ADVANCE FAULT DIAGNOSIS OF POWER-PLANT COMPONENTS USING VIBRATION SIGNATURE ANALYSIS

THESIS

submitted in fulfillment of the requirement of the degree of

DOCTOR OF PHILOSOPHY

to

J.C. BOSE UNIVERSITY OF SCIENCE & TECHNOLOGY, YMCA

by

CHHAYA GROVER

YMCAUST/ Ph 15/ 2K12

Under the Supervision of

Dr. NEELAM TURK

PROFESSOR

Department of Electronics Engineering

Faculty of Engineering & Technology

J.C. Bose University of Science & Technology, YMCA

Sector-6, Mathura Road, Faridabad, Haryana, India

AUGUST, 2021

This thesis is dedicated

to

my beloved daughters

for their constant support and unconditional love

DECLARATION

I hereby declare that this thesis entitled **ADVANCE FAULT DIAGNOSIS OF POWER-PLANT COMPONENTS USING VIBRATION SIGNATURE ANALYSIS** by **CHHAYA GROVER**, being submitted in fulfillment of the requirements for the Degree of Doctor of Philosophy in ELECTRONICS ENGINEERING under Faculty of Engineering and Technology of J.C. Bose University of Science and Technology, YMCA, during the academic year 2020-2021, is a bona fide record of my original work carried out under the guidance and supervision of **Dr. NEELAM TURK, PROFESSOR, DEPARTMENT OF ELECTRONICS ENGINEERING** and has not been presented elsewhere.

I further declare that the thesis does not contain any part of any work which has been submitted for the award of any degree either in this university or in any other university.

(Chhaya Grover)

Registration No. YMCAUST/ Ph 15/2K12

CERTIFICATE

This is to certify that this thesis entitled **ADVANCE FAULT DIAGNOSIS OF POWER-PLANT COMPONENTS USING VIBRATION SIGNATURE ANALYSIS** by **CHHAYA GROVER**, submitted in fulfillment of the requirement for the Degree of Doctor of Philosophy in **ELECTRONICS ENGINEERING** under Faculty of Engineering & Technology of J.C. Bose University of Science & Technology, YMCA, Faridabad, during the academic year 2020-2021, is a bonafide record of work carried out under my guidance and supervision.

I further declare that to the best of my knowledge, the thesis does not contain any part of any work which has been submitted for the award of any degree either in this university or in any other university.

Dr. NEELAM TURK
PROFESSOR
Department of Electronics Engineering
Faculty of Engineering & Technology
J.C. Bose University of Science & Technology
YMCA, Faridabad

Dated:

ACKNOWLEGMENT

Through my years of Ph.D. work, I had an unvarying light of support and guidance in the form of my Supervisor **Dr. Neelam Turk**, Professor, Department of Electronic Engineering , J.C. Bose University of Science & Technology, YMCA, Faridabad. I am immensely grateful to her for her invaluable supervision, advice, direction and intelligence while concurring the milestones of my academic research. Her wealth of experience and insightful feedback pushed me to sharpen my thinking and bring my work to a higher level. I would also like to acknowledge her for providing administrative support as the Chairperson, Department of Electronics Engineering, in the recent years.

I would like to express my gratitude to **Prof. Dinesh Kumar**, Vice-Chancellor, J.C. Bose University of Science & Technology, YMCA, Faridabad, for his constant motivation and support.

My sincere thanks to **Dr. Munish Vashishtha**, Professor, Department of Electronic Engineering, J.C. Bose University of Science & Technology, YMCA, Faridabad, for his thoughtful suggestions and recommendations that helped me widen my research from various perspectives.

I am also thankful to **Dr. Kshitij Gupta**, Professor, IIT Delhi for giving me the opportunity to conduct experiments in "Vibration Research Lab" at IIT Delhi.

Last but not the least, I would like to thank my family: my **husband**, my **daughters** and my **parents** for supporting me spiritually throughout my Ph.D. and my life in general. Without their wise counsel, tremendous understanding, encouragement, and support in the past few years, it would have been impossible for me to complete my research work.

(Chhaya Grover)

Registration No. YMCAUST/ Ph 15/2K12

ABSTRACT

Power plants, generators of electric power, are lifelines of every country. They majorly consist of driving units (e.g. motor, turbine), transmission units (e.g. gearbox, coupling), processing machines (e.g. generator) and auxiliary units (compressor, pump). Many of these units are rotating equipment which make use of bearings, gears and rotors. Out of these, bearings are the key mechanical components of rotating machines in power plants. Whenever a rotating machine of a power plant goes through unprecedented downtime, a common belief throughout the industry is that the bearing failed. Failure of a bearing can be disastrous as it can bring huge capital losses. Early fault detection and diagnosis of bearings in rotating machinery is crucial for increasing reliability and lifetime of modern power-plant equipment.

