT	WT	WT	WT
88	WT	WT	WT	WT	**ND**	WT
89	WT	WT	WT	WT	WT	WT
90	WT	WT	WT	WT	WT	WT

Table 7.5 Elements fault class matrix for the ND fault data condition

Fault matrix / Elements	Time domain $\left[A_{T_{i,j}}\right]$	Frequency domain $\left[A_{F_{i,j}}\right]$	Wavelet packet $\left[A_{Wp_{i,j}}\right]$	Wavelet Mexican hat $\left[A_{Wm_{i,j}}\right]$	Wavelet Morlet $\left[A_{Wmo_{i,j}}\right]$	Unified $\left[A_{Oi.j}\right]$
91	ND	ND	ND	ND	ND	ND
92	ND	ND	ND	ND	ND	ND
93	ND	ND	ND	ND	ND	ND
94	ND	ND	ND	ND	ND	ND
95	ND	ND	ND	ND	ND	ND
96	ND	ND	ND	ND	ND	ND
97	**CT**	ND	ND	ND	ND	ND
98	ND	ND	ND	ND	ND	ND
99	ND	ND	ND	ND	ND	ND
100	ND	ND	ND	ND	ND	ND
101	ND	ND	ND	ND	ND	ND
102	ND	ND	ND	ND	ND	ND
103	ND	ND	ND	ND	ND	ND
104	**WT**	ND	**WT**	ND	**WT**	**WT**
105	ND	ND	ND	ND	ND	ND
106	ND	ND	ND	ND	ND	ND
107	ND	ND	ND	ND	ND	ND
108	**CT**	ND	ND	ND	ND	ND
109	ND	ND	ND	ND	ND	ND
110	ND	ND	ND	ND	ND	ND
111	ND	ND	ND	ND	ND	ND
112	ND	ND	ND	ND	ND	ND
113	ND	ND	ND	ND	ND	ND
114	ND	ND	ND	ND	ND	ND
115	ND	ND	ND	ND	ND	ND
116	ND	ND	ND	ND	ND	ND
117	ND	ND	ND	ND	ND	ND
118	ND	ND	ND	ND	ND	ND
119	ND	ND	ND	ND	ND	ND
120	ND	ND	ND	ND	ND	ND

For example, at element number 104 of Table 7.5, using the time domain, wavelet packet and wavelet Morlet the fault prediction is WT, on the other hand by using frequency domain and wavelet Mexican hat the fault prediction is ND, and hence the unified fault prediction is

assigned as WT based on more number of voting for WT (i.e., 3) as compared to ND (i.e., 2 only). On the similar lines all other unified fault prediction is obtained and in Table 7.5 for all other element number the fault prediction is ND. Based on this now the prediction accuracy (all classes) is calculated according to Eqn. (3.20), which is 99.17% for the example discussed.

For the same rotational speed case, out of 5 different rotation speeds for 3 rotational speeds the fault prediction is perfect, and for 10 Hz and 15 Hz rotational speeds it is 99.17 % and 95.00%, respectively. For the interpolation rotational speed the prediction accuracies out of 7 speeds only one speed (15 Hz at range 10 Hz with prediction of 42.50 %) less than 50 % fault prediction. For the extrapolation rotational speed the prediction accuracies out of 6 speeds only one speed (20 Hz at range 10 Hz with prediction of 42.50%) the prediction accuracy is less than 50 %.

***Unified Fault Prediction Accuracies using ν-SVC Classifier*:** Table 7.6 illustrates unified prediction accuracies (all classes) for the same speed as well as for the interpolation and extrapolation rotational speeds, which are based on the ν-SVC classifier. The best prediction accuracies for a rotational speed are tabulated for a particular domain in row wise, which are the best values found in Chapter 4-6 by the three optimization techniques (GSM, GA, ABCA).

Unified prediction accuracies (all classes) are found by voting strategies, which are explained in above paragraph. For the same rotational speed case, out of 5 different rotation speeds for 4 rotational speeds the prediction is perfect and for only 15 Hz rotational speed it is 97.50%. At the interpolation rotational speed the prediction accuracy out of 7 speeds for 15 Hz speed at the range of 10 Hz the improved prediction accuracy (65.00%) is lowest. For the extrapolation rotational speed the prediction out of 6 speeds for 20 Hz speed at the range of 10 Hz the unified prediction accuracy (64.17%) is lowest.

Table 7.6 Unified fault prediction (all classes) accuracies for ν-SVC

Speed (Hz)	Best accuracies (%) using					Unified accuracy (%)
	Time domain	Frequency domain	Wavelet packet	Wavelet Morlet	Wavelet Mexican hat	
For same speed						
10	99.17	100.00	100.00	96.67	94.17	**100.00**
15	90.00	97.58	95.83	99.17	96.67	**97.50**
20	97.50	100.00	99.17	100.00	97.50	**100.00**
25	100.00	97.50	100.00	100.00	95.83	**100.00**
30	100.00	100.00	100.00	99.17	95.83	**100.00**
For interpolation speed (range 5 Hz)						
12.5	89.17	85.83	89.17	97.50	77.50	**97.50**
17.5	94.17	88.33	95.00	100.00	94.17	**100.00**
22.5	96.67	84.17	95.83	100.00	96.67	**98.33**
27.5	100.00	76.67	100.00	100.00	93.33	**100.00**
For interpolation speed (range 10 Hz)						
15	59.17	68.33	39.17	65.83	48.33	**65.00**
20	85.83	70.83	88.33	89.17	70.00	**94.17**
25	94.17	80.00	90.83	98.33	82.50	**96.67**
For extrapolation speed (range 5 Hz)						
15	75.83	75.83	70.83	76.67	76.67	**76.67**
25	90.83	75.00	90.83	91.67	85.00	**93.33**
30	100.00	88.33	97.50	97.50	90.83	**100.00**
For extrapolation speed (range 10 Hz)						
20	62.50	58.33	66.67	64.17	49.17	**64.17**
25	69.17	76.67	66.67	66.67	54.17	**68.33**
30	77.50	74.17	73.33	83.33	75.00	**90.00**

***Unified Fault Prediction Accuracies using C-SVC and ν-SVC Classifier*:** Table 7.7 illustrated overall unified prediction accuracies (all classes) for the same speed as well as for the interpolation and extrapolation rotational speeds. The best prediction accuracies for a rotational speed are tabulated for a particular domain in row wise, which are the best fault

predictions found in Chapter 4-6 from three optimization techniques (i.e., GSM, GA, ABCA). Unified prediction accuracies are found by voting strategies. But in this case the result of *C*-SVC and ν-SVC is considered together.

Table 7.7 Overall unified fault prediction accuracies (all classes)

Speed (Hz)	C-SVC best accuracy (%) using					ν-SVC best accuracy (%) using					Over all unified Accu. (%)
	Time domain	Frequency domain	Wavelet packet	Wavelet Morlet	Wavelet Mexican hat	Time domain	Frequency domain	Wavelet packet	Wavelet Morlet	Wavelet Mexican hat	
For same speed											
10	97.50	100.00	98.33	96.67	92.50	99.17	100.00	100.00	96.67	94.17	**100.00**
15	90.83	97.50	90.00	99.17	94.17	90.00	97.58	95.83	99.17	96.67	**96.25**
20	98.33	98.33	98.33	100.00	95.83	97.50	100.00	99.17	100.00	97.50	**100.00**
25	100.00	95.00	100.00	99.17	95.00	100.00	97.50	100.00	100.00	95.83	**100.00**
30	100.00	88.33	100.00	99.17	91.67	100.00	100.00	100.00	99.17	95.83	**100.00**
For interpolation speed (range 5 Hz)											
12.5	87.50	90.00	88.33	95.83	75.00	89.17	85.83	89.17	97.50	77.50	**97.50**
17.5	94.17	84.17	95.00	99.17	93.33	94.17	88.33	95.00	100.00	94.17	**99.17**
22.5	95.83	67.50	96.67	99.17	96.67	96.67	84.17	95.83	100.00	96.67	**98.33**
27.5	100.00	82.50	99.17	99.17	92.50	100.00	76.67	100.00	100.00	93.33	**100.00**
For interpolation speed (range 10 Hz)											
15	48.33	66.66	25.00	50.83	25.00	59.17	68.33	39.17	65.83	48.33	**59.17**
20	65.83	66.67	75.83	92.50	56.67	85.83	70.83	88.33	89.17	70.00	**94.17**
25	93.33	74.17	90.00	98.33	82.50	94.17	80.00	90.83	98.33	82.50	**96.67**
For extrapolation speed (range 5 Hz)											
15	74.17	76.67	75.00	70.83	70.00	75.83	75.83	70.83	76.67	76.67	**80.83**
25	80.83	71.67	82.50	91.67	79.17	90.83	75.00	90.83	91.67	85.00	**92.50**
30	99.17	90.83	96.67	97.50	90.00	100.00	88.33	97.50	97.50	90.83	**100.00**
For extrapolation speed (range 10 Hz)											
20	45.00	54.17	42.50	49.17	39.17	62.50	58.33	66.67	64.17	49.17	**54.17**
25	55.00	67.50	55.00	68.33	57.50	69.17	76.67	66.67	66.67	54.17	**63.33**
30	72.50	70.83	70.83	77.50	70.00	77.50	74.17	73.33	83.33	75.00	**83.33**

At the same rotational speed case, out of 5 different rotation speeds for 4 rotational speeds the fault prediction is perfect and for 15 Hz rotational speed it is 96.25%. For the interpolation rotational speed the fault prediction, out of 7 speeds the 15 Hz speed at range of 10 Hz the prediction (59.17%) is the lowest. For the extrapolation rotational speed the fault prediction out of 6 speeds for 20 Hz speed at range of 10 Hz the prediction (54.17%) is the lowest. It is observed that in all the cases the prediction accuracies is more than 50%. There are two cases

(speeds of 20 Hz and 25 Hz) at 10 Hz range of extrapolation speed the fault predictions are 54.17% and 63.33%, respectively.

Table 7.8 Unified percentage fault predictions (individual classes) at various speeds

Speed (Hz)	ND (%)	CT (%)	MT (%)	WT (%)
For same speed				
10	100.00	100.00	100.00	100.00
15	100.00	100.00	86.67	100.00
20	100.00	100.00	100.00	100.00
25	100.00	100.00	100.00	100.00
30	100.00	100.00	100.00	100.00
For interpolation speed (range 5 Hz)				
12.5	96.67	96.67	100.00	96.67
17.5	100.00	100.00	96.67	100.00
22.5	100.00	100.00	93.33	100.00
27.5	100.00	100.00	100.00	100.00
For interpolation speed (range 10 Hz)				
15	66.67	43.33	70.00	56.67
20	93.33	100.00	90.00	93.33
25	86.67	100.00	100.00	100.00
For extrapolation speed (range 5 Hz)				
15	96.67	66.67	63.33	96.67
25	83.33	86.67	100.00	100.00
30	100.00	100.00	100.00	100.00
For extrapolation speed (range 10 Hz)				
20	13.33	33.33	93.33	76.67
25	13.33	43.33	100.00	96.67
30	50.00	100.00	100.00	83.33

Based on the overall fault prediction (all classes) the individual fault predictions are illustrated in Table 7.8. The individual fault prediction for all the cases is 100.00% except for 15 Hz rotational speed the fault prediction at MT case is 86.67%. In the interpolation case the lowest individual speed prediction accuracy is 43.33% at 15 Hz speed (range 10 Hz) for the CT case.

Similarly for the extrapolation case the lowest prediction accuracy is 13.33%, which appears in two places at 20 Hz and 25 Hz speed (10 Hz range) for the ND case.

7.4 Comparison of Unified Fault Prediction Accuracy

The overall unified prediction accuracies (all classes) found from two classifiers and three domains are compared with the best prediction accuracies in time, frequency and time-frequency (wavelet) (which is tabulated in Table 7.7). The best prediction accuracies (all classes) at a particular speed and domain are considered here, for example at speed of 10 Hz the best prediction accuracies in time domain for two classifiers is 99.17% (which is selected as maximum of two classifier prediction accuracies of 97.50% and 99.17%).

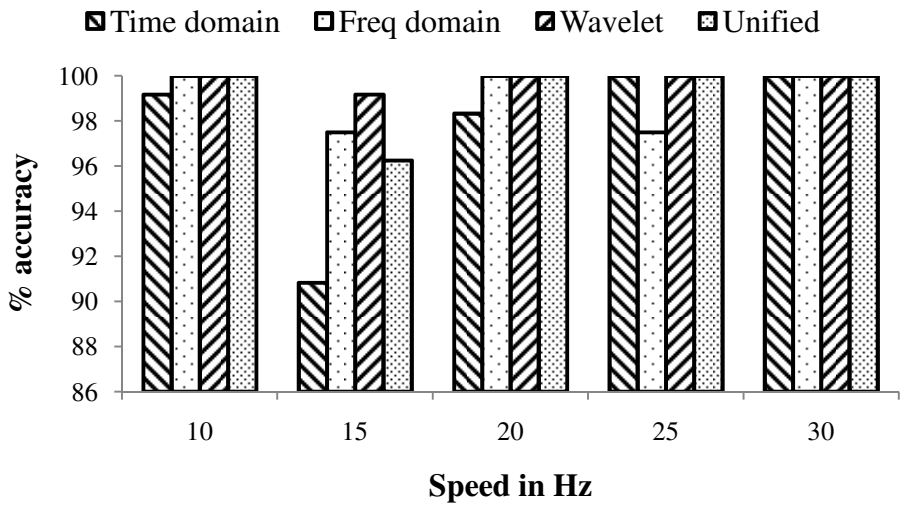
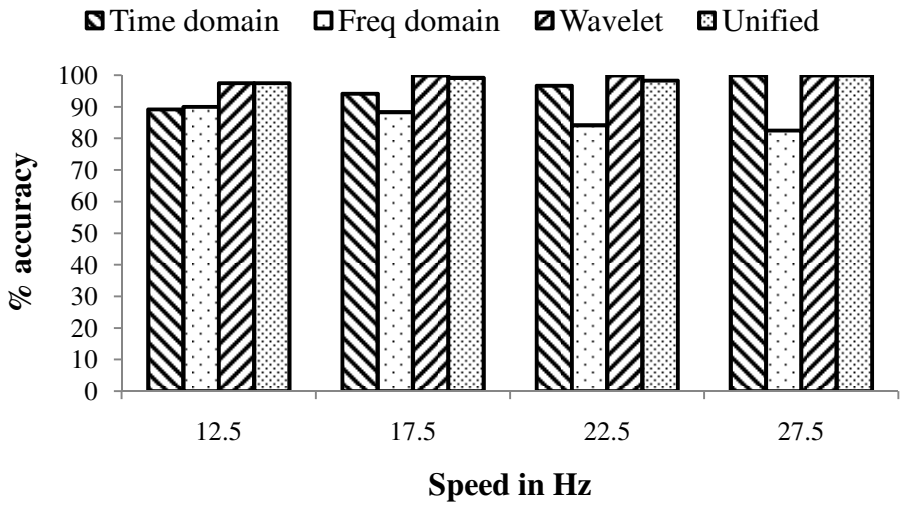


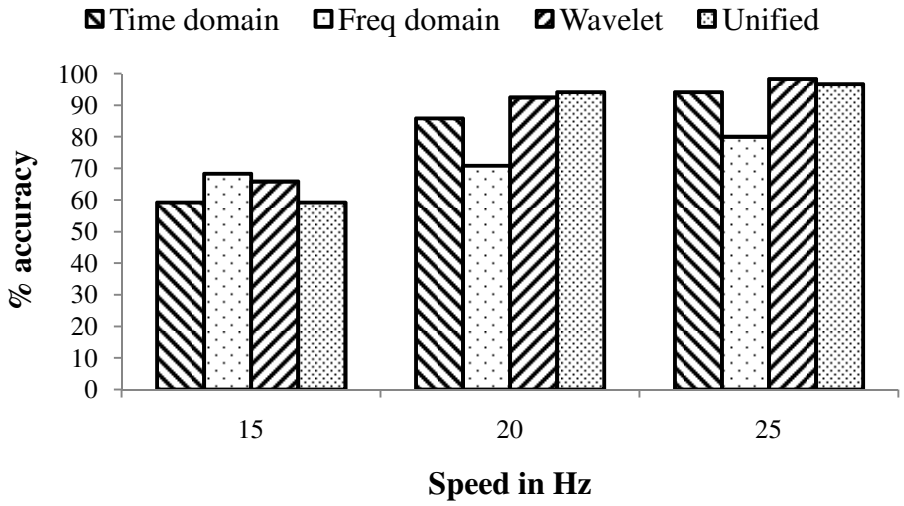
Figure 7.6 Comparison of the unified accuracy (all classes) with the best accuracy (all classes) using the time, frequency and time-frequency domain data at the same rotational speed

Figure 7.6 illustrates the comparison of prediction accuracies (all classes) at the same speed condition. Overall unified prediction accuracies (all classes) are equal or better to the best accuracy (all classes) for all speeds considered except at 15 Hz, in which the overall unified prediction accuracy (all classes) is 96.25% and the time, frequency and time-frequency domain

best prediction accuracies are 90.83%, 97.58% and 99.17%, respectively. In the above discussion it is found that 15 Hz rotational speed predictions are poor. This may happen due to the participation of structural modes, which diminishes the effect of faults in signals. However, to find the possible causes and overcome this, a thorough investigation is required.

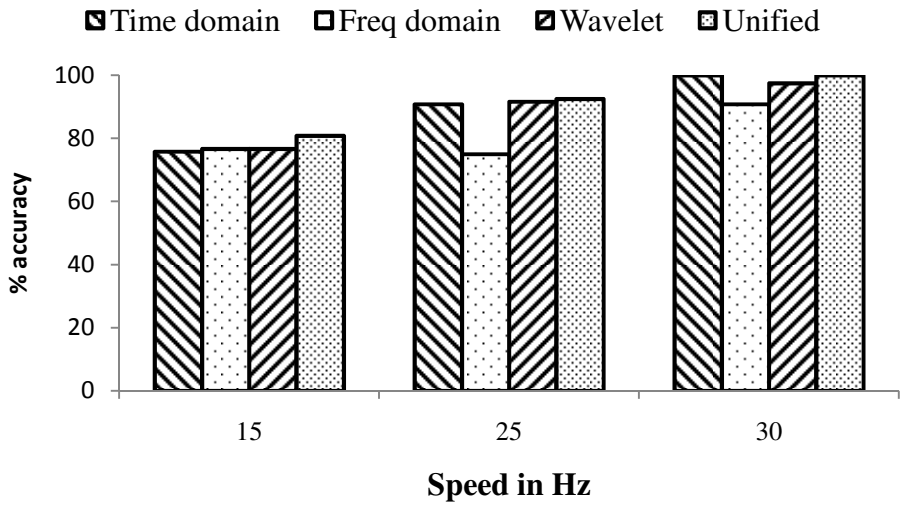


(a) Accuracy at the interpolation speed range of 5 Hz

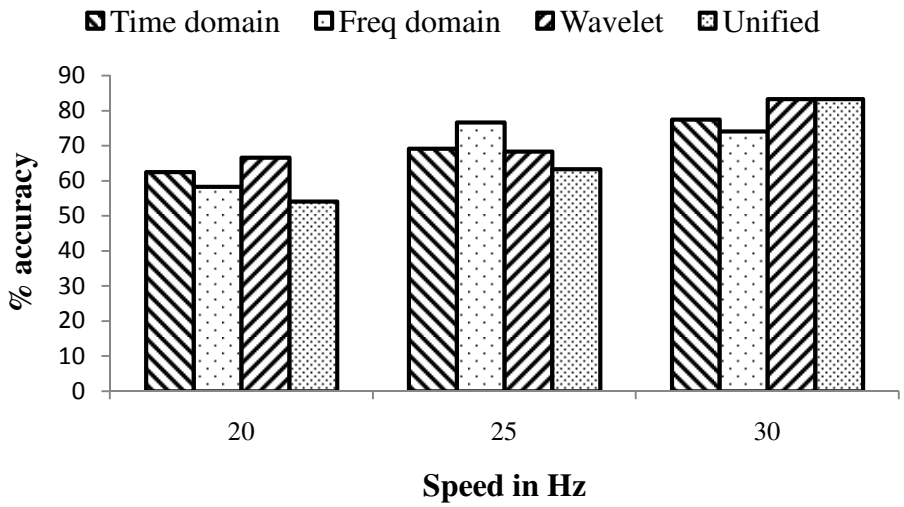


(b) Accuracy at the interpolation speed range of 10 Hz

Figure 7.7 Comparison of the unified accuracy (all classes) with the best accuracy (all classes) using the time, frequency and time-frequency domain data at the interpolation speed



(a) Accuracy at the extrapolation speed range of 5 Hz



(b) Accuracy at the extrapolation speed range of 10 Hz

Figure 7.8 Comparison of the unified accuracy (all classes) with the best accuracy (all classes) using the time, frequency and time-frequency domain data at the extrapolation speed

Figure 7.7 and Figure 7.8 illustrate the comparison of overall unified prediction accuracies (all classes) and the best prediction accuracies (all classes) in all three domains for the interpolation and extrapolation cases. Figures marked as (a) and (b) are for speed ranges of 5 Hz and 10 Hz, respectively. It is seen from the figure that in the case of interpolation speed range of 5 Hz at rotational speeds of 17.5 Hz and 22.5 Hz the interpolation unified prediction accuracy (all

classes) (99.17% and 98.33%, respectively) is lower than the wavelet best accuracy (both 100.00%). For other two cases the overall unified accuracy (all classes) is better. In the case of interpolation speed range of 10 Hz at rotational speed of 15 Hz and 25 Hz the unified prediction accuracy (all classes) (59.17% and 96.67%, respectively) is slightly lower than best prediction accuracies for that particular speed (68.33% in frequency domain and 65.83% in wavelet domain for 15 Hz rotational speed, and 98.33% of wavelet domain for 25 Hz rotational speed).

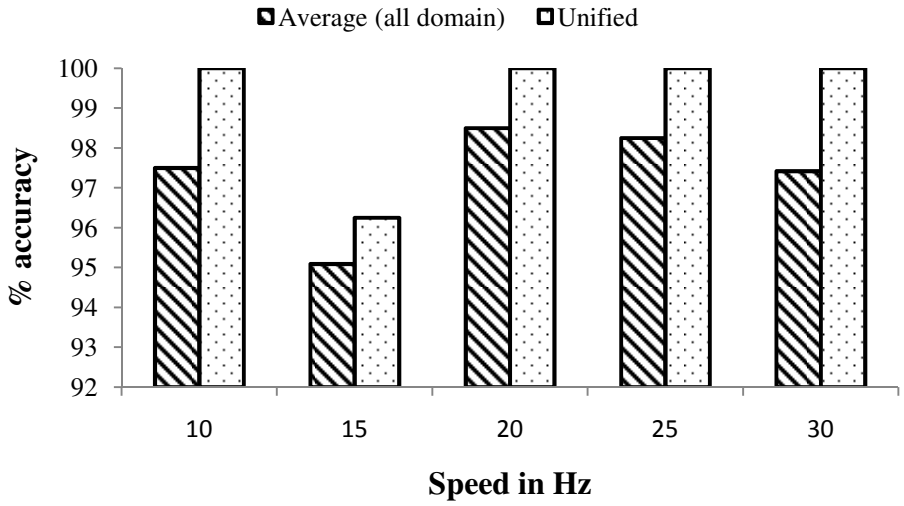
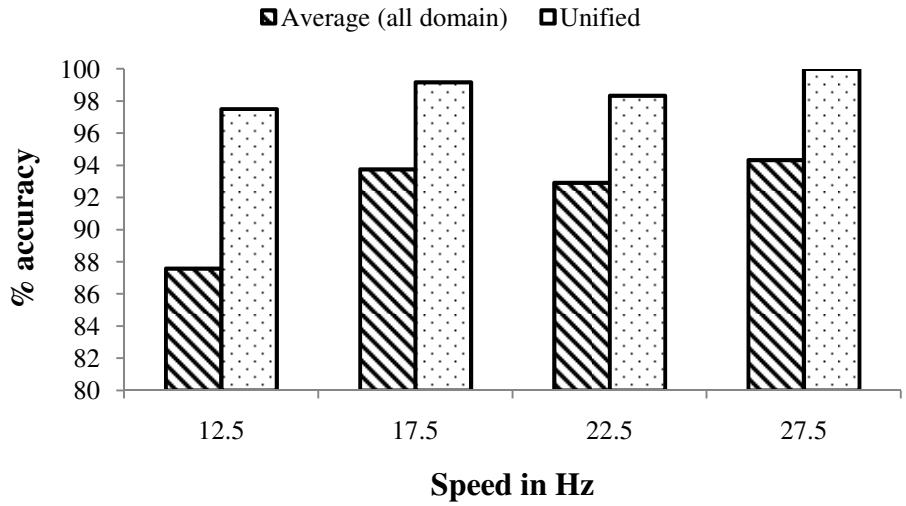
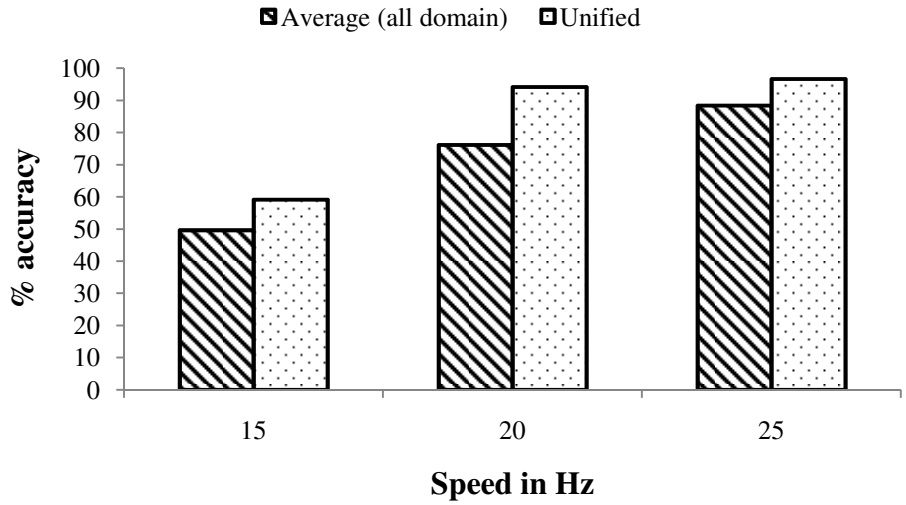


Figure 7.9 Comparison of the unified prediction accuracy with the average accuracy using the time, frequency and time-frequency domain data at the same rotational speed

In 5 Hz extrapolation speed range, unified prediction accuracies (all classes) are better than other best prediction accuracies (all classes) using the three domain data. But for 10 Hz extrapolation speed range at 20 Hz and 25 Hz rotational speeds the unified prediction accuracies (all classes) (54.17% and 63.33%, respectively) is slightly lower than the other three domain best prediction accuracies (all classes) (62.50%, 58.33%, 66.67% for 10 Hz speed and 69.17%, 76.67%, 68.33% for 20 Hz speed using the time, frequency and wavelet domain data, respectively).

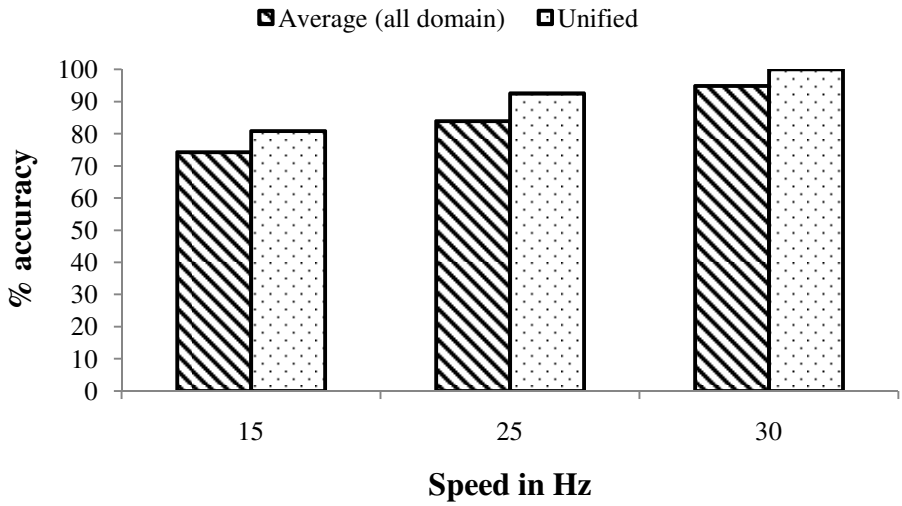


(a) Accuracy at the interpolation speed range of 5 Hz

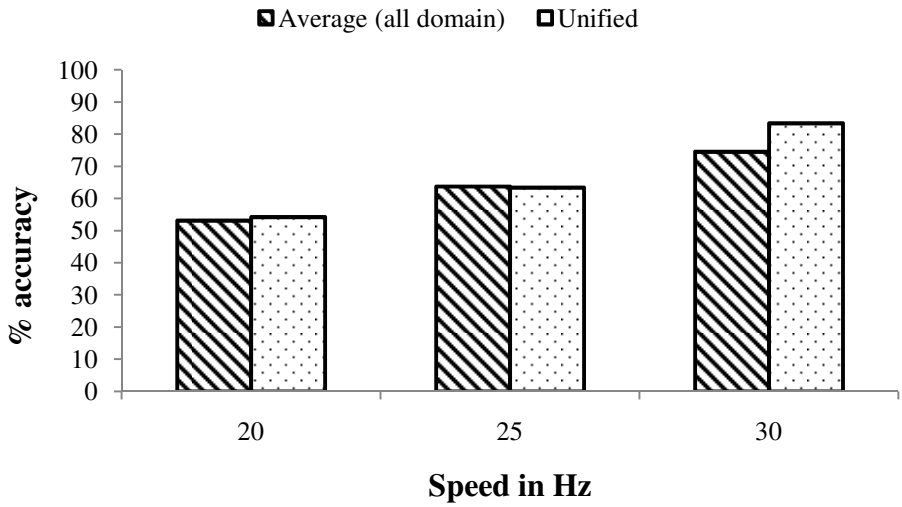


(b) Accuracy at the interpolation speed range of 10 Hz

Figure 7.10 Comparison of the unified accuracy (all classes) with the average accuracy using the time, frequency and time-frequency domain data at the interpolation speed



(a) Accuracy at the extrapolation speed range of 5 Hz



(b) Accuracy at the extrapolation speed range of 10 Hz

Figure 7.11 Comparison of the unified prediction accuracy (all classes) with the average accuracy using the time, frequency and time-frequency domain data at the extrapolation speed

The comparison is also performed by taking the average of the best prediction accuracies (instead of based on the voting) found using the three domain data and two classifiers with the unified prediction accuracies (all classes). The average of the best prediction accuracies is calculated considering values of the best prediction accuracies for a particular speed (row wise

values) from Table 7.7. For the same speed, interpolation speed and extrapolation speed the comparison is illustrated in Figure 7.9, Figure 7.10 and Figure 7.11, respectively. It is observed that in all cases the unified prediction accuracies (all classes) are better than average prediction accuracies except in 10 Hz extrapolation speed range at 25 Hz rotational speed it is slightly lower (the unified accuracy is 63.33% and the average accuracy is 63.67%).

From the above discussion, it is observed that unified prediction accuracies (all classes) found from the voting strategy is comparatively sound in taking the effect of average prediction accuracies found using the three domains and two classifies, except one case for 10 Hz extrapolation speed range with 25 Hz rotational speed. On the other hand unified prediction accuracies (all classes) are better than best prediction accuracies (all classes) of various domains and it is better in 4 cases out of 5 cases at the same speed case, 3 cases out of 7 cases in interpolation speed and 4 cases out of 6 cases in extrapolation speed. This is the indication that the voting strategy is one of the promising strategies instead of simply taking the average of best prediction accuracies (all classes) for finding the unified prediction accuracies for better decision making.

7.5 Summary

A comparison of fault predictions estimated in the time, frequency and time-frequency domains has been presented. It is apparent that the wavelet based fault predictions is more or less leading and among them the Morlet wavelet family fault prediction are the best. Subsequently, an attempt to find out new unified prediction accuracies (all classes) from prediction accuracies obtained using the different domain data and optimization techniques. Unified prediction accuracies (all classes) are presented individually for the C-SVC and the ν-SVC as well as in their combination. The unified prediction accuracies (all classes) are calculated based on voting strategies and results are presented. Unified prediction accuracies (all classes) are compared

with the best prediction accuracies (all classes) of a particular domain (with consideration of both classifiers) and with the average prediction accuracies found from the three domains and two classifies, separately. Fault prediction results indicate the soundness of the voting strategy. In the next chapter overall conclusions and future scopes will be presented.