Even though considerable prior research has been conducted, most of the research conducted so far has been tested on simulated data from laboratory setups. Moreover, majority of the previous studies on Vibration Signal Processing, have focused more on optimising fault diagnosis classifiers, and few have discussed the reason behind the use of a set of particular statistical features. Furthermore, many nonlinear signals or nonlinearities in signals, cannot be examined properly by second order statistical methods, and this led to researchers exploring higher order statistics (HOS) methods which help in preserving phase and amplitude of signals.

This research aims to fill this research gap by devising vibration-based methodologies for detecting embryonic defects in rotating equipment bearing heterogeneous data sets from Vibration Labs, Online Repositories and Real industrial system in a 500 MW power plant. It develops a framework of Intelligent Fault Diagnosis for Vibration based condition monitoring of bearings based on Signal processing and Artificial Intelligence (AI) Techniques. It mainly proposes two methodologies: 1. Vibration Signature Analysis using Statistical Features and Machine Learning and 2. Vibration Signature Analysis using Bispectrum Images and Deep Transfer Learning.

To achieve this, this research presents a review on the early fault diagnosis of bearings wherein a number of techniques for rotating machinery are surveyed in terms

of AI-based intelligent fault diagnosis. It then proposes novel statistical features following which it develops a strategy to conduct an extensive analysis of new and old statistical features, in order to derive an optimal set of fault-sensitive features to be fed as input to Machine Learning classifiers. It also proposes a novel HOS based method for analyzing Bispectrum of vibration signals using advanced AI techniques of Deep Transfer Learning.

The main contribution of this research work is three-fold: Firstly, proposing Hjorth Parameters and Normal negative log likelihood for Gaussian Mixture Model (GMM) for rolling element ball bearing fault detection and severity estimation; secondly, devising a strategy to probe a comprehensive set of statistical time domain features with the objective of identifying optimal feature subset to feed as input to Machine Learning classifiers for rolling element ball bearing fault detection and severity estimation on data acquired from an operational power plant; third, in the frequency domain, sensitivity and effectiveness of Bispectrum for fault diagnosis has been investigated using deep transfer learning neural networks. The third contribution eliminates the need for manual feature extraction and also attempts to solve the most omnipresent problem which is faced with diagnosing most of the data sets - shortage of data samples - in this case, ball bearing vibration dataset.

The outcomes of this research can be applied for intelligent fault diagnosis of bearings in rotating machinery of power plants. Hence, this research can serve as a handbook on advanced statistical techniques in association with artificial intelligence techniques for researchers working on bearing fault diagnosis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
AM	Arithmetic Mean
ANN	Artificial Neural Network
AUC	Accuracy
AUPRC	Area Under Precision Recall Curve
AUROC	Area Under the ROC curve
BB	Ball Bearing
CBD	City-Block distance
CF	Crest Factor
CM	Condition Monitoring
CNN	Convolution Neural Network
CWRU	Case Western Reserve University
DBVDM	Density-Based Value Distance Measure
DL	Deep Learning
EEMD	Ensemble Empirical Mode Decomposition
EMD	Empirical Mode Decomposition
FFT	Fast Fourier Transform
FN	False Negative
FP	False Positive
G-Mean	Geometric Mean
GMM	Gaussian Mixture Model
HA	Hjorth Parameter 1(Activity)
HC	Hjorth Parameter 3(Complexity)
HD	Hamming Distance
HHT	Hilbert Huang Transform
HL	Histogram Lower Bound
HM	Hjorth Parameter 2(Mobility)
HOS	Higher Order Statistics
HOSA	Higher Order Spectral Analysis

HU	Histogram Upper bound
IF	Impulse Factor
IMF	Intrinsic Mode Function
IR	Inner Race
IVDM	Interpolated Value Difference Metric
KNN	K- Nearest Neighbours
KF	Kurtosis Factor
KU	Kurtosis
MAD	Mean Absolute Deviation
MAE	Mean Absolute Error
MCC	Matthews Correlation Coefficient
MeAD	Median Absolute Deviation
MF	Margin Factor
MFPT	Machinery Failure Prevention Technology
ML	Machine Learning
NN	Neural Network
NNLL	Normal Negative Log Likelihood
OR	Outer Race
PCC	Pearson Correlation Coefficient
PNN	Probabilistic Neural Network
PSD	Power Spectral Density
Q1	First Quartile
Q2	Second Quartile
Q3	Third Quartile
ReLU	Rectified Linear Unit
REP	Reduced Error Pruning
ResNet	Residual Neural Network
RF	Random Forest
RIPPER	Repeated Incremental Pruning to Produce Error Reduction
RMS	Root Mean Square
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristic

RP	Recurrence Plot
RSSQ	Root-Sum-of-Squares
SD	Standard Deviation
SF	Skewness Factor
SGD	Stochastic Gradient Descent
SHF	Shape Factor
SIFT	Scale Invariant Feature Transform
SK	Skewness
SMOTE	Synthetic Minority Oversampling Technique
STFT	Short-Time Fourier Transform
SURF	Speeded Up Robust Features
SVM	Support Vector Machine
T1	Time domain statistical feature 1
T30	Time domain statistical feature 30
TLM	Transfer Learning Method
TN	True Negative
TP	True Positive
VAR	Variance
VDM	Value Difference Metric
WNLL	Weibull Negative Log Likelihood
ZCR	Zero Crossing Rate

CHAPTER I

INTRODUCTION

1.1 PREAMBLE

Electricity is essential for modern life, making Power Plants a necessity for any country. The main components of power plants are driving units (e.g. electrical or hydraulic motor, turbine), transmission units (e.g. gearbox, coupling), processing machines (e.g. generator) and auxiliary units (compressor, pump). Majority of these components are rotating equipment and hence rely on bearings, gears, rotors, etc., among which bearings are the main facilitators of rotatory motion. In the event of a rotating machine failure in a power plant, the cause is usually attributed to a bearing failure. Bearing failure is intriguing in the sense that in some cases, a single event can cause the failure and in others, occurrence of multiple conditions such as irrational human decisions, harsh environmental conditions and substandard maintenance can contribute to early failure. This ultimately results in large scale failure of the rotating component and additional damage to related equipment. In fact, there is a bearing failure manual released by Electric Power Research Institute for discussing common bearing faults found in the rotating equipment installed in electric power plants [1]. Bearing failure can be catastrophic, bringing massive monitory losses. Hence, early fault detection and diagnosis of bearings in rotating machinery is essential to increase reliability of modern power plants.

Vibration based condition monitoring is highly popular and well-accepted for power plant rotating machines because the machine vibration response is sensitive to any small structural or process parameter change. Industry and academia have conducted massive amounts of research work in the last two decades involving Vibration Signal Processing and Artificial Intelligence for early failure detection of bearings, thereby increasing the lifetime of power-plant equipment and avoiding potential damages to other connected equipment.

Despite considerable prior research being conducted, there is one caveat to it: seldom, sufficient field data is available from real power plant equipment and hence most of the research conducted so far has been tested on simulated data from

laboratory setups. However, laboratory setups being different from industrial equipment with regards to scale and operating conditions, there is no guarantee that research solutions validated with simulated data will perform equally well on actual systems. While conducting experiments the validity of the proposed solution or algorithm is best confirmed when it's tested on data from real industrial systems. Also, continuing with the discussion on the ongoing research involving Vibration Signal Processing and Artificial Intelligence, researchers are able to utilise the power of statistical features by applying popular supervised machine learning techniques. Although all of the surveyed studies have applied various time-domain parameters, whether conventional or novel and dimensional or dimensionless, for bearing fault diagnosis, either exclusively or in addition to other frequency domain and time-frequency features, none of the studies discusses and provides the reason behind the use of a set of particular statistical features. Majority of the previous studies, till now, have focused more on optimising fault diagnosis classifiers, and few have compared established features with novel ones.

Moreover, even though in signal processing, first and second order statistics have shown their significance in fault diagnosis, many nonlinear signals or non-linearities in signals, cannot be examined accurately by second order statistical methods, resulting in researchers to explore higher order statistical (HOS) methods. HOS methods help in preserving phase and amplitude of signals hence enabling reconstruction of signals and thereby being more suitable than second-order statistics for fault diagnosis of vibration signals.

The information gathered on the current limitations and already conducted research work on bearing fault diagnosis, led on to start this research journey with the motivation of filling this research gap by devising vibration-based methodologies for detecting nascent defects in rotating equipment bearings heterogeneous data sets from small-scale test beds and a real industrial system in a 500 MW power plant. In a nutshell, this research presents a review on the early fault diagnosis of bearings wherein a number of techniques for rotating machinery are surveyed in terms of time and frequency domain statistical features and Artificial Intelligence (AI) based diagnosis. It then emphasises on features by proposing novel statistical features and developing a strategy to conduct an extensive analysis of new and old statistical features, so as to derive the most optimal

set of fault-sensitive features. It also devises a methodology involving HOS methods specifically Bispectrum analysis of vibration signals using advanced AI techniques of Deep Learning. This methodology eliminates the need for manual feature extraction and also attempts to solve the most omnipresent problem faced in diagnosing ball bearing vibration datasets- shortage of data samples. Finally, this research can also serve as a handbook for researchers in the field of early fault diagnosis of bearings using advanced statistical techniques along with Artificial Intelligence.