CHAPTER 8

Conclusions and Future Scopes of Work

8.1 Overview of the Present Work

It has been established that the SVM has a capability to classify gear faults at the same rotational speed, but it is rare to see the fault classification ability at the intermediate and extrapolated speeds, at which speed the vibration data is not available. Taking into consideration this point, the present work attempts to study the fault prediction based on the SVM for the multiple fault classification of gearbox by training the classifier at the beginning and at the end of rotational speeds in a particular band of rotational speed, and then faults are classified at the intermediate speed and out of the speed band. The potential of two SVM classifiers (C-SVC and ν-SVC) with respect of its tuning of parameters by using optimizing tools, like the GA, and the ABCA; and their performance are studied and compared with the classical optimisation technique that is the GSM. Finally, a novel technique is suggested to find the unified prediction accuracies from prediction accuracies obtained by using three domain data (i.e., the time, frequency and time-frequency), individually.

8.2 Major Conclusions of the Present Work

Major conclusions of the thesis have been summarised as follows.

- Fault classification ability to classify four gear faults (CT, MT, WT and ND) at the intermediate speed and extrapolated speed as well as at the same rotational speed by the SVM is demonstrated and it is found to be promising.
- The potential of two SVM classifiers (C-SVC and ν-SVC) with respect to tuning of the SVM kernels (i.e. RBF) parameters by using two optimizing tools (i.e. the GA and the ABCA) and comparison with the classical technique (i.e., GSM) show consistency in fault predictions. However, the GA and the ABCA show its ability in the form of the improved accuracy as compared to the GSM for several occasions.

- Gear fault features are extracted from the vibration data of a gear box. Only three statistical features (standard deviation, kurtosis and skewness) are used from the time, frequency and time-frequency (wavelet) domain data for effective fault classification.
- The time and FFT data is used for calculation of statistical features. The prediction accuracy results of frequency domain show improvement of fault predictions with respect to time domain data in several occasions.
- Wavelet features are calculated from time domain vibration data. The continuous wavelet transform (CWT) and wavelet packet transform (WPT) is used for obtaining the wavelet coefficients of vibration signals.
- Faults are classified by using the vibration data from three domains and fault predictions are compared. Results show an increasing trend of fault prediction performance while using the time, frequency and time-frequency domain data, respectively. Among the wavelet family, the Morlet wavelet shows the best performance over the other wavelet family as well as for time and frequency domain data.
- Unified prediction accuracies have been obtained on the basis of a novel voting strategy, which is carried out individually for two classifiers for vibration data of three domains. Finally an over all unified accuracy has been found.

8.3 Contribution of the Present Work

- Fault classifier parameters are optimized by the ABCA technique to get better optimization values.
- Classification of faults at interpolation/extrapolation rotational speeds by the classifier for which the training data is not available.
- The three statistical features are calculated from whole FFT data points and wavelet coefficients for improvement of classifications**.**

- Prediction accuracies found from different classifiers and domains are unified for better decision making by a novel voting strategy.

8.4 Applicability and Limitations

The algorithm developed in the present work has been used for the classification of the gear faults with vibration data in time domain or frequency domain or both. The method also gives a direction to choose an optimal accuracy among different prediction accuracies found by using different domain data as well with different classification methods. Methods developed can be used in condition monitoring of other kind of machine elements, like bearing, shaft misalignment, motor faults, shaft cracks, etc. There is a lot of post processing of the vibration data is required, and hence it can be used in off-line condition only. In the actual situation, bearings, shafts, couplings etc. are also included in the gearbox. The processing of vibration signatures generated from faults of several machinery elements are still a challenging issue.

8.5 Future Scopes of Work

- Fault predictions in present study are based on SVM kernel of the radial basis function (RBF) with the time, frequency and time-frequency domain data. The study can be extended with the use of other SVM kernels, like the linear, polynomial or combination of different kernels.
- In present study, it is observed that the fault prediction increases when the noise is reduced in signals. A proper pre-conditioning and removal of noise at the time of acquisition of signals (or by post processing) by using filters can be studied.
- In present study, features were calculated from vibration signatures of the gear box. Other signatures like acoustics can be used for finding features. Not only the three statistical features as discussed in present work, other statistical features (like higher statistical moments and probability density function) can be included in the fault classification.

- The optimization technique has been used for finding optimal parameters of the SVM. The optimization technique could also be used for the selection of effective features for fault classification. The selection of features can also be studied by conventional technique like distance evaluation technique and fault predictions could be compared with that from optimization techniques.
- The entire discussions and procedures developed are off-line and tested on the gearbox faults only. The procedure can be used in condition monitoring of bearings, misalignments of shaft, motor faults, etc. The procedure can be modified and tested for online condition monitoring by using the regression model.
- In the above discussion it is found that 15 Hz rotational speed predictions are poor. This may happen due to the participation of structural modes, which diminishes the effect of faults in signals. However, to find the possible causes and overcome this, a thorough investigation is required.
- The unified prediction technique explained in Chapter 7 is based on the majority voting strategy. The study can be extended by considering a dynamically weighted voting strategy based on prior classification performance of the classifier/optimizer/feature set combination.
- The entire study is based on the vibration signature collected by the accelerometer and three features. The collection of signatures from other sources like thermal, noise, acoustics is one of the interesting aspects to be investigated. Also selections of optimal features and sensor fusion (temperature, vibration, acoustics etc.) are to be investigated.

BIBLIOGRAPHY

Ayoubi M. and Isermann R., 1997, "Neuro-fuzzy systems for diagnosis", Fuzzy Sets and Systems, 89, 289-307.

Austerlitz H., 2003, "Data Acquisition Techniques using PCs", Academic Press, San Diego, CA.

Ali S., Mohammad J. M., Masoud S., 2009. "Fault classification in gears using support vector machines (SVMs) and signal processing". IEEE Transaction, 9781-4244-3428-2/09.

Boser B. E., Guyon I. and Vapnik V., 1992. "A training algorithm for optimal margin classifiers". Proceedings of the Fifth Annual Workshop on Computational Learning Theory, ACM Press, pages 144-152.

Bishop C.M., 1995, "Neural Networks for Pattern Recognition", Oxford University Press, Oxford.

Burgess C. J. C., 1998. "A tutorial on support vector machines for pattern recognition". Data Mining and Knowledge Discovery, 2, 955–974.

Collacott R. A., 1977, "Mechanical fault diagnosis and condition monitoring", Chapman & Hall, London, UK.

Coifman R. R. and Wickerhauser M.V., 1992, "Entropy-based algorithms for best basis selection", IEEE Transaction on Information Theory, 38 (2), 713-718.

Cameron B. G. and Stuckey M. J., 1994. "A review of transmission vibration monitoring at Westland Helicopter Ltd.". In Proceedings of the 20th European Rotorcraft Forum, 16-116.

Cohen L., 1995, "Time-frequency analysis", Englewood Cliffs, Prentice Hall, NJ.

Cortes C. and Vapnik V., 1995. Support-vector network, machine learning, 20, 273-297.

Chang, S.H., Kang K.S., Choi S.S., Kim H.G., Jeong H.K. and Yi C.U., 1995, "Development of the on-line operator aid system OASYS using a rule-based expert system and fuzzy logic for nuclear power plants," Nuclear Technology, Vol. 112, 266-294.

Crisp D. J. and Burges C. J. C., 2000. "A geometric interpretation of ν-SVM classifiers". In S. Solla, T. Leen, and K.-R. Muller, editors, Advances in Neural Information Processing Systems, volume 12, Cambridge, MA, MIT Press.

Chang C.-C. and Lin C.-J., 2001. "Training ν-support vector classifiers: Theory and algorithms". Neural Computation, 13(9), 2119-2147.

Chen Z.Y., He Y.Y., Chu F.L., and Huang J.Y., 2003, "Evolutionary strategy for classification problems and its application in fault diagnostics," Engineering Applications of Artificial Intelligence, Vol. 16, 31-38.

Cho S., Asfour S., Onar A., Kaundinya N., 2005, "Tool breakage detection using support vector machine learning in a milling process", International Journal of Machine Tools & Manufacture, 45, 241–249.

Cheng Junsheng, Yu Dejie, and Yu Yang, 2008. "A fault diagnosis approach for gears based on IMF AR model and SVM". EURASIP Journal on Advances in Signal Processing, 1-7.

Chang C.-C. and Lin C.-J., 2011. "LIBSVM: A library for support vector machines". ACM Transactions on Intelligent Systems and Technology, 2:27:1-27:27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Downham E, 1976. "Vibration in rotating machinery: Ma;fuction diagnosis-Art and science", Proceedings of the Institute of Mechanical Engineers-Vibration in rotating Machinery, 1-6.

Dalpiaz G., Rivola A. and Rubini R., 2000, "Effectiveness and sensitivity of vibration processing techniques for local fault detection in gears", Mechanical System and Signal Processing, 3, 387–412.

Duda R.O., 2001, "P.E. Hart, D.G. Stock, Pattern Classification", Wiley, New York.

Deb K., 2003. "Multi-Objective Optimization using Evolutionary Algorithms". John Willey & Sons, Ltd., Chichester.

Deb K., 2005. "Optimization for Engineering Design: algorithms and examples". Printice-Hall of India Pvt. Ltd, New Delhi.

Du R. and Yeung K., 2004, "Fuzzy transition probability: A new method for monitoring progressive faults. Part 1: The theory," Engineering Applications of Artificial Intelligence, Vol. 17, pp. 457-467.

Fang R., 2006, "Induction machine rotor diagnosis using support vector machines and rough set", Lecture Notes in Artificial Intelligence, 4114, 631–636.

Fodor I.K., A Survey of Dimension Reduction Techniques, US Department of Energy, available in (http://www.llnl.gov/tid/lof/documents/pdf/240921.pdf).

GA software (GAOT, 1996) website http://www.ise.ncsu.edu/kay/.

Goldberg D, 1989, "Genetic Algorithms in search, Optimization, and Machine Learning", Addison-Wesley.

Gao Lixin, Ren Zhiqiang, Tang Wenliang, Wang Huaqing and Chen Peng, 2010. "Intelligent gearbox diagnosis methods based on SVM, wavelet lifting and RBR". Journal of Sensor, 10, 4602-4621.

Hill J.W. and Baines N.C., 1988. "Application of an expert system to rotating machinery health monitoring", Proceedings of the Institution of Mechanical Engineers-Vibrations in Rotating Machinery, 449-454.

Houck C. R., Joines J. and Kay M., 1996. "A genetic algorithm for function optimization: A Matlab implementation". ACM Transactions on Mathematical Software. Software available at http://www.ise.ncsu.edu/kay.

Hsu C.-W. and Lin C.-J., 2002a. "A comparison of methods for multi-class support vector machines". IEEE Transactions on Neural Networks, 13(2), 415-425.

Huang, Y.C. and Huang, C.M., 2002, "Evolving wavelet networks for power transformer condition monitoring," IEEE Trans. Power Delivery, Vol. 17, 412-416.

Han L., Ding L., Yu J., Li Q., Liang Y., 2004, "Power plant boiler air preheater hot spots detection system based on least square support vector machine", Lecture Notes in Computer Science, 12, 598–604.

Huang C-L., Wang C-J., 2006, "A GA-based feature selection and parameters optimization for support vector machines". Expert Systems with Applications, 31, 231–240.

Hu Q., He Z., Zhang Z., Zi Y., 2007, "Fault diagnosis of rotating machinery based on improved wavelet package transform and SVM ensemble", Mechanical System and Signal Processing, 21 (2), 688–705.

Hsu C-W, Chang C-C, Lin C-J, 2010, "A Practical Guide to Support Vector Classification", Department of Computer Science National Taiwan University, Taipei 106, Taiwan (http://www.csie.ntu.edu.tw/~cjlin).

Jolliffe I.T.,1986. "Principal Component Analysis". Springer-Verlag.

Joines J. and Houck C., 1994. "On the use of non-stationary penalty functions to solve constrained optimization problems with genetic algorithms", In 1994 IEEE International Symposium Evolutionary Computation, Orlando, Florida, 579-584.

Jack L. B. and Nandi A. K., 2002. "Fault Detection using Support Vector Machines and Artificial Neural Networks, augmented by Genetic Algorithms". Mechanical Systems and Signal Processing, 16(2–3), 373–390.

Jardine Andrew K.S., Lin Daming, Banjevic Dragan, 2006, "A review on machinery diagnostics and prognostics implementing condition-based maintenance", International Journal of Mechanical Systems and Signal Processing, 20, 1483–1510.

Knerr S., Personnaz L. and Dreyfus G., 1990. "Single-layer learning revisited: a stepwise procedure for building and training a neural network". Neurocomputing: Algorithms, Architectures and Applications, NATO ASI, Springer-Verlag, Berlin, 41-50.

Kressel U. H. G., 1998. "Pairwise classification and support vector machines". In Advances in Kernel Methods–Support Vector Learning, B. Scholkopf, C. J. C. Burges, and A. J. Smola, Eds., MIT Press, Cambridge, MA, 255–268.

Kirianaki N.V., Yurish S.Y., Shpak N.O., Deynega V.P., 2002. "Data Acquisition and Signal Processing for Smart Sensors", Wiley, Chichester, West Sussex, England.

Karaboga D., Basturk B., 2007. "A powerful and Efficient Algorithm for Numerical Function Optimization: Artificial Bee Colony (ABC) Algorithm". Journal of Global Optimization, 39(3), 459-171. Software available at http://www.mf.erciyes.edu.tr/abc.

Karaboga D., Basturk B., 2008. "On the performance of artificial bee colony (ABC) algorithm", Applied Soft Computing 8(1), 687–697.

Leblanc, J. F. A., Dube, J. R. F., & Devereux, B., 1990. "Helicopter gearbox vibration analysis in the Canadian forces – applications and lessons". In Proceedings of the first international conference, gearbox noise and vibration, IMechE Cambridge, UK, 173-177.

Liu T.I., Singonahalli J.H., Iyer N.R., 1996, "Detection of roller bearing defects using expert system and fuzzy logic", Mechanical Systems and Signal Processing, Vol. 10, pp. 595-614.

Lin J. and Qu L., 2000, "Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis", Journal of Sound and Vibration, 234, 135–148.

Lou X. and Loparo K.A., 2004, "Bearing fault diagnosis based on wavelet transform and fuzzy inference," Mechanical Systems and Signal Processing, Vol. 18, 1077-1095.

Li Y., Chai Y.Z., Yin R.P., Xu X.M., 2005, "Fault diagnosis based on support vector machine ensemble", in: Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, 3309–3314.

Liu S., Jia C-Y., Ma H., 2005, "A new weighted Support Vector Machine with GA-based parameter selection". Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 18-21 August 2005.

LIBSVM (2011), Version 3.1. http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Mallat S., 1989**, "**A theory for multiresolution signal decomposition: the wavelet representation", IEEE Pattern Analysis and Machine Intelligence, 11(7), 674-693.

Michalewicz Z., 1994, "Genetic Algorithms + Data Structures = Evolution Programs", AI Series, Springer-Verlag, New York.

Mallat S., 1998, "A wavelet tour of signal processing", Academic Press.

Mechefske C.K., 1998, "Objective machinery fault diagnosis using fuzzy logic," Mechanical Systems and Signal Processing, Vol. 12, 855-862.

MATLAB (2009), Version 2, Package is used for the ABC algorithm. http://www.mf.erciyes.edu.tr/abc .

Rioul O. and Vetterli M., 1991, "Wavelet and signal processing", in IEEE Signal Processing Magazine, 84(4), 14-38.

Ramesh R., Mannan M.A., Poo A.N., Keerthi S.S., 2003, "Thermal error measurement and modeling in machine tools, Part II: hybrid Bayesian network-support vector machine model", International Journal of Machine Tools & Manufacture, 43, 405–419.

Rajas S. A., Fernandez-Reyes D., 2005, "Adapting Multiple Kernel Parameters for Support Vector Machines using Genetic Algorithms". IEEE Congress on Evolutionary Computation (CEC-2005), 626-631.

Ren Q., Ma X., Miao G., 2005, "Application of support vector machine in reciprocating compressor valve fault diagnosis", Lecture Notes in Computer Science 12, 81–84.

Stewart, R. M., 1976, "Vibration analysis as an aid to the detection and diagnosis of faults in rotating machinery", Proceedings of the Institution of Mechanical Engineers-Vibrations in Rotating Machinery, 223-229.

Smith D. M., 1980. "Recognition of causes of rotor vibration in turbomachinery", Proceedings of the Institution of Mechanical Engineers_Vibration in Rotating Machinery, 1-4.