1.2 RESEARCH OBJECTIVES

Vibration signature analysis techniques are popularly used for fault diagnosis of rotating machines and are improving with time due to the advances in disciplines such as Statistics, Signal Processing and Artificial Intelligence. This research is on the application of vibration signature analysis for fault diagnosis of rotating components of power plants. The objectives of this research are outlined and listed below:

1. To review and analyze the condition monitoring of power plant components using vibration signature analysis.

2. To pre-process vibration signals acquired from vibration labs and real power plant using advanced signal processing techniques capable of handling the non-stationary and nonlinear character of the vibration signals.

3. To extensively research a comprehensive list of time-domain statistical features and to propose new advanced statistical features having more sensitivity to faults.

4. To propose a methodology to rank all extracted features in order to generate an optimal feature subset and apply effective machine learning classifiers for accurate bearing fault classification and fault severity estimation using vibration signature analysis.

5. To validate the proposed methodology on real power plant data and verify the robustness of the proposed method.

6. To investigate the sensitivity and effectiveness of higher order statistics, Bispectrum, for fault diagnosis by designing a transfer learning based deep learning methodology.

1.3 ORGANIZATION OF THESIS

This thesis is organized in 9 chapters as explained below. The thesis organization covering proposed methodologies involving Machine Learning and Deep Transfer Learning Frameworks is also depicted in the form of a flow chart in Figure 1.1.

Chapter 2 covers the literature review of Vibration-based Fault Diagnosis of Power-plant Components; Vibration Signature analysis using feature extraction in time, frequency and time-frequency domains; and Vibration Signature analysis using artificial intelligence techniques.

Chapter 3 discusses Vibration Data Acquisition from Vibration Research Lab, Vibration Data Repositories and the Power Plant. Experimental Setup and Dataset Description for every case is elaborated.

Chapter 4 deals with pre-processing of raw vibration signals prior to feature extraction. Signal pre-processing technique, Empirical Mode Decomposition (EMD) and best intrinsic mode function (IMF) selection technique along with the obtained results are presented in this chapter.

Chapter 5 deals with extraction of fault sensitive features from pre-processed vibration signals. Conventional Statistical Features as well as proposed Novel Statistical Features are discussed in this chapter.

Chapter 6 covers Feature Subset Selection from the whole set of extracted features for the purpose of dimensionality reduction. Theoretical aspects, proposed methodology and results of Filter Based Feature Ranking and Conventional Search & Swarm Search Techniques based Feature Subset selection have been discussed in this chapter.

Chapter 7 deals with Fault Diagnosis and Classification using optimal statistical features subset and machine learning classifiers. Experimental validation involving three case studies has been detailed.

Chapter 8 examines Fault Detection and Diagnosis using Higher Order Spectrum Analysis. Theoretical frameworks of Bispectrum, Image based fault diagnosis, Deep Neural Network Architectures and Transfer Learning have been discussed along with the results from the experimental study on bearing dataset fault diagnostics.

Chapter 9 summarizes the findings of this research work and gives suggestions for future work. It also covers the detailed point-by-point comparison of proposed Machine Learning and Deep Learning frameworks.

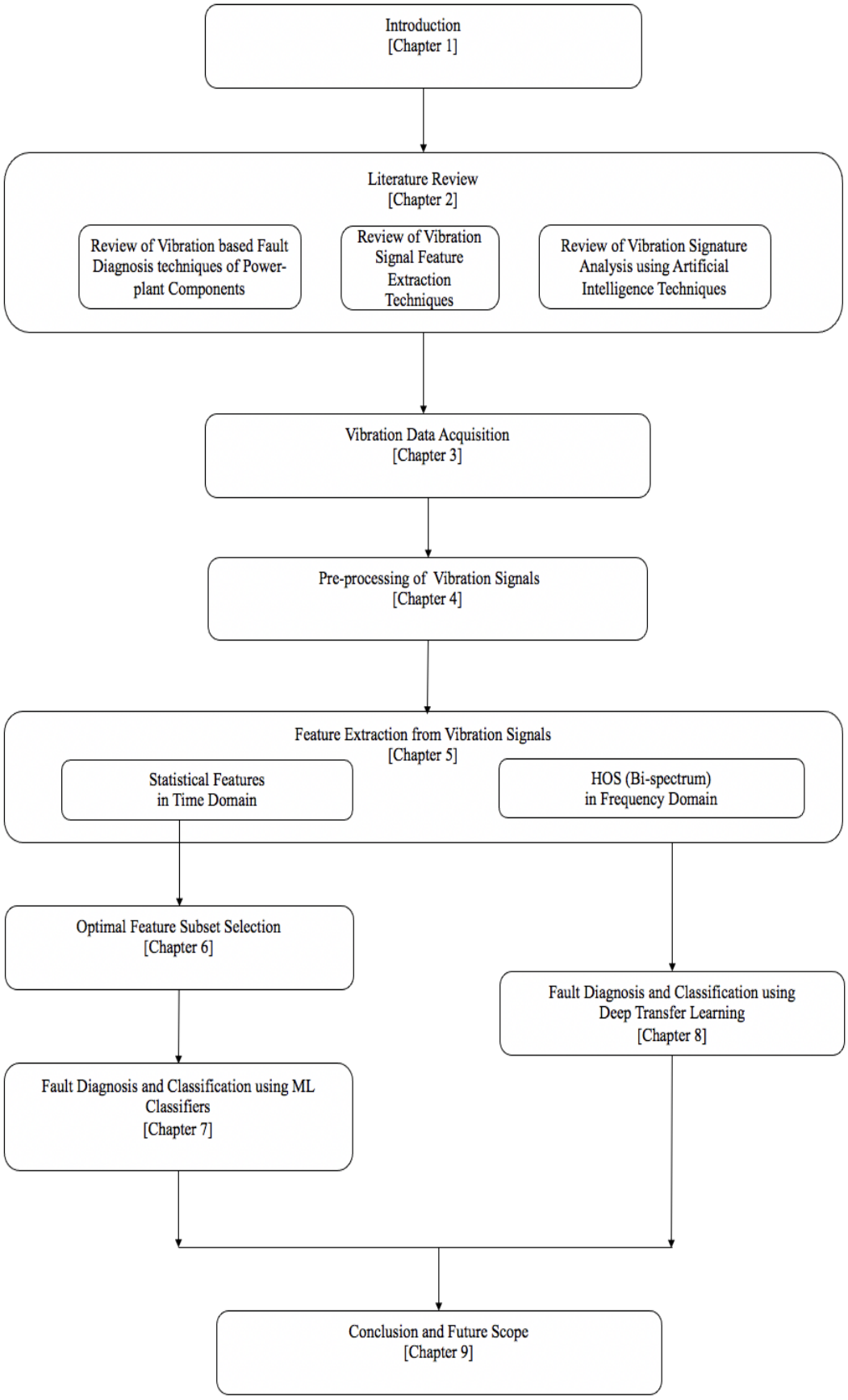


Figure 1.1: Organization of thesis

CHAPTER II

LITERATURE REVIEW

2.1 INTRODUCTION

In this chapter, the literature on condition monitoring of power plant components using vibration signal processing is reviewed. The literature review is divided into three parts.

Unscheduled outages of power plants due to sudden failures of key-components like motors, turbines and pumps may result in severe economic and environmental implications. Hence monitoring and maintenance methodology capable of providing the best technical solution to ensure safe operation of the plant is very important. **In the first part**, vibration-based fault diagnosis of Power-plant components is covered. A detailed survey of the techniques used for vibration signature analysis by the researchers is reviewed and presented.

Extraction of fault sensitive features for vibration based condition monitoring is an extensively researched domain. **In the second part**, extraction of fault sensitive features from the vibration signal in time, frequency and time-frequency domain has been reviewed.

Fault diagnosis and prognosis in power-plant components is often a labour-intensive and time-consuming practice, hence researchers are increasingly interested in automating the diagnosis procedure using artificial intelligence techniques along with advanced signal processing techniques. **In the third part** of this chapter, Tools and techniques from machine learning and deep learning domains, that can be used in conjunction with effective signal processing techniques for effective decision making are reviewed.

2.2 VIBRATION-BASED CONDITION MONITORING OF ROTATING MACHINERY OF POWER-PLANT

A power plant is an industrial facility for the generation of electrical energy from energy such as heat energy (thermal), chemical energy (nuclear) and mechanical energy (hydro, wind). Figure 2.1. is a schematic diagram of a Thermal Power Plant. A

thermal power plant mainly consists of Boiler, Turbine, Condenser and Pump. Firstly, coal is burnt to produce high temperature and high pressure steam in a boiler. The steam is passed through a steam turbine to produce rotational motion. The generator, mechanically coupled to the turbine, thus rotates producing electricity. After imparting energy to the turbine rotor, the steam passes out of the turbine blades into the condenser where it condenses to cold water which is then fed to the boiler using a pump. The figure just depicts a high-level description of a power plant, there are a number of pumps, like feed-water pumps and circulating pumps, which are indispensable to an industrial power plant pipeline. All rotating machines like turbines (e.g. low pressure/ high pressure Turbines), motors, pumps and compressors have bearings installed inside them in order to reduce frictional losses incurred during rotational motion.