Staszewski W. J. and Tomlinson G. R., 1994, "Application of the wavelet transform to fault detection in a spur gear", Mechanical System and Signal Processing, 8, 289–307.

Smalley A.J., Baldwin R. M., Mauney D.A. and Millwater H. R., 1996. "Towards risk based criteria fro rotor vibration", Proceedings of the Institute of Mechanical Engineers-Vibration in Rotating Machinery, 517-527.

Scholkopf B., Smola A., Williamson R. C. and Bartlett P. L., 2000. "New support vector algorithms". Neural Computation, 12, 1207-1245.

Samanta B., Al-Balushi K.R., Al-Araimi S.A., 2003. "Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection". Engineering Applications of Artificial Intelligence, 16, 657–665.

Siddique A., Yadava G.S. and Singh B., 2003, "Applications of artificial intelligence techniques for induction machine stator fault diagnostics: Review," Proceedings of the IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, New York, 29-34.

Samanta B., 2004. "Gear fault detection using artificial neural networks and support vector machines with genetic algorithms", Mechanical Systems and Signal Processing, 18, 625–644.

Sun J., Rahman M., Wong Y.S., Hong G.S., 2004, "Multiclassification of tool wear with support vector machine by manufacturing loss consideration", International Journal Machine Tools & Manufacture, 44, 1179–1187.

Sun J., Hong G.S., Rahman M., Wong Y.S., 2004, "The application of nonstandard support vector machine in tool condition monitoring system", in: Proceedings of the Second IEEE International Workshop on Electronic Design, Test and Applications, 1–6.

Soman K. P. and Ramachandran K. I., 2005, "Insight into wavelets from theory to practice", Prentice-Hall of India Private Limited.

Sugumaran V., Muralidharan V., Ramachandran K.I., 2007, "Feature selection using decision tree and classification proximal support vector machine for fault diagnostic of roller bearing", Mechanical System and Signal Processing 21, (2), 930–942.

Saravanan N., Siddabattuni V.N.S. Kumar, Ramachandran K.I., 2008. "A comparative study on classification of features by SVM and PSVM extracted using Morlet wavelet for fault diagnosis of spur bevel gear box", Expert Systems with Applications, 35, 1351–1366.

Saravanan N., Cholairajan S., Ramachandran K.I., 2009. "Vibration-based fault diagnosis of spur bevel gear box using fuzzy technique". Expert Systems with Applications, 36, 3119–3135.

Saravanan N., Siddabattuni V.N.S. Kumar, Ramachandran K.I., 2010. "Fault diagnosis of spur bevel gear box using artificial neural network (ANN) and proximal support vector machine (PSVM)". Applied Soft Computing, 10, 344–360.

Samadzadegan F., Soleymani A. and Abbaspour R. Ali, 2010, "Evaluation of Genetic Algorithms for Tuning SVM Parameters in Multi-Class Problems". 11th IEEE International Symposium on Computational Intelligence and Informatics , 18–20 November, 2010, Budapest, Hungary.

Thomas D. L., 1984, "Vibration monitoring strategy for large turbo-generators", Proceedings of the Institution of Mechanical Engineers-Vibrations in Rotating Machinery, 91-99.

Taylor J. I., 1995. "Back to the basics of rotating machinery vibration analysis". Sound and Vibration, 29(2), 12-16.

Tsang, A.H.C., 1995, "Condition-based maintenance: tools and decision making," Journal of Quality in Maintenance Engineering, Vol. 1, 3-17.

Toyota T., Niho T., Chen P., 1999, “Failure Detection and Diagnosis of Rotating Machinery by Orthogonal Expansion of Density Function of Vibration Signal”, IEEE Transaction, 886-891.

Tax D.M.J. , Ypma A., Duin R.P.W., 1999, “Pump failure determination using support vector data description”, Lecture Notes in Computer Science 415–425.

Tiwari R., Tanwar N., and Mishra S.,2009. “Application of support vector machine techniques for heath monitoring of gears”. National Conference on Condition Monitoring (NCCM-2009), NSTL, Visakhapatnam, India.

Venkatasubrsmanian, V., 1944, “Towards integrated process supervision: current status and future directions,” Proceedings of the IFAC International Conference on Computer Software Structures, Sweden, pp. 1-13.

Vapnik V. N., 1995. “The Nature of Statistical Learning Theory”. New York, Springer.

Wang W. Q., Ismail F. and Golnaraghi M. F.,2001, “Assessment of gear damage monitoring techniques using vibration measurements”, Mechanical System and Signal Processing, 5, 905–922.

Weiwu Yan, Huihe Shao, “Application of Support Vector Machine Nonlinear Classifier to Fault Diagnoses”, 2002. Proceedings *of* the 4th World Congress on Intelligent Control and Automation, June 10-14,, Shanghai, P.R.China.

Waeyenbergh G., and Pintelon, L., 2002, “A framework for maintenance concept development,” International Journal of Production Economics, Vol. 77, 299-313.

Widodo A., Yang B.S., 2006, “Faults detection and classification of induction motor using wavelet support vector machine”, in: Korean Society of Power System Engineering (KSPSE) Conference, 79–84.

Widodo A., Yang B.S., 2007, "Wavelet support vector machine for machine faults classification", Lecture Series on Computer and Computational Sciences, 8, 1–5.

Widodo Achmad, Yang Bo-Suk, 2007. "Support vector machine in machine condition monitoring and fault diagnosis". Mechanical Systems and Signal Processing, 21, 2560-2574.

Xuan Jianping, Jiang Hanhong, Shi Tielin, and Liao Guanglan, 2005. "Gear fault classification using genetic programming and support vector machines". International Journal of Information Technology, 9, 19-27.

Xu Y., Wang L., 2005, "Fault diagnosis system based on rough set theory and support vector machine", Lecture Notes in Computer Science, 980–988.

Yang B.S., Lim D.S, An J.L., 2000, Vibration diagnostic system of rotating machinery using artificial neural network and wavelet transform, in: Proceedings of 13th International Congress on COMADEM, Houston, USA, 12–20.

Yan G.T. and Ma G.F., 2004, "Fault diagnosis of diesel engine combustion system based on neural networks," Proceedings of the 2004 International Conference on Machine Learning and Cybernetics, Vol. 5, Shanghai, China, 3111-3114.

Yang B.S., Han T., An J.L., 2004, "ART-KOHONEN neural network for faults diagnosis of rotating machinery", Mechanical System and Signal Processing, 18 (3), 645–657.

Yang B.S., Jeong S.K., Oh Y.M., Tan A.C.C., 2004, "Case-based reasoning system with Petri nets for induction motor fault diagnosis", Expert Systems with Applications, 27 (2), 301–311.

Yang B.S., Han T., Hwang W.W., 2005, “Fault diagnosis of rotating machinery based on multi-class support vector machines”, Journal of Mechanical Science and Technology, 19 (3), 845–858.

Yang B.S., Hwang W.W., Kim D.J., Tan A.C.C., 2005, “Condition classification of small reciprocating compressor for refrigerators using artificial neural networks and support vector machines”, Mechanical System and Signal Processing, 19 (2), 371–390.

Yang B.S., Hwang W.W., Ko M.H., Lee S.J., 2005, “Cavitation detection of butterfly valve using support vector machines”, Journal of Sound Vibration, 287 (1–2), 25–43.

Yuan S.F., Chu F.L., 2006, “Support vector machines-based fault diagnosis for turbo-pump rotor”, Mechanical System and Signal Processing, 20 (4), 939–952.

Yuan S.F., Chu F.L., 2007, “Fault diagnosis based on support vector machine with parameter optimization by artificial immunization algorithm”, Mechanical System and Signal Processing, 21 (3), 1318–1330.

Zhang S., Asakura T., Xu X.L., and Xu B.J., 2003, “Fault diagnosis system for rotary machine based on fuzzy neural networks,” JSME International Journal, Mechanical Systems, Machine Elements and Manufacturing, Vol. 46, 1035-1041.

Zamanian A. H. and Ohadi A., 2010. “Gearbox fault detection through PSO Exact Wavelet Analysis and SVM Classifier”. 18th Annual International Conference on Mechanical Engineering-ISME 2010, Sharif University of Technology, Tehran, Iran.

APPENDIX A: User's Guide for Signal Processing Software

A.1 Processing the Raw Vibration Data

1. Copy and paste the data from the text files into different excel files, and name the files accordingly.
2. Make a new excel file containing the names of the excel files and the multiclass label (1, 2, 3, 4) assigned to each of the file. For example names.xlsx.

A.2 Training and Testing Data

1. MATLAB™ program cv.m or SVM_GA.m or runABC.m is used to select the training and testing data separately and entering it onto the SVM classifier using the optimizing technique GSM or GA or ABCA, respectively.
2. Open the functions trainingdata.m and testingdata.m to change the location to folder containing the data files. For example

   b=load(fullfile('F:','djb','d','bordoloi','PhD_work','myworkPhD','libsvm3.1_matlab',foldername,dataname,name));
3. In the main program cv.m or SVM_GA.m or runABC.m, enter the name of folder in the 'foldername' and the subfolder 'dataname' containing the *.mat files of the data obtained from the processing of the raw data.
4. Enter the name of files to use as the arguments of 'trainingdata.m' function, separated by comma. Similar settings hold for testing data too.
5. Results are evaluated based on fault prediction accuracies as well as individual fault prediction accuracies in training and testing. These results are shown in the command window.
6. The command screen data can be saved by appropriately in a tabular form.

A.3. Programs

Table A.1 illustrates the list of MATLAB™ computer program used for the generation of results.

Table A.1 List of computer codes used for the present work

Sl No.	Program name	Description
1.	trainingdata.m	Function to select the training data for SVM
2.	testingdata.m	Function to select the testing data for SVM
3.	cv.m	Program to run SVM software for multiclass classification on Gear defects using GSM
4.	SVM_GA.m	Program to run SVM software for multiclass classification on Gear defects using GA
5.	runABC.m	Program to run SVM software for multiclass classification on Gear defects using ABCA
6.	featureextraction.m	Program to extract statistical features from time domain data
7.	featureextraction_fft.m	Program to extract statistical features from frequency domain data
8.	wavelet_features.m	Program to extract statistical features using Mexican hat and Morlet wavelet family from time domain data
9.	wavelet_tree.m	Program to extract statistical features using wavelet packet from time domain data

APPENDIX B: Plots of Frequency and Time domain data

B.1 Calculation of the Gear Mesh Frequency

The motor spindle speed is reduced with the pulley and belt by 2.6 times (approximately) and number of teethes on the pinion is 18 and gear is 27 (illustrated in Figure 2.5). The gear mesh frequency is the equal to the teeth multiply by the rpm of the pinion. For example, the gear mesh frequency (GMF) for the 30 Hz rotational speed is $(30\times18)/2.6=207.69$ Hz and the 1x, 2x, 3x rotational speeds are 11.53 Hz, 23.08 Hz, and 34.62 Hz, respectively. Table B.1 represents the 1x, 2x, 3x and gear mesh frequencies for different rotational speeds.

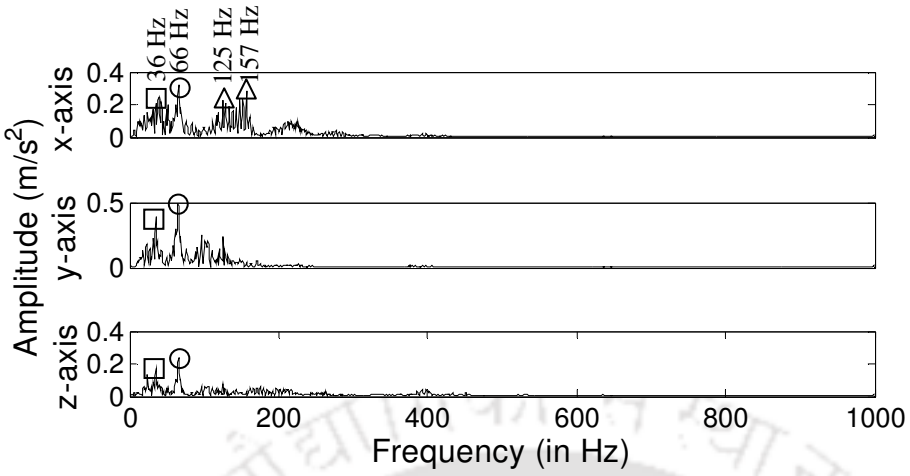
Table B.1 1x, 2x, 3x and gear mesh frequencies for different rotational speed

Spindle speed (A) in Hz	1x $(B=A/2.6)$ in Hz	2x $2B$ in Hz	3x $3B$ in Hz	4x $4B$ in Hz	5x $5B$ in Hz	GMF $18B$ in Hz
10	3.85	7.69	11.53	15.4	19.25	69.23
15	5.77	11.54	17.31	23.08	28.85	103.85
20	7.69	15.38	23.08	30.76	38.45	138.46
25	9.62	19.23	28.85	38.48	48.10	173.08
30	11.53	23.08	34.62	46.12	57.65	207.69

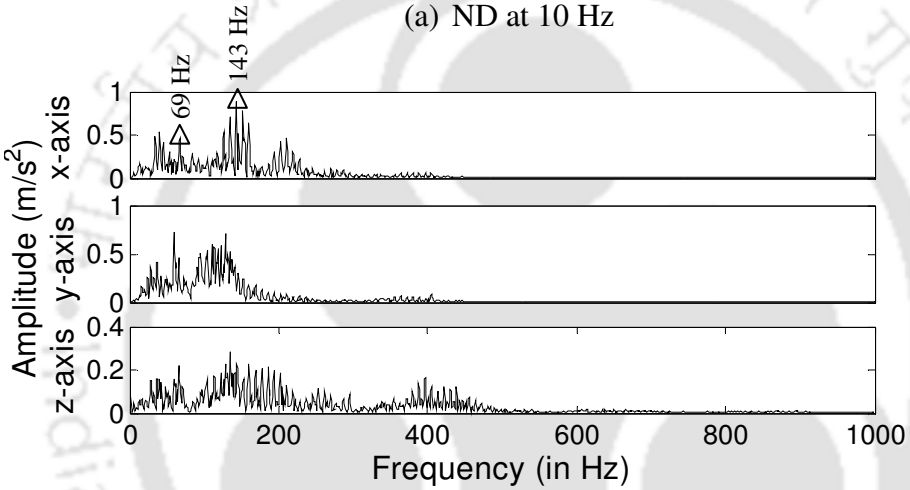
B.2 Plots for Frequency Domain and Time Domain Data

The ND, CT, MT and WT case frequency domain data for three orthogonal directions and 10 Hz, 15 Hz and 20 Hz rotational speed are plotted in Figure B.1 (a)-(c), Figure B.2 (a)-(c), Figure B.3 (a)-(c), Figure B.4 (a)-(c) respectively. In the similar way time domain data are plotted in Figure B.5 (a)-(c), Figure B.6 (a)-(c), Figure B.7 (a)-(c) and Figure B.8 (a)-(c) respectively. The peak of frequency are marked on Figures and it is indicated in Table B.2

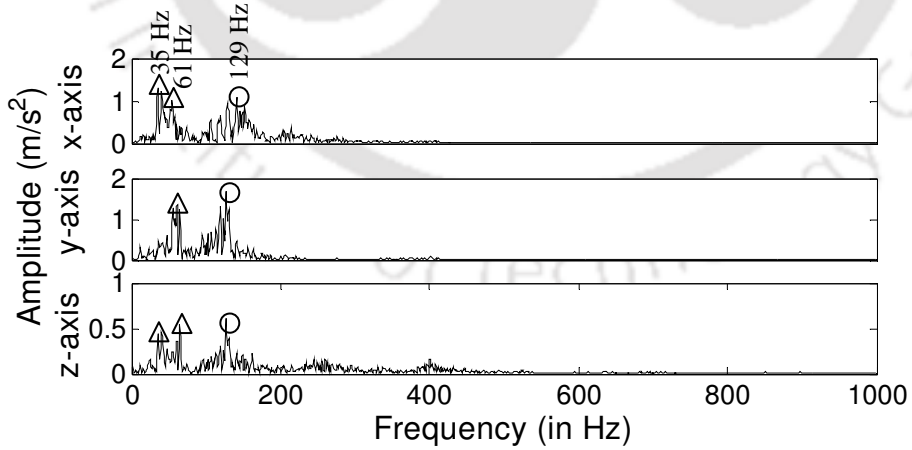
○ GMF □ nX component of rotational frequency △ Frequency of structural mode



(a) ND at 10 Hz

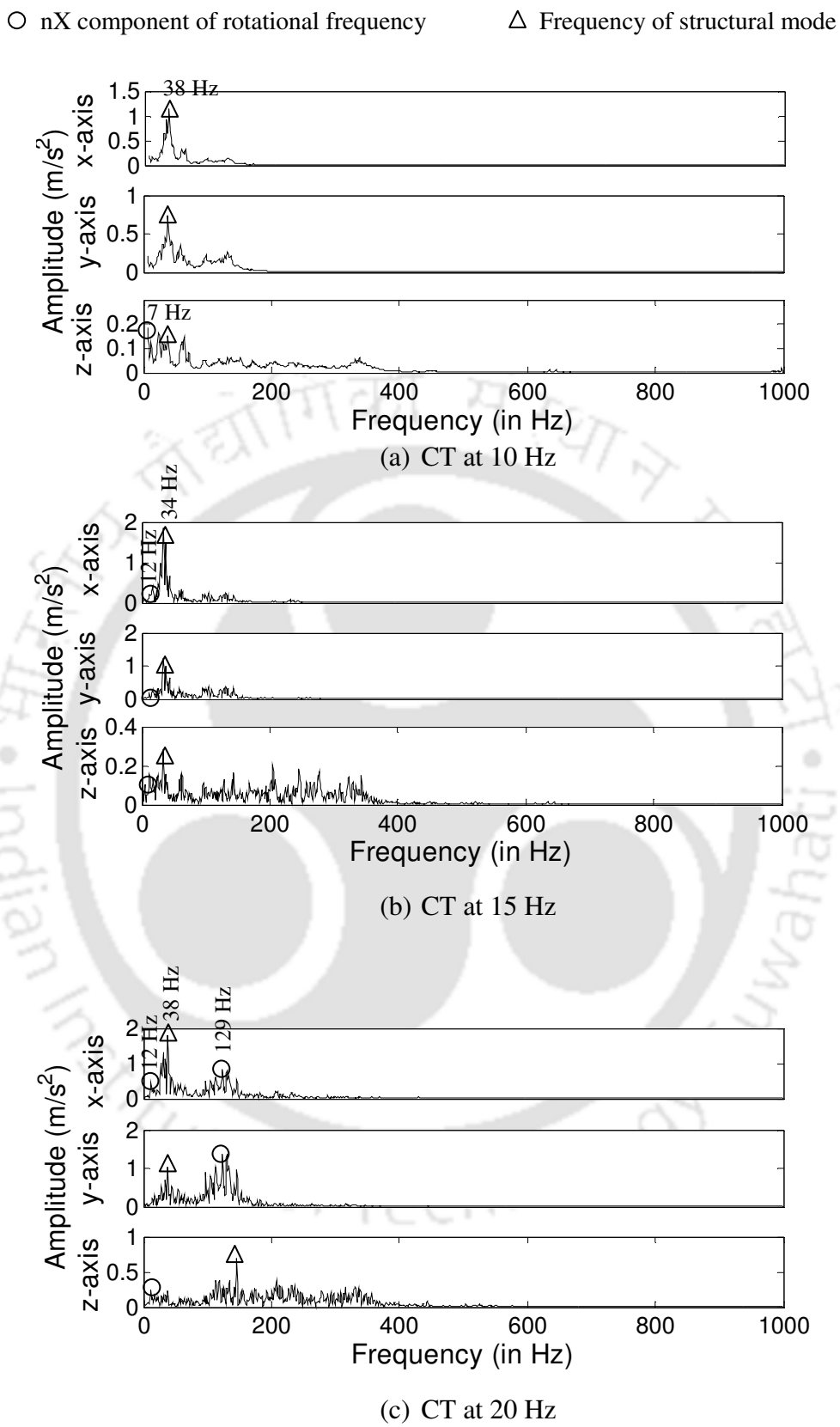


(b) ND at 15 Hz



(c) ND at 20 Hz

Figure B.1 Response in three orthogonal directions in frequency domain at (a) 10 Hz, (b) 15 Hz and (c) 20 Hz rotational speed for ND



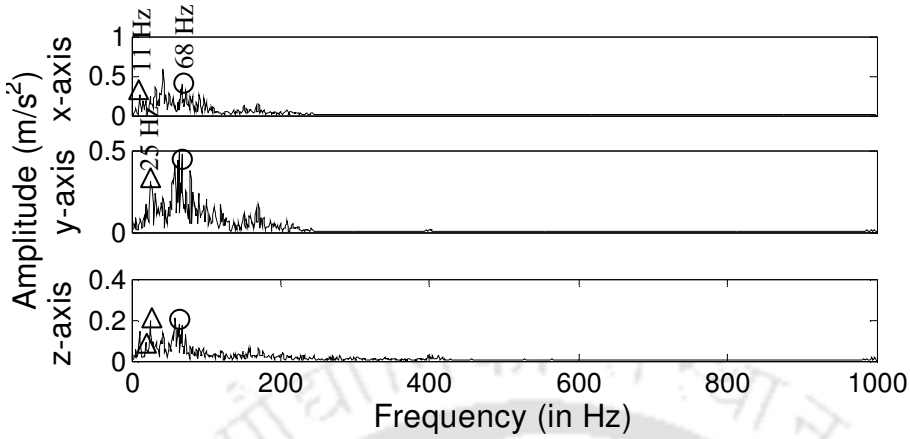
(a) CT at 10 Hz

(b) CT at 15 Hz

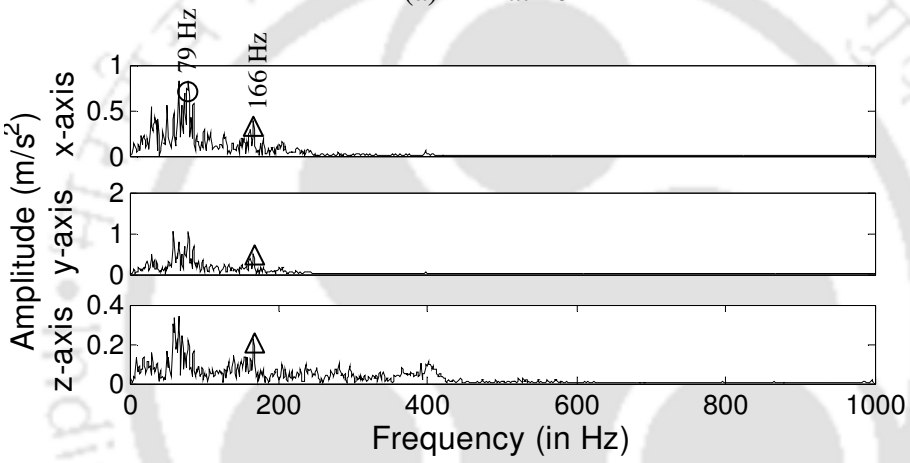
(c) CT at 20 Hz

Figure B.2 Response in three orthogonal directions in frequency domain at (a) 10 Hz, (b) 15 Hz and (c) 20 Hz rotational speed for CT

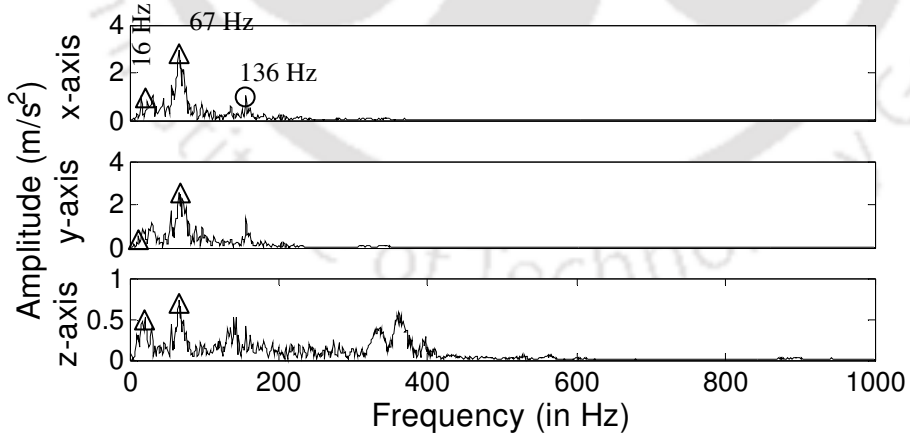
○ nX component of rotational frequency △ Frequency of structural mode



(a) MT at 10 Hz



(b) MT at 15 Hz



(c) MT at 20 Hz

Figure B.3 Response in three orthogonal directions in frequency domain at (a) 10 Hz, (b) 15 Hz and (c) 20 Hz rotational speed for MT

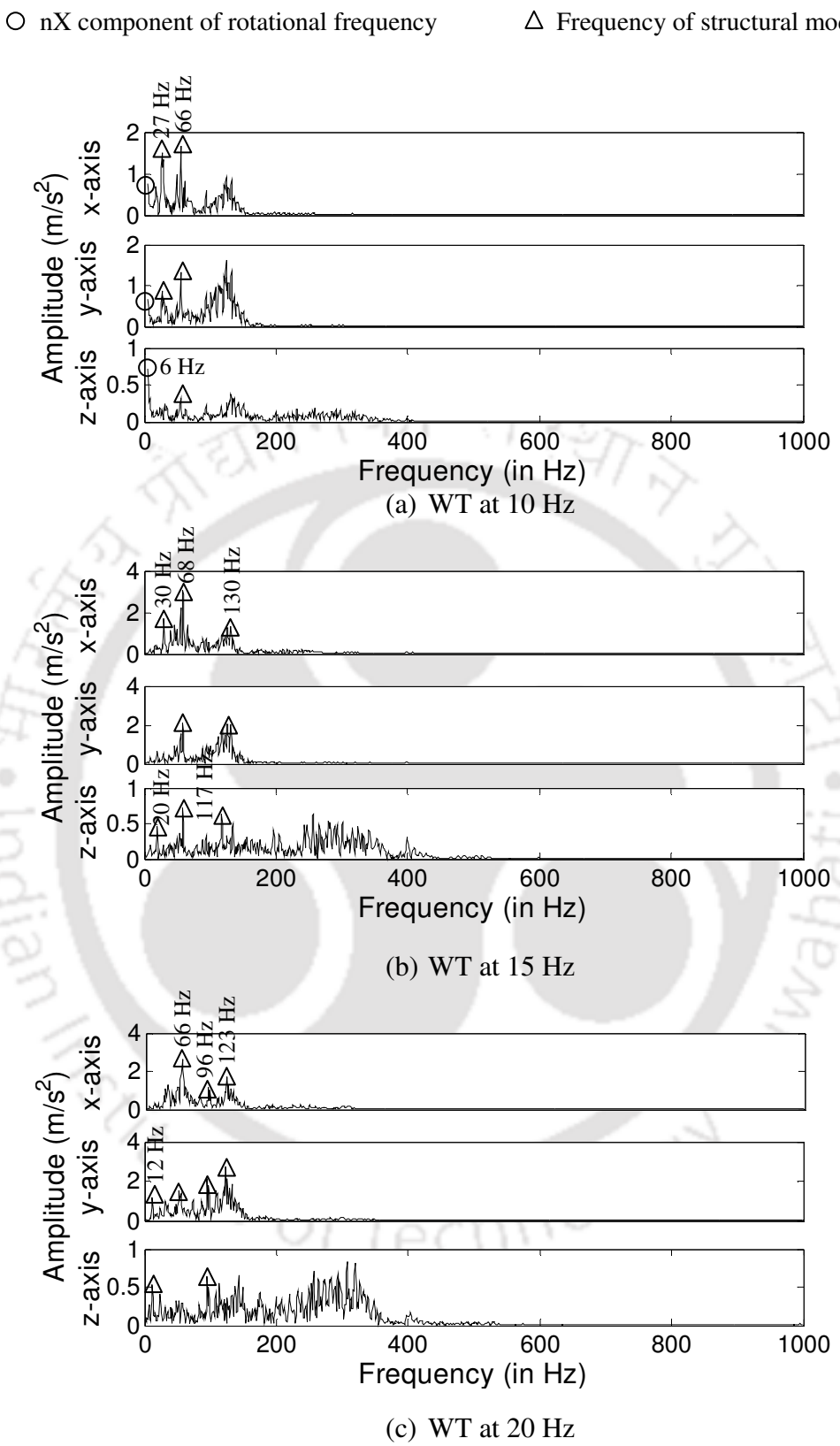
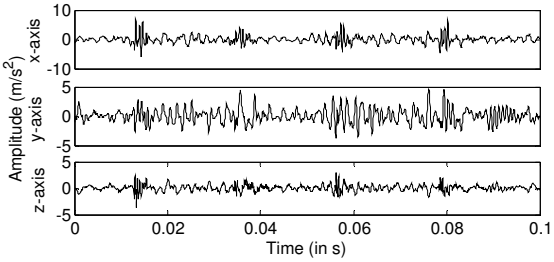
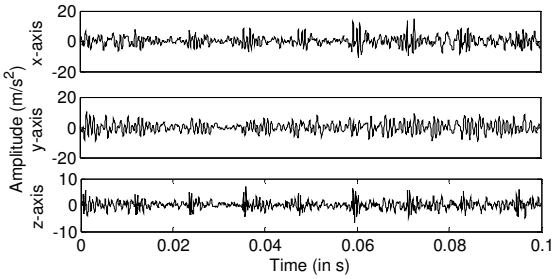


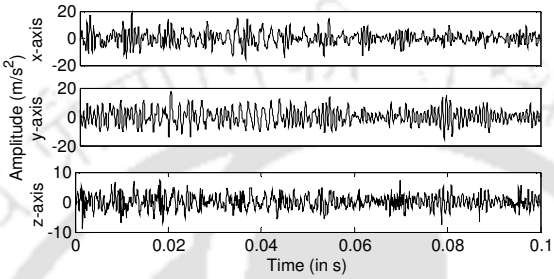
Figure B.4 Response in three orthogonal directions in frequency domain at (a) 10 Hz, (b) 15 Hz and (c) 20 Hz rotational speed for WT



(a) ND at 10 Hz

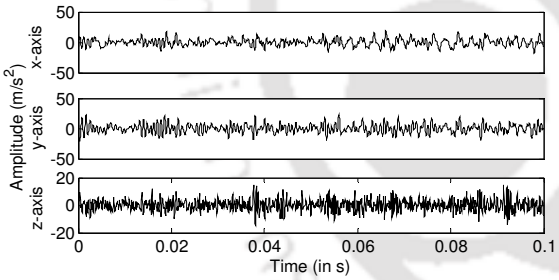


(b) ND at 15 Hz

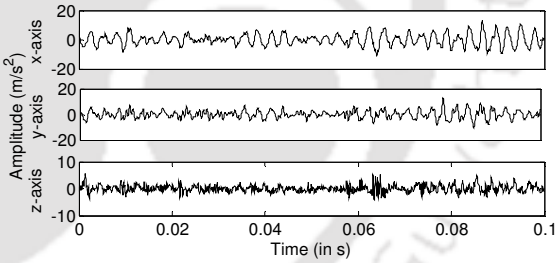


(c) ND at 20 Hz

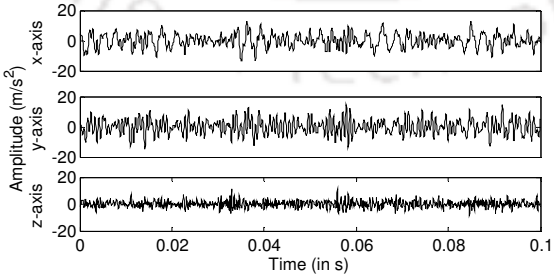
Figure B.5 Response in three orthogonal directions in time domain at (a) 10 Hz, (b) 15 Hz and (c) 20 Hz rotational speed for ND