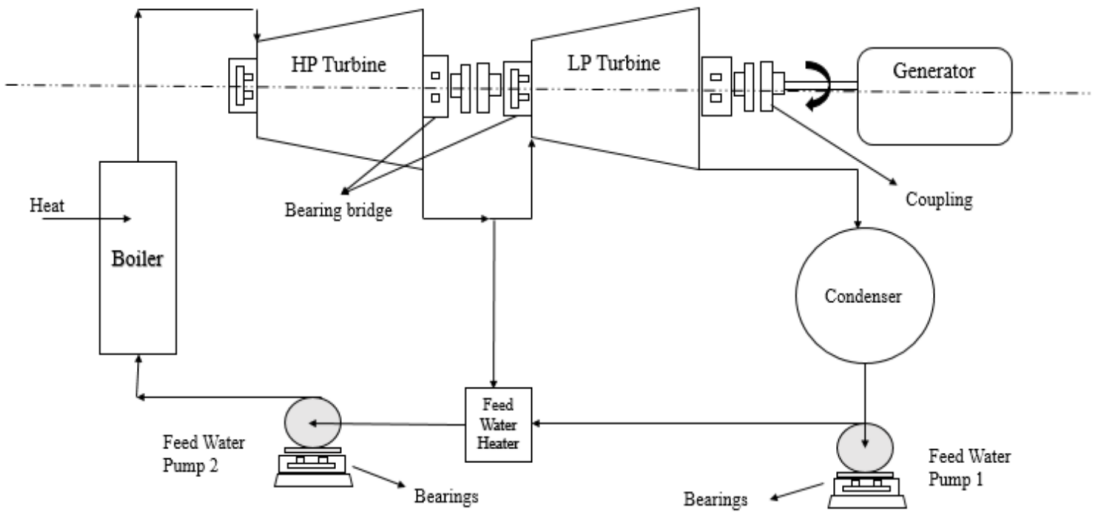


Figure 2.1: Schematic diagram of Thermal Power Plant

Predictive maintenance is crucial for power plants because it helps in preventing severe economic and environmental losses which can be incurred due to unprecedented failures of key-components like motors, turbines and pumps. This is achieved by 'Condition Monitoring'. Condition monitoring is the continuous tracking of the state of an equipment using parameters such as vibration, current, temperature etc. without interrupting its operation in the power plant. This monitoring is done at regular intervals in order to detect abnormal changes in the parameter values which is helpful in indicating the occurrence of faults in these equipment. Since failures occur

to individual components, the monitoring measurements need to focus on the particular fault types and severity of the critical components. The goal is to get close to the ideal target of 100% availability of power plant equipment during operational time by early fault diagnosis and ordering for spare parts.

The monitoring also helps in avoiding secondary damages to the connected components which get affected by the faulty component and also to plan the maintenance activity for increasing the operational time of the plant. C. De Michelis et al. described condition monitoring and assessment of power plant components in their work. They presented damage mechanisms and fault diagnostic methods for four main components of power plants: steam turbine, gas turbine, boiler and heat recovery steam generator. Current developments and future trends in these areas are also discussed [2].

Effective predictive maintenance of rotary machinery includes several electrical and non-electrical techniques as shown in Figure 2.2. A detailed study on four types of condition monitoring techniques i.e. vibration monitoring, noise monitoring, motor current signature analysis and wear debris analysis of systems in thermal power plant and also the methods for data analysis has been carried out in their work by Kurien C., Srivastava A.K. [3]. A case study on the effectiveness of condition monitoring techniques (vibration analysis, motor current signature analysis, noise monitoring and wear debris analysis) for fault diagnosis of three pumps (boiler feed water pump, auxiliary cooling water pump and condensate extraction pump) in thermal power plant has been presented by Kurien C., Srivastava A.K. [4].

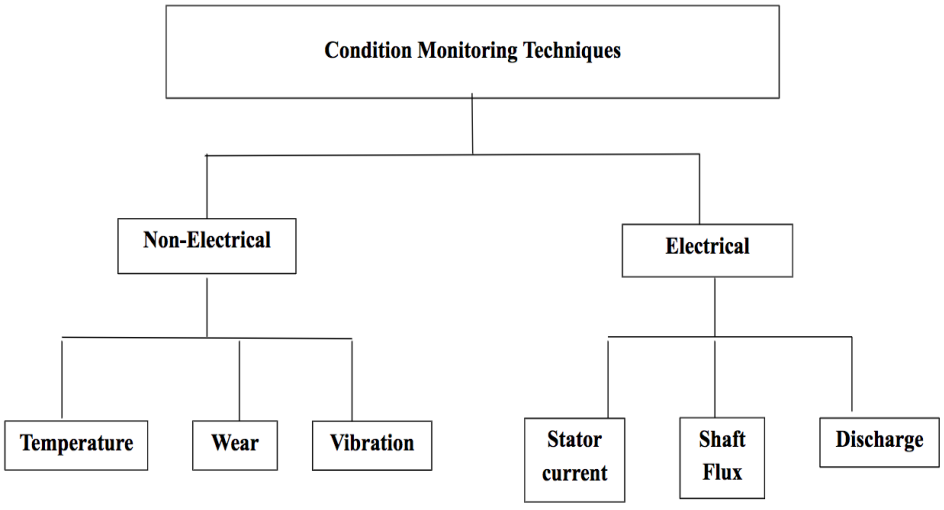


Figure 2.2: Condition monitoring techniques

Out of these techniques, vibration is a key part of a strong condition monitoring regime for a power plant as it contains several units which make use of rotational equipment for which, performing vibration analysis can help increase the lifetime of power-plant equipment [5–7]. Vibration monitoring techniques are found to be highly efficient in determining the defects of the rotating machine at an early stage. Each rotating machine presents a characteristic vibration that uniquely differentiates it, and is commonly known as Vibration Signature. When any component of a machine is subjected to some kind of fault, its measured vibration characteristics will exhibit changes from normal levels i.e. the vibration signature of the machine will change. This vibration signature can be processed to give information on the condition of the machine. Vibration monitoring can detect the presence of unbalanced forces generated due to misalignment, damaged bearing, unbalance in rotors, looseness of rotating parts, rubbing of rotating parts, cracked teeth in gear, damaged rotor bar, electrical defects, resonance, etc. The book "Industrial Approaches in Vibration-Based Condition Monitoring" by J.K. Sinha is a tutorial for the issues faced by vibration-based condition monitoring professionals, including good industrial examples [8]. Vibration based fault diagnosis and prognosis has been an active area of research for many years [9–12].

The forced shutdown of a power plant or its load reduction is usually due to failures in its rotating machinery. Therefore its vital that the condition of all rotating machinery such as steam turbine, gas turbine, generator and major auxiliaries such as pumps are maintained using a predictive maintenance strategy such as vibration monitoring, in order to eliminate unexpected breakdowns and ensuring long-term reliability and availability of machinery .

Steam or gas turbines are the most critical and the most expensive individual rotating machinery. T. Barszcz has presented Vibration-Based Condition Monitoring of Wind Turbines in his book [13]. When the turbine condition is known, repairs can be planned in a timely manner so that there's no unexpected long production downtime and the power plant's electricity consumers are not affected. It also enables monetary savings and financial planning as the required spares and tools can be arranged ahead of downtime.

Another category of important rotating machinery is pump. There are a number of

pumps installed in thermal power plants for effective generation of electricity. Various pumps used in thermal power plants include boiler feed water pump , auxiliary cooling water pump, seal oil pump, condensate extraction pump, ash disposal pump and others. Major reasons for abnormal vibrations in the pumps include damage of the bearings, imbalance in the rotor, misalignment of the rotating shaft.

Researchers have reviewed state-of-the-art research results that cover basic signal processing techniques; fault detection, diagnostic and prognostics techniques for bearings [14–16] and gears [17]. Recently, Wei, Yu et al. reviewed the research works on early fault diagnosis approaches of gears, rotors, and bearings to present a guide map for researchers in the field of fault diagnosis in rotating machinery [18]. Othman A., Damanhuri N.S., Hamzah N. used feature extraction from vibration signals for effective Fault diagnosis in power plant rotating machinery [19], on the other hand H. Oh et al. converted vibration signals to images and used deep learning techniques for fault diagnosis of power plant's rotor system in their work [20].

Research shows that in rotating machinery, the root cause of faults is often the faulty bearing. A bearing is a mechanical device designed to reduce friction in a part of a machine where another part turns or slides. Bearings are frequently used to facilitate the rotational movements in power plant turbines. These bearings can also be found in other equipment such as pumps, compressors, and electric motors in thermal, wind, and hydro power plants. Figure 2.1. shows some of the main bearing positions in a standard thermal power plant. There are several different types of bearings, including ball and roller bearings, linear bearing and journal bearings. Rolling bearings are among the most important components in the vast majority of machines, that are used for rotating or linear shaft applications. They consist of rolling elements, inner and outer races.

Turbine bearing failures in electric utilities have been responsible for outages amounting to 1.1% to 1.8% of theoretical power output. In addition to failures in turbine generators, bearing failures in other rotating equipment, including pumps, fans, and auxiliary gas turbines and motors, can also lead to plant outages. The Electric Power Research Institute (EPRI)'s bearing failure manual discusses bearing problems typical of the rotating equipment in electric utility power plants. The main purpose of the manual is to provide power plant engineering and service personnel with ways to properly identify specific modes of bearing damage, as well as root causes, followed

by a list of remedial actions applicable to each specific case [1].

Given the serious consequences of such breakdowns, determination of the causes of bearing failure and methods of effective repair are of paramount importance. When a little defect such as a crack or chip at the inner or outer surface of the bearing happens, it could lead to a major disaster to the rotating machine. Mass imbalance, shaft misalignment and the improper surface of the bearing itself could be the major factors in the faulty bearing that need to be treated specially. Exacting demands are made on their load carrying capability, running accuracy, noise levels, friction and frictional heat, life and reliability. Faults in rolling element bearings are considered a distinct indicator of decreasing pump performance and failure. Therefore, over the years bearings have been the subject of extensive research and continuous improvements.

A detailed manual of bearing failures and their repair in rotating equipment of power plants is given by Pinkus, Oscar [1]. Hu, Qiao et al. proposed a method using empirical mode decomposition (EMD), fuzzy feature extraction and support vector machine (SVM) for fault diagnosis of a 50 MW steam turbo-generator set in a power plant. They validated their methodology by collecting vibration data from acceleration transducers mounted on bearing bridges of the low-pressure turbine of the 50 MW steam turbo-generator set [21]. He, Yongyong et al. investigated the damage and failure of journal bearing of a power plant's turboset [22]. A case study on failure analysis of 320,000 kW steam turbine journal bearing was presented by Mehdizadeh, M. and Khodabakhshi, F. in their research [23]. Tazi N. and Chatelet E., Bouzidi Y. discussed wear analysis of wind turbine bearings in their work [24]. Pino G., Ribas J.R., Guimarães L.F. used HMM with orbit curves and Wavelet Packet Transform for fault diagnosis of hydro turbine components' (bearings) degradation [25].

Hence, whenever a machine is experiencing unscheduled downtime, the common phrase that is used throughout the industry - "The bearing failed", makes total sense.

2.3 VIBRATION SIGNATURE ANALYSIS USING FEATURES FROM DIFFERENT DOMAINS

For vibration signature analysis, researchers have identified and extracted many fault sensitive features in different domains. A review of various vibration feature

extraction techniques in time domain, frequency domain, and joint time frequency domain for fault diagnosis in rotating machines is presented in their studies by Yang H. et al. and Engin S.N. et al. [26, 27]. K. F. Tom has written a detailed primer on feature xtraction from vibraition signals for prognostics and diagnostics applications [28]. W.Caesarendra and T. Tjahjowidodo presented a review of feature extraction methods from vibration signals captured from low-speed slew bearing [29].

2.3.1 Vibration Signature Analysis in Time Domain

Vibration signals are initially obtained as a series of digital values representing proximity, velocity, or acceleration in the time domain. The research on vibration techniques in the time domain for various types of machinery can be categorized into the groups as shown in Table 2.1.

Table 2.1: Overview of Time Domain Vibration Feature Extraction techniques.

Raw Signal Based Methods	Statistical Parameters (e.g. RMS, Mean, Variance, Kurtosis, Skewness, Crest Factor etc.) Time Synchronous Averaged Based Methods (TSA Signal, Residual Signal, Difference Signal)
Filter Based Methods	Demodulation, Adaptive Noise Cancelling
Stochastic and Advanced Methods	Chaos, Blind Deconvolution, Thresholding, Autoregressive Model Based Methods

As far as statistical parameters are concerned, several types of statistical features have been used so far by the researchers for their classification approach. The conventional time domain statistical parameters, namely, mean value, maximum value, root mean square value, peak-to-peak value, root, standard deviation, variance, kurtosis and skewness are the most commonly used dimensional parameters [30–35]. A wide set of dimensionless statistical parameters like Crest factor, Shape factor, Impulse Factor, Clearance Factor, Skewness Factor and Kurtosis Factor have also been used to enhance the performance of fault diagnosis systems [36–39].

A major boost in bearing fault diagnosis research was seen when novel statistical features began to be applied in artificial intelligence techniques. Sreejith B. et al. proposed a method for pattern recognition by extracting Normal negative

log-likelihood and Kurtosis values of time domain signals and using them as input to an artificial neural network [40]. Abbasiona S. et al. used Weibull negative log-likelihood function of denoised signals and support vector machine (SVM) classifier for fault classification of rolling bearings [41], while researchers also applied zero crossings, mean absolute deviation and histogram features in their respective works [42–44].

The application of a new group of statistical parameters, called as "Hjorth Parameters", was explored for diagnosing 4 types of faults in rolling element ball bearings [45]. Lastly, different kinds of entropies have also been used for intelligent fault diagnosis in various researches [26], [46–48]. Some