(a) CT at 10 Hz

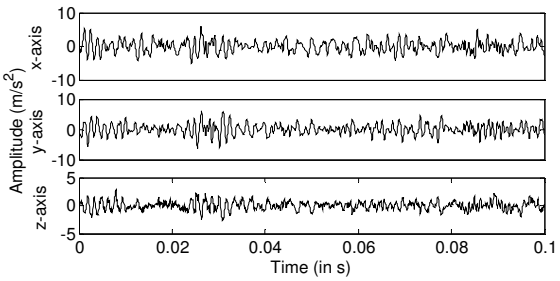


(b) CT at 15 Hz

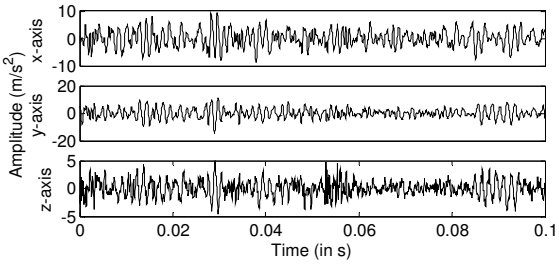


(c) CT at 20 Hz

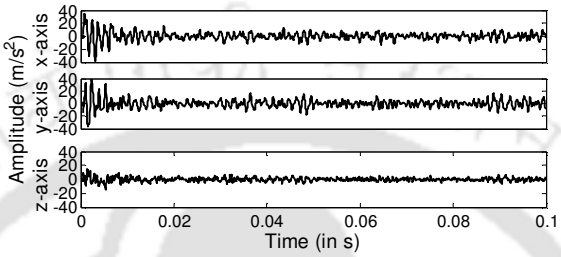
Figure B.6 Response in three orthogonal directions in time domain at (a) 10 Hz, (b) 15 Hz and (c) 20 Hz rotational speed for CT



(a) MT at 10 Hz

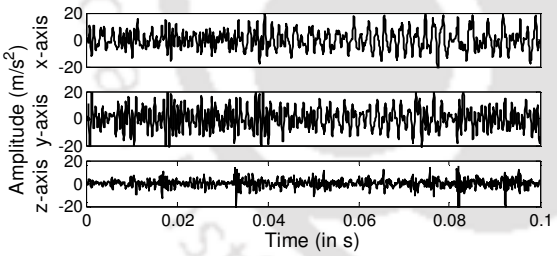


(b) MT at 15 Hz

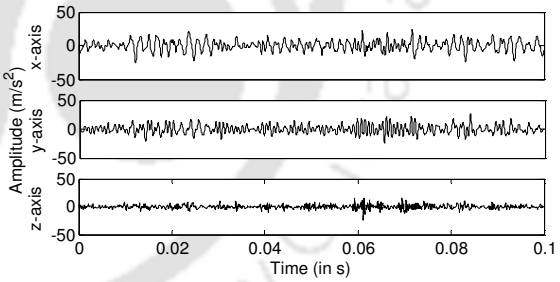


(c) MT at 20 Hz

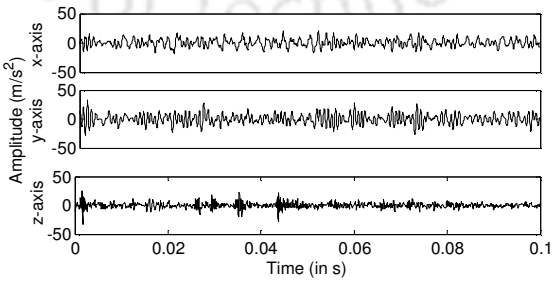
Figure B.7 Response in three orthogonal directions in time domain at (a) 10 Hz, (b) 15 Hz and (c) 20 Hz rotational speed for MT



(a) WT at 15 Hz



(b) WT at 15 Hz



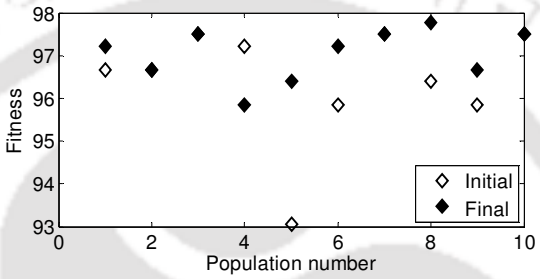
(c) WT at 20 Hz

Figure B.8 Response in three orthogonal directions in time domain at (a) 10 Hz, (b) 15 Hz and (c) 20 Hz rotational speed for WT

APPENDIX C: Effect of Populations in GA and ABCA

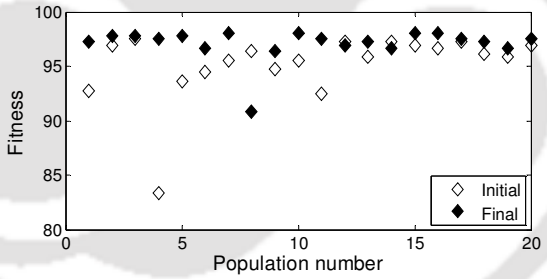
C.1 Variation of Fitness with Different Population for GA

The fitness values are plotted (in Figure C.1 (a)-(g)) for the different population (10 to 70 in the step of 10) in GA. In the case of ABCA the populations are taken 10 to 110 in the step of 10 and shown in Figure C.2 (a)-(k). In Figures *C*-SVC conditions are considered. Time feature for the same speed at 10 Hz rotational speed are used.



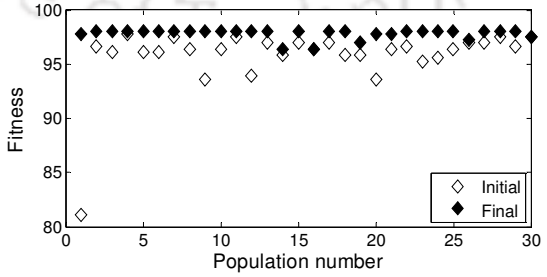
(a) 10 Hz with *C*-SVC and 10 population

(one out of 10 is conversed (10%) to 97.78% fitness value)



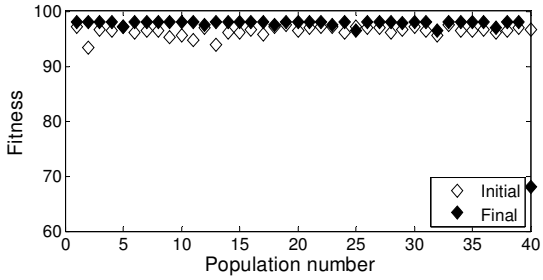
(b) 10 Hz with *C*-SVC and 20 population

(four out of 20 is conversed (20%) to 97.50% fitness value)



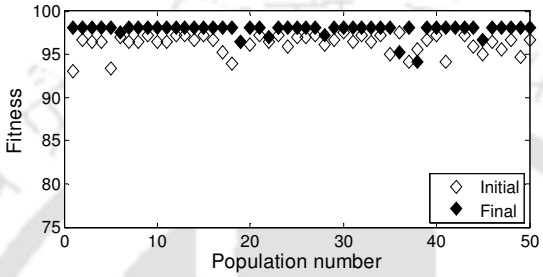
(c) 10 Hz with *C*-SVC and 30 population

(22 out of 30 is conversed (73.33%) to 98.06 fitness value)



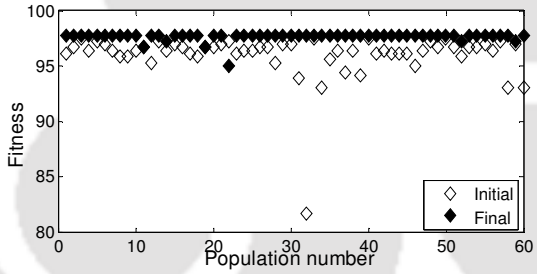
(d) 10 Hz with *C*-SVC and 40 population

(31 out of 40 is conversed (77.50%) to 98.06 fitness value)



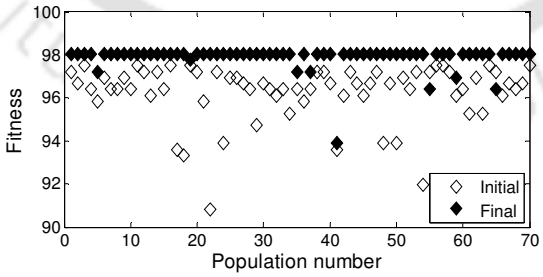
(e) 10 Hz with *C*-SVC and 50 population

(43 out of 50 is conversed (86.00%) to 98.06 fitness value)



(f) 10 Hz with *C*-SVC and 60 population

(57 out of 60 is conversed (95.00%) to 97.22 fitness value)

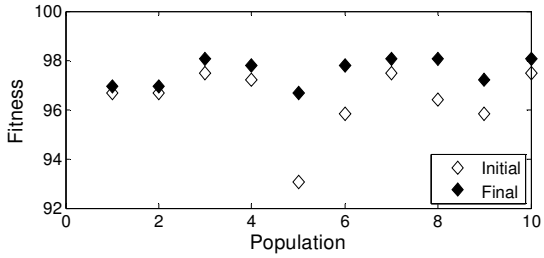


(g) 10 Hz with *C*-SVC and 70 population

(62 out of 70 is conversed (88.57%) to 98.06 fitness value)

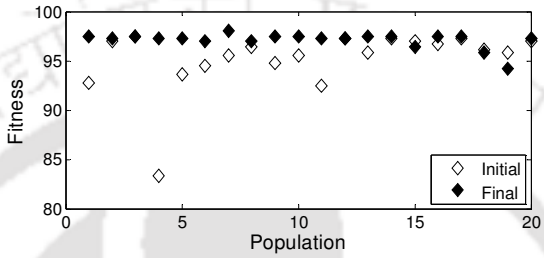
Figure C.1 Variation of the fitness values with different population for GA

In Figure C.1, it is clear that the fitness value 98.06 is achieved by the 22 population among the 30 population.



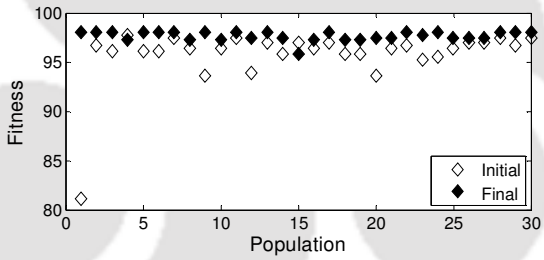
(a) 10 Hz with C-SVC and 10 population

(4 out of 10 is conversed (40%) to 98.06 fitness value)



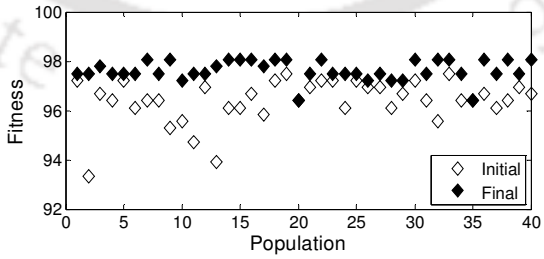
(b) 10 Hz with C-SVC and 20 population

(1 out of 20 is conversed (5%) to 98.06 fitness value)



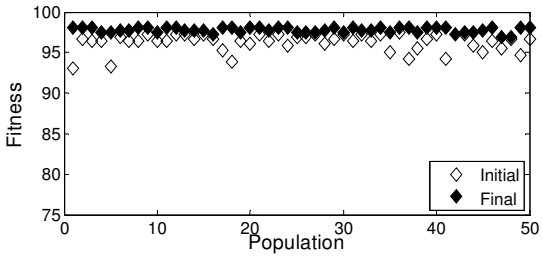
(c) 10 Hz with C-SVC and 30 population

(15 out of 30 is conversed (50%) to 98.06 fitness value)



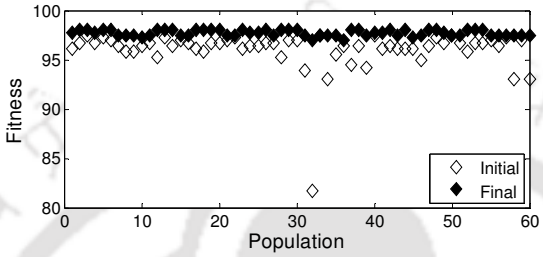
(d) 10 Hz with C-SVC and 40 population

(14 out of 40 is conversed (35.00%) to 98.06 fitness value)



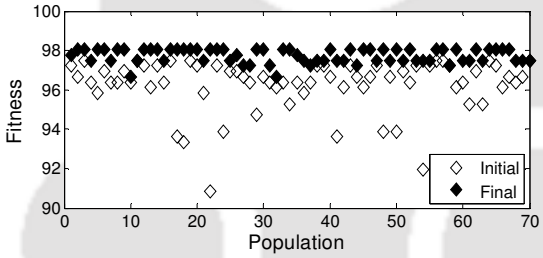
(e) 10 Hz with C-SVC and 50 population

(24 out of 50 is conversed (48.00%) to 98.06 fitness value)



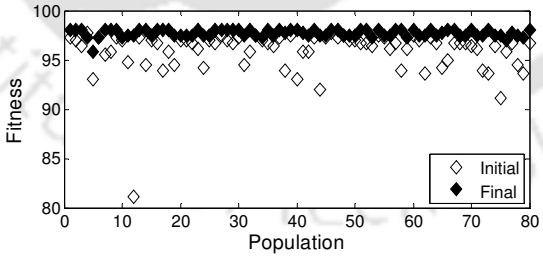
(f) 10 Hz with C-SVC and 60 population

(25 out of 60 is conversed (41.67%) to 98.06 fitness value)



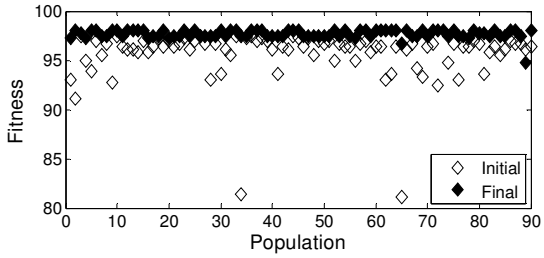
(g) 10 Hz with C-SVC and 70 population

(36 out of 70 is conversed (51.43%) to 98.06 fitness value)



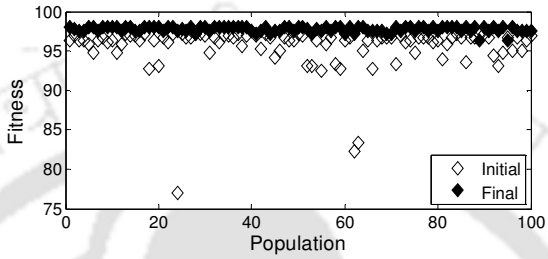
(h) 10 Hz with C-SVC and 80 population

(37 out of 80 is conversed (46.25%) to 98.06 fitness value)



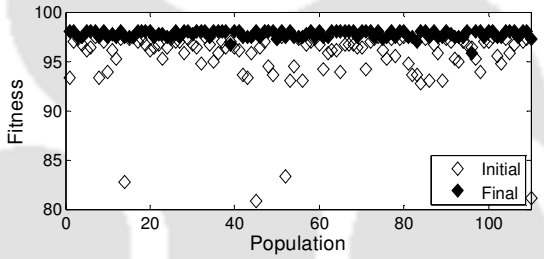
(i) 10 Hz with C-SVC and 90 population

(41 out of 90 is conversed (45.55%) to 98.06 fitness value)



(j) 10 Hz with C-SVC and 100 population

(53 out of 100 is conversed (53.00%) to 98.06 fitness value)



(k) 10 Hz with C-SVC and 110 population

(58 out of 110 is conversed (52.72%) to 98.06 fitness value)

Figure C.2 Variation of the fitness values with different population for ABCA

In Figure C.1, it is clear that the fitness value of 98.06 is achieved by 22 populations among 30 populations. In Figure C.2, it is clear that the fitness value of 98.06 is achieved by 4 populations among 10 populations. It is clear from Figures C.2 ((a) – (k)) that among chosen population sizes (10 to 110), the convergence of more than 50% population size is achieved when the population size of 100 is chosen (Figure C.2. (j)).

APPENDIX D: Specification of PC, MATLAB™ and Tabulation of Optimized Parameters with Computation Time

D.1 Specification of PC and MATLAB™ Version

PC specification: Intel core i3, S30 processor, Intel H57 chipset, 4 GB DDR3 RAM.

MATLAB™ specification: Version 7.6.0.324 (R2008a), February 10, 2008.

D.2 Tabulation of Optimal Parameters with the Computation Time for Optimization and Classification

Table D.1 Optimized parameters with computation time using time domain features (for the same speed)

Speed (Hz)	Optimization method	*C*-SVC classifier			ν-SVC		
		C	γ	Time (sec.)	ν	γ	Time (sec.)
For same speed							
10	GSM	1.41421	0.17678	429.41	0.04419	0.02210	159.53
	GA	0.92253	0.19546	89.45	0.03242	0.01964	109.13
	ABCA	1.49996	0.46600	765.21	0.03628	0.02485	824.87
15	GSM	1.41421	0.25000	400.82	0.17678	0.00138	163.95
	GA	1.48250	0.27835	115.11	0.09795	0.02472	134.25
	ABCA	1.50000	0.27145	858.58	0.09746	0.03167	952.76
20	GSM	1.41421	0.12500	379.84	0.08839	0.03125	175.74
	GA	0.43415	0.05789	106.21	0.08395	0.13742	139.07
	ABCA	1.37000	0.10845	832.46	0.15300	0.01177	1024.73
25	GSM	0.70711	0.03125	371.53	0.01105	0.02210	196.69
	GA	0.53318	0.02049	101.99	0.09554	0.04750	142.28
	ABCA	0.25620	0.06479	901.70	0.13693	0.05752	1002.71
30	GSM	0.12500	0.01105	354.27	0.00098	0.00098	201.80
	GA	0.92253	0.03431	111.51	0.03704	0.01964	129.72
	ABCA	0.25620	0.06479	856.16	0.17080	0.01920	949.11

Table D.2 Optimized parameters with computation time using time domain features (for the interpolation and extrapolation speed)

Speed (Hz)	Optimization method	*C*-SVC classifier			ν-SVC		
		C	γ	Time (sec.)	ν	γ	Time (sec.)
For interpolation speed (range 5 Hz)							
12.5	GSM	1.41421	0.35355	1676.24	0.06250	0.06250	686.38
	GA	1.48250	0.09860	155.36	0.05213	0.15960	160.37
	ABCA	1.50000	0.09767	1071.42	0.02920	0.13096	1155.69
17.5	GSM	1.41421	0.25000	1569.41	0.08839	0.08839	694.53
	GA	1.31361	0.02472	177.04	0.11176	0.01776	174.20
	ABCA	0.62134	0.04302	1211.58	0.11535	0.03048	1335.21
22.5	GSM	1.41421	0.12500	1454.42	0.04419	0.04419	712.73
	GA	0.92253	0.05789	156.09	0.06478	0.08774	177.17
	ABCA	0.71057	0.16583	1161.64	0.08006	0.12859	1527.93
27.5	GSM	1.41421	0.06250	1343.97	0.01105	0.03125	774.60
	GA	1.22977	0.01921	128.46	0.05171	0.01964	168.08
	ABCA	1.29666	0.03363	1159.58	0.03359	0.03694	1453.51
For interpolation speed (range 10 Hz)							
15	GSM	1.41421	0.25000	1711.37	0.04419	0.12500	745.19
	GA	1.21405	0.13800	164.15	0.06478	0.19033	173.96
	ABCA	1.30049	0.13948	1107.59	0.10953	0.06370	1329.51
20	GSM	1.41421	0.06250	1594.73	0.06250	0.06250	722.60
	GA	0.85381	0.05789	160.32	0.01092	0.14633	204.15
	ABCA	0.94651	0.05064	1134.35	0.01079	0.14626	1471.35
25	GSM	1.41421	0.12500	1525.94	0.04419	0.03125	748.67
	GA	0.92253	0.05789	167.32	0.06741	0.01964	173.19
	ABCA	1.39685	0.07651	1229.68	0.04330	0.01937	1388.97
For extrapolation speed (range 5 Hz)							
15	GSM	1.41421	0.12500	1653.48	0.04419	0.08839	631.66
	GA	0.92253	0.13522	138.58	0.06478	0.14092	150.17
	ABCA	1.00422	0.14220	1000.69	0.04641	0.12200	1136.37
25	GSM	1.41421	0.12500	1397.79	0.04419	0.06250	689.27
	GA	0.96216	0.18980	187.92	0.06728	0.10039	191.57
	ABCA	0.78353	0.15242	1139.68	0.06707	0.17116	1298.05
30	GSM	1.00000	0.08839	1316.99	0.00098	0.03125	765.23
	GA	0.92253	0.19420	172.52	0.06478	0.16766	198.48
	ABCA	1.03560	0.17701	1310.10	0.14555	0.00001	1591.73
For extrapolation speed (range 10 Hz)							
20	GSM	1.41421	0.35355	1687.76	0.06250	0.06250	692.58
	GA	1.48250	0.09860	145.93	0.05213	0.15960	159.43
	ABCA	1.50000	0.09751	1035.59	0.02683	0.12825	1105.11
25	GSM	1.41421	0.25000	1545.06	0.08839	0.08839	683.77
	GA	1.31361	0.02472	170.76	0.11176	0.01776	172.64
	ABCA	1.32237	0.02415	1157.55	0.13127	0.03792	1275.58
30	GSM	1.41421	0.12500	1469.07	0.04419	0.04419	723.43
	GA	0.92253	0.05789	150.89	0.06478	0.08774	178.63
	ABCA	0.61127	0.12905	1157.40	0.07941	0.12425	1463.79

Table D.3 Optimized parameters with computation time using frequency domain features (for the same speed)

Speed (Hz)	Optimization method	*C*-SVC classifier			ν-SVC		
		C	γ	Time (sec.)	ν	γ	Time (sec.)
For same speed							
10	GSM	1.41421	0.00195	384.74	0.00276	0.00098	103.30
	GA	0.88932	0.24543	181.73	0.01869	0.30462	184.22
	ABCA	1.26581	0.28442	1214.34	0.11200	0.37042	1301.50
15	GSM	1.00000	0.01563	314.02	0.00781	0.00276	147.33
	GA	1.29002	0.01850	60.41	0.08361	0.01964	108.71
	ABCA	1.29462	0.01851	620.53	0.02648	0.00415	870.75
20	GSM	1.41421	0.70711	352.82	0.00552	0.01105	160.35
	GA	1.14687	0.03431	86.47	0.01176	0.01776	110.86
	ABCA	0.72685	0.03691	632.00	0.01924	0.00211	729.73
25	GSM	1.41421	0.06250	337.33	0.02210	0.01563	148.34
	GA	1.28850	0.23869	108.71	0.01176	0.08774	115.58
	ABCA	0.96123	0.29873	831.58	0.02213	0.06250	784.82
30	GSM	1.41421	0.12500	355.93	0.00781	0.08839	155.12
	GA	1.48250	0.15695	93.86	0.01436	0.01104	115.46
	ABCA	1.50000	0.15403	622.16	0.14608	0.00257	709.40

Table D.4 Optimized parameters with computation time using frequency domain features (for the interpolation and extrapolation speed)

Speed (Hz)	Optimization method	*C*-SVC classifier			ν-SVC		
		C	γ	Time (sec.)	ν	γ	Time (sec.)
For interpolation speed (range 5 Hz)							
12.5	GSM	1.41421	0.01105	409.01	0.01105	0.00552	224.72
	GA	0.92253	0.18965	167.18	0.08800	0.30462	196.64
	ABCA	0.25620	0.06479	1023.88	0.19089	0.25751	1173.25
17.5	GSM	1.41421	0.04419	370.60	0.00781	0.01105	170.93
	GA	1.36838	0.18367	124.60	0.04297	0.01104	126.20
	ABCA	1.50000	0.39238	900.78	0.04011	0.00108	919.54
22.5	GSM	1.41421	0.17678	361.46	0.03125	0.00391	169.51
	GA	1.47537	0.40907	137.90	0.02317	0.01171	113.64
	ABCA	1.50000	0.37838	782.37	0.08184	0.00110	764.65
27.5	GSM	1.41421	0.12500	368.23	0.01105	0.01563	169.88
	GA	1.38301	0.71053	155.36	0.08408	0.01415	130.31
	ABCA	1.45627	0.00316	1052.23	0.04244	0.00987	886.56
For interpolation speed (range 10 Hz)							
15	GSM	1.41421	0.01105	457.56	0.00552	0.01105	233.99
	GA	0.92253	0.19769	167.55	0.01176	0.01404	141.57
	ABCA	1.15193	0.13890	998.59	0.04239	0.00449	1007.51
20	GSM	1.41421	0.12500	364.75	0.01563	0.01105	167.37
	GA	1.10283	0.54974	161.60	0.17414	0.46728	184.60
	ABCA	1.07383	0.52260	1041.09	0.02719	0.49377	1128.41
25	GSM	1.41421	0.06250	378.62	0.01563	0.00781	172.09
	GA	1.48250	0.11301	103.78	0.02270	0.01318	111.55
	ABCA	1.50000	0.09935	947.10	0.05026	0.00193	1031.20
For extrapolation speed (range 5 Hz)							
15	GSM	1.41421	0.00276	416.72	0.01105	0.00195	227.28
	GA	0.52075	0.18965	188.07	0.03736	0.01512	218.40
	ABCA	0.25620	0.06479	1218.65	0.23190	0.13653	1319.69
25	GSM	1.41421	0.08839	359.36	0.00781	0.01105	158.67
	GA	1.48250	0.05145	138.66	0.03524	0.01964	119.74
	ABCA	1.26058	0.06468	956.53	0.08109	0.00434	907.31
30	GSM	1.41421	0.12500	373.41	0.02210	0.00391	174.74
	GA	1.32021	0.93942	247.25	0.00388	0.91063	228.82
	ABCA	1.49144	0.98834	1613.90	0.00079	0.84512	1484.49
For extrapolation speed (range 10 Hz)							
20	GSM	1.41421	0.01105	461.66	0.01105	0.00552	218.39
	GA	0.92253	0.18965	172.31	0.08800	0.30462	191.29
	ABCA	0.25620	0.06479	1046.30	0.19089	0.25751	1197.52
25	GSM	1.41421	0.04419	380.32	0.00781	0.01105	168.60
	GA	1.48250	0.18965	126.21	0.03408	0.01415	128.33
	ABCA	1.50000	0.39238	921.18	0.04011	0.00108	926.96
30	GSM	1.41421	0.17678	367.75	0.03125	0.00391	169.71
	GA	1.47537	0.40907	137.89	0.02892	0.01318	121.19
	ABCA	1.50000	0.37838	814.07	0.08184	0.00110	787.39

Table D.5 Optimized parameters with computation time using Morlet wavelet family features (for the same speed)

Speed (Hz)	Optimization method	*C*-SVC classifier			ν -SVC		
		C	γ	Time (sec.)	ν	γ	Time (sec.)
For same speed							
10	GSM	1.41421	0.03125	388.99	0.01105	0.01105	174.93
	GA	1.22696	0.04553	106.67	0.00580	0.04011	122.05
	ABCA	1.50000	0.05289	750.12	0.00600	0.00993	791.30
15	GSM	1.00000	0.04419	350.19	0.00098	0.00276	177.52
	GA	0.92253	0.03764	112.81	0.00504	0.03710	125.78
	ABCA	1.29462	0.03143	739.90	0.04408	0.01502	896.93
20	GSM	1.00000	0.02210	358.05	0.00195	0.00276	200.65
	GA	0.92253	0.01464	117.75	0.00580	0.01964	135.52
	ABCA	0.25620	0.06479	874.18	0.03331	0.01227	918.61
25	GSM	0.25000	0.02210	376.65	0.01563	0.00391	234.79
	GA	0.83692	0.01527	129.27	0.01176	0.01964	152.61
	ABCA	0.01851	1.14055	1059.23	0.04976	0.04449	1097.98
30	GSM	1.41421	0.01563	409.91	0.00098	0.00276	268.81
	GA	0.83692	0.01850	143.76	0.05603	0.01697	168.45
	ABCA	0.31885	0.01965	1165.60	0.03878	0.01177	1245.63

Table D.6 Optimized parameters with computation time using Morlet wavelet family features (for the interpolation and extrapolation speed)

Speed (Hz)	Optimization method	*C*-SVC classifier			ν -SVC		
		C	γ	Time (sec.)	ν	γ	Time (sec.)
For interpolation speed (range 5 Hz)							
12.5	GSM	1.41421	0.03125	738.29	0.12500	0.00781	369.09
	GA	1.43418	0.05789	134.34	0.02908	0.01964	136.59
	ABCA	1.50000	0.06596	945.14	0.08613	0.03798	1193.05
17.5	GSM	1.41421	0.03125	689.00	0.01105	0.00098	384.14
	GA	0.83692	0.01464	132.61	0.05950	0.01964	146.05
	ABCA	1.48512	0.02596	972.39	0.09116	0.00001	1063.82
22.5	GSM	1.00000	0.02210	682.39	0.02210	0.00781	401.01
	GA	0.52465	0.01850	133.17	0.01992	0.01964	149.51
	ABCA	1.50000	0.02588	1105.69	0.01502	0.00733	1064.33
27.5	GSM	0.50000	0.00781	678.68	0.00781	0.00781	464.48
	GA	0.57186	0.00986	140.06	0.00580	0.01318	167.14
	ABCA	1.50000	0.01295	1115.03	0.02213	0.01250	1061.86
For interpolation speed (range 10 Hz)							
15	GSM	1.41421	0.03125	757.01	0.02210	0.00781	402.97
	GA	1.34280	0.03395	139.04	0.10336	0.01104	160.08
	ABCA	1.50000	0.03591	983.88	0.06533	0.00496	1104.14
20	GSM	1.41421	0.03125	718.92	0.17678	0.00276	440.32
	GA	1.29002	0.02224	143.64	0.34977	0.08774	202.09
	ABCA	1.21875	0.03746	1159.61	0.21067	0.01965	1341.05
25	GSM	1.41421	0.02210	743.37	0.12500	0.00195	453.73
	GA	0.92253	0.00986	145.85	0.15096	0.00986	176.41
	ABCA	0.97509	0.00416	1193.71	0.15312	0.01005	1350.86
For extrapolation speed (range 5 Hz)							
15	GSM	1.41421	0.03125	688.59	0.06250	0.00781	321.68
	GA	0.92253	0.05789	124.63	0.08408	0.01104	135.02
	ABCA	0.95717	0.06715	800.22	0.08756	0.00852	904.03
25	GSM	1.41421	0.06250	662.70	0.00552	0.01563	377.52
	GA	0.83692	0.01850	127.69	0.05799	0.01964	149.39
	ABCA	0.35898	0.01177	1077.93	0.11072	0.00001	1050.11
30	GSM	1.41421	0.02210	664.78	0.00138	0.02210	420.48
	GA	0.92253	0.01286	134.26	0.00580	0.01415	154.09
	ABCA	0.83730	0.01477	1060.06	0.00843	0.01239	1056.58
For extrapolation speed (range 10 Hz)							
20	GSM	1.41421	0.03125	708.74	0.12500	0.00781	355.91
	GA	1.43418	0.05789	131.10	0.02908	0.01964	135.20
	ABCA	1.50000	0.06596	940.24	0.08613	0.03798	1180.37
25	GSM	1.41421	0.03125	670.27	0.01105	0.00098	374.46
	GA	0.83692	0.01464	130.67	0.05950	0.01964	140.26
	ABCA	1.48512	0.02596	982.60	0.09116	0.00001	1048.84
30	GSM	1.00000	0.02210	665.97	0.02210	0.00781	397.24
	GA	0.52465	0.01850	128.05	0.01992	0.01964	146.12
	ABCA	1.50000	0.02588	1109.00	0.01502	0.00733	1071.57

Table D.7 Optimized parameters with computation time using Mexican hat wavelet family features (for the same speed)

Speed (Hz)	Optimization method	*C*-SVC classifier			ν -SVC		
		C	γ	Time (sec.)	ν	γ	Time (sec.)
For same speed							
10	GSM	1.41421	0.35355	402.90	0.17678	0.01563	178.84
	GA	0.94389	0.17377	126.04	0.25996	0.29437	138.35
	ABCA	1.23720	0.02409	889.84	0.13837	0.00995	1055.66
15	GSM	1.41421	0.06250	374.16	0.06250	0.01563	170.35
	GA	0.92253	0.05789	108.09	0.06478	0.06810	136.86
	ABCA	0.62307	0.04702	760.36	0.15300	0.01177	942.43
20	GSM	1.00000	0.03125	359.97	0.02210	0.03125	188.05
	GA	0.92253	0.01286	102.95	0.07831	0.04829	151.09
	ABCA	0.31885	0.01965	821.49	0.17080	0.01920	996.80
25	GSM	1.00000	0.04419	383.32	0.08839	0.03125	232.30
	GA	0.92253	0.05789	151.92	0.07019	0.04595	167.48
	ABCA	1.11685	0.06378	1088.44	0.10323	0.05790	1126.97
30	GSM	0.70711	0.04419	406.28	0.06250	0.03125	255.37
	GA	1.34019	0.04641	162.56	0.01568	0.05816	172.96
	ABCA	1.34267	0.01965	1167.23	0.01502	0.05601	1243.48

Table D.8 Optimized parameters with computation time using Mexican hat wavelet family features (for the interpolation and extrapolation speed)

Speed (Hz)	Optimization method	*C*-SVC classifier			ν -SVC		
		C	γ	Time (sec.)	ν	γ	Time (sec.)
For interpolation speed (range 5 Hz)							
12.5	GSM	1.00000	0.12500	424.82	0.06250	0.03125	215.41
	GA	1.46962	0.08774	144.86	0.06728	0.04553	156.11
	ABCA	1.50000	0.09033	1105.26	0.06863	0.03674	1073.50
17.5	GSM	1.41421	0.12500	399.01	0.12500	0.02210	214.93
	GA	0.92253	0.05789	133.70	0.01176	0.02472	139.90
	ABCA	0.71472	0.05007	1109.27	0.26912	0.05199	1234.46
22.5	GSM	1.41421	0.04419	387.48	0.02210	0.01563	231.02
	GA	1.27568	0.09013	150.20	0.04844	0.04421	165.07
	ABCA	1.30830	0.08902	1229.74	0.09353	0.13637	1205.76
27.5	GSM	1.41421	0.04419	394.24	0.04419	0.01563	260.34
	GA	1.31404	0.01255	134.22	0.05366	0.01340	165.95
	ABCA	1.46241	0.01242	1170.91	0.05414	0.01277	1168.28
For interpolation speed (range 10 Hz)							
15	GSM	1.41421	0.06250	434.47	0.06250	0.02210	223.61
	GA	0.72770	0.05789	149.79	0.08408	0.01964	149.94
	ABCA	1.50000	0.04219	1142.15	0.08052	0.00449	1119.92
20	GSM	1.41421	0.08839	428.59	0.02210	0.03125	239.27
	GA	1.34047	0.04634	161.48	0.04939	0.01964	157.64
	ABCA	1.50000	0.04184	1217.36	0.00676	0.02818	1402.07
25	GSM	1.41421	0.06250	433.32	0.08839	0.01563	262.25
	GA	0.70086	0.01850	143.28	0.04270	0.01964	163.67
	ABCA	1.15974	0.03925	1191.63	0.07978	0.01177	1344.88
For extrapolation speed (range 5 Hz)							
15	GSM	1.41421	0.12500	416.47	0.12500	0.04419	191.56
	GA	1.45689	0.09781	125.42	0.13401	0.01120	170.45
	ABCA	1.50000	0.10378	914.54	0.19717	0.00625	1054.07
25	GSM	1.41421	0.08839	371.97	0.03125	0.02210	211.18
	GA	1.09067	0.05789	134.29	0.10356	0.10850	179.45
	ABCA	0.95490	0.06715	1061.54	0.09696	0.00449	1161.63
30	GSM	1.00000	0.03125	394.19	0.12500	0.00138	254.47
	GA	1.05227	0.07632	172.96	0.17193	0.07290	183.10
	ABCA	1.43722	0.02922	1233.32	0.04859	0.00944	1248.90
For extrapolation speed (range 10 Hz)							
20	GSM	1.00000	0.12500	432.86	0.06250	0.03125	214.78
	GA	1.46962	0.08774	139.20	0.06728	0.04553	159.50
	ABCA	1.50000	0.09033	1099.60	0.06863	0.03674	1105.70
25	GSM	1.41421	0.12500	411.65	0.12500	0.02210	214.60
	GA	0.92253	0.05789	133.10	0.01176	0.02472	141.34
	ABCA	0.71472	0.05007	1106.71	0.26912	0.05199	1289.49
30	GSM	1.41421	0.04419	398.16	0.02210	0.01563	238.88
	GA	1.27568	0.09013	154.32	0.04844	0.04421	164.77
	ABCA	1.30830	0.08902	1234.01	0.09353	0.13637	1196.08

Table D.9 Optimized parameters with computation time using Wavelet packet transform features (for the same speed)

Speed (Hz)	Optimization method	*C*-SVC classifier			ν -SVC		
		C	γ	Time (sec.)	ν	γ	Time (sec.)
For same speed							
10	GSM	1.41421	0.35355	397.89	0.17678	0.01563	171.43
	GA	1.48250	0.19012	106.53	0.03408	0.02472	118.30
	ABCA	1.49144	0.14479	634.97	0.05154	0.01054	765.87
15	GSM	1.41421	0.06250	364.68	0.06250	0.01563	171.49
	GA	1.34233	0.15010	132.05	0.21819	0.19073	153.03
	ABCA	0.80285	0.14249	882.77	0.11200	0.11554	1036.72
20	GSM	1.00000	0.03125	354.52	0.02210	0.03125	187.76
	GA	1.03048	0.27177	155.04	0.19914	0.01104	152.03
	ABCA	1.29841	0.27145	951.89	0.14316	0.03878	1104.55
25	GSM	1.00000	0.04419	372.69	0.08839	0.03125	235.12
	GA	1.11983	0.01514	111.14	0.01176	0.01964	137.72
	ABCA	1.47936	0.01851	927.58	0.03451	0.01506	1023.49
30	GSM	0.70711	0.04419	393.70	0.06250	0.03125	252.75
	GA	0.92253	0.01850	116.44	0.01195	0.01964	142.66
	ABCA	0.35898	0.01177	933.13	0.17080	0.01920	936.21

Table D.10 Optimized parameters with computation time using Wavelet packet transform features (for the interpolation and extrapolation speed)

Speed (Hz)	Optimization method	*C*-SVC classifier			ν -SVC		
		C	γ	Time (sec.)	ν	γ	Time (sec.)
For interpolation speed (range 5 Hz)							
12.5	GSM	1.00000	0.12500	457.09	0.06250	0.03125	209.83
	GA	1.03182	0.13769	137.99	0.07701	0.03030	146.12
	ABCA	1.50000	0.09651	945.09	0.06853	0.01337	1042.55
17.5	GSM	1.41421	0.12500	423.98	0.12500	0.02210	214.04
	GA	0.92253	0.19769	152.49	0.29405	0.19618	182.53
	ABCA	1.50000	0.04886	1011.55	0.20822	0.00478	1137.89
22.5	GSM	1.41421	0.04419	417.01	0.02210	0.01563	230.23
	GA	1.18093	0.06012	141.12	0.16298	0.08774	181.90
	ABCA	1.36260	0.05966	1043.62	0.17764	0.08202	1201.82
27.5	GSM	1.41421	0.04419	436.32	0.04419	0.01563	256.52
	GA	0.92253	0.01527	124.71	0.00580	0.01104	149.18
	ABCA	1.29462	0.01851	1022.99	0.01432	0.01177	1024.71
For interpolation speed (range 10 Hz)							
15	GSM	1.41421	0.06250	456.46	0.06250	0.02210	219.90
	GA	0.82539	0.19886	151.25	0.13439	0.08774	177.19
	ABCA	1.15113	0.15727	1059.67	0.13256	0.07216	1118.37
20	GSM	1.41421	0.08839	544.73	0.02210	0.03125	238.06
	GA	1.19223	0.01323	140.76	0.10238	0.08774	181.98
	ABCA	1.14055	0.01851	1058.12	0.14451	0.00541	1161.99
25	GSM	1.41421	0.06250	436.29	0.08839	0.01563	257.62
	GA	1.12253	0.05789	149.06	0.13401	0.02472	200.59
	ABCA	1.50000	0.03304	1100.95	0.14512	0.04702	1250.97
For extrapolation speed (range 5 Hz)							
15	GSM	1.41421	0.12500	422.01	0.12500	0.04419	189.97
	GA	1.48250	0.08774	109.86	0.04900	0.08774	133.19
	ABCA	1.50000	0.09011	924.21	0.05703	0.08079	897.11
25	GSM	1.41421	0.08839	385.31	0.03125	0.02210	226.87
	GA	1.36838	0.06614	139.20	0.21639	0.13586	180.84
	ABCA	1.43933	0.07089	1028.69	0.07533	0.01199	1078.85
30	GSM	1.00000	0.03125	395.27	0.12500	0.00138	253.26
	GA	0.92253	0.01527	125.24	0.00580	0.11726	154.44
	ABCA	0.25620	0.06479	1054.00	0.12432	0.00001	1040.70
For extrapolation speed (range 10 Hz)							
20	GSM	1.00000	0.12500	431.56	0.06250	0.03125	209.39
	GA	1.03182	0.13769	134.90	0.07701	0.03030	147.09
	ABCA	1.50000	0.09651	948.15	0.06853	0.01337	1046.89
25	GSM	1.41421	0.12500	417.15	0.12500	0.02210	217.18
	GA	0.92253	0.19769	159.69	0.29405	0.19618	181.89
	ABCA	1.50000	0.04886	1010.01	0.20822	0.00478	1159.89
30	GSM	1.41421	0.04419	413.08	0.02210	0.01563	231.88
	GA	1.18093	0.06012	146.77	0.16298	0.08774	180.55
	ABCA	1.36260	0.05966	1044.84	0.17764	0.08202	1214.02

APPENDIX E: Parametric Sensitivity Analysis

E.1 Parametric Sensitivity Analysis

It is importance to observe the performance measures with respect to parametric variations for better evaluation of optimized SVM parameters. Variation in parameters has been done by generating random points of ±1% variation in the neighbourhood of corresponding optimised values. A uniform distribution was used to generate these random points. For example, when performing the sensitivity analysis with respect to C, it is varied by 1% around its optimum value and other parameter γ is kept constant. The study is limited to the same speed with time domain feature.

Table E.1 Parametric sensitivity values of optimized parameters for C–SVC

+1% deviation of optimized design variables		Average % change in accuracy	−1% deviation of optimized design variables		Average % change in accuracy
Opt. C	Opt. γ		Opt. C	Opt. γ	
10 Hz speed					
Yes	No	-0.25	Yes	No	0.06
No	Yes	0.06	No	Yes	0.06
Yes	Yes	-0.45	Yes	Yes	0.06
15 Hz speed					
Yes	No	0.33	Yes	No	0.33
No	Yes	0.33	No	Yes	0.33
Yes	Yes	0.33	Yes	Yes	0.33
20 Hz speed					
Yes	No	0.06	Yes	No	0.06
No	Yes	0.06	No	Yes	0.06
Yes	Yes	0.06	Yes	Yes	0.06
25 Hz speed					
Yes	No	0.0	Yes	No	0.0
No	Yes	0.0	No	Yes	0.0
Yes	Yes	0.0	Yes	Yes	0.0
30 Hz speed					
Yes	No	0.0	Yes	No	0.0
No	Yes	0.0	No	Yes	0.0
Yes	Yes	0.0	Yes	Yes	0.0

It is observed that there is not much variation of average accuracy on ±1% deviation of the parameter C and γ individually for C–SVC. It is observed that at 10 Hz speed the decrease of accuracy 0.25% with increase of C, and 0.06% increase with decrease of C. On the other hand, minor increase of accuracy by 0.06% with increase or decrease of γ. In case of 15 Hz speed the accuracy is increased by 0.33% with increase or decrease of C or γ. Moreover, the 20 Hz speed the increase of 0.06%. Similarly, it is observed that the variation of accuracy on ±1% deviation of parameters γ and ν (individually) decreases by 3.22% at 10 Hz rotational speed, but this decrement are 2.33% to 1.58%, respectively, at 15 Hz and 20 Hz rotational speeds for the ν –SVC.

Table E.2 Parametric sensitivity values of optimized parameters for ν–SVC

+1% deviation of optimized design variables		Average % change in accuracy	−1% deviation of optimized design variables		Average % change in accuracy
Opt. ν	Opt. γ		Opt. ν	Opt. γ	
10 Hz speed					
Yes	No	-3.22	Yes	No	-3.22
No	Yes	-3.22	No	Yes	-3.22
Yes	Yes	-3.22	Yes	Yes	-3.22
15 Hz speed					
Yes	No	-2.33	Yes	No	-2.33
No	Yes	-2.33	No	Yes	-2.33
Yes	Yes	-2.33	Yes	Yes	-2.33
20 Hz speed					
Yes	No	-1.58	Yes	No	-1.58
No	Yes	-1.58	No	Yes	-1.58
Yes	Yes	-1.58	Yes	Yes	-1.58
25 Hz speed					
Yes	No	0.0	Yes	No	0.0
No	Yes	0.0	No	Yes	0.0
Yes	Yes	0.0	Yes	Yes	0.0
30 Hz speed					
Yes	No	0.0	Yes	No	0.0
No	Yes	0.0	No	Yes	0.0
Yes	Yes	0.0	Yes	Yes	0.0

In the above analysis, we discussed only the effect of a single parameter variation, the simultaneous variation in parameters are discussed now. At 10 Hz speed the accuracy is decreased by 0.45% and is increased by 0.06% for variation of both C and γ in the positive and negative signs, respectively, for the case of $C-$SVC. The increase of accuracy by 0.33% and 0.06% for 15 Hz and 20 Hz speeds, respectively, in both signs of simultaneous parameter variation. It is observed in the case of $\nu-$SVC, that the decrements of accuracy are 3.22%, 2.33% and 1.58% for variation of both parameters γ and ν in both direction during 10 Hz, 15 Hz and 20 Hz speed, respectively.

In both the SVMs at higher rotational speeds (25 Hz and 30 Hz), the variation is negligible small. A quantitative approach has been chosen for the sensitive analysis, therefore we have to compute sensitivity of performance measure around a given solution point. The summary of the parametric sensitivity analysis of optimized design parameters for the $C-$SVC and $\nu-$SVC are tabulated in Table E.1 and E.2. It is observed that there is not much variation of average accuracy on ±1% deviation of the parameter C and γ individually for $\tilde{C}$SVC. It is observed that at 10 Hz speed the decrease of accuracy 0.25% with increase of C, and 0.06% increase with decrease of C. On the other hand, minor increase of accuracy by 0.06% with increase or decrease of γ. In case of 15 Hz speed, the accuracy is increased by 0.33% with increase or decrease of C or γ. Moreover, the 20 Hz speed the increase of 0.06%. Similarly, it is observed that the variation of accuracy on ±1% the deviation of parameters γ and ν (individually) have decrease of 3.22% at 10 Hz rotational speed, but this decrements are 2.33% to 1.58%, respectively, at 15 Hz and 20 Hz rotational speeds for the $\nu-$SVC.

In the above analysis, we discussed only the effect of a single parameter variation, the simultaneous variation in parameters are discussed now. At 10 Hz speed, the accuracy is

decreased by 0.45% and is increased by 0.06% for variation of both C and γ in the positive and negative signs, respectively, for the case of C-SVC. The increase of accuracy by 0.33% and 0.06% for 15 Hz and 20 Hz speeds, respectively, have been found in both signs of simultaneous parameter variation.

From the sensitive analysis, it could be concluded that the variation of both parameters on $C-$SVC is not significant but it decreases by 3.22% in $\nu-$SVC. The same procedure can be followed for the interpolation and extrapolation speed as well using other domain features.

Publication from the Present Thesis

Journals:

1. **D. J. Bordoloi** and R. Tiwari, 2013, 227(11), 2428-2439, *Proceedings of the Institution of Mechanical Engineers, Part C, Journal of Mechanical Engineering Science*, Optimization of Support Vector Machine based Multi-Fault Classification with Evolutionary Algorithms from the Time Domain Vibration Data of Gears.
2. **D. J. Bordoloi** and R. Tiwari, 2014, 73, 49-60, *Mechanism and Machine Theory*, Optimum Multi-fault Classification of Gears with Integration of Evolutionary and SVM Algorithms.
3. **D. J. Bordoloi** and R. Tiwari, 2014, 55, 1-14, *Measurement*, Support Vector Machine Based Optimisation of Multi-Fault Classification of Gears with Evolutionary Algorithms from Time-Frequency Vibration Data.
4. **D. J. Bordoloi** and R. Tiwari, 2014, 17(3), 39-49, *International Journal of Condition Monitoring and Diagnostic Engineering (COMADEM)*, Unification of Multi-class Fault Classification from Diverse Domain Features of Gear Using SVM Algorithms.

Conferences:

1. **D. J. Bordoloi** and R. Tiwari, 2012, Health Monitoring of Gears Based on Vibrations by Support Vector Machine Algorithms, *ASME 2012 Gas Turbine India Conference*, December 1, 2012, Mumbai, India (GT India 2012-9586).
2. **D. J. Bordoloi** and R. Tiwari, 2013, Health Monitoring of Gear Elements Based on Time-Frequency Vibration by Support Vector Machine Algorithms, *ASME 2013 Gas Turbine India Conference*, December 5-6, 2013, Bangalore, Karnataka, India (GT India 2013-3772).
3. **D. J. Bordoloi** and R. Tiwari, 2013, Health Monitoring of Gear Elements Based by SVM Algorithms, third National Symposium on Rotor Dynamics (*NSRD 2014)*, February 12-14, 2014, Bangalore, Karnataka, India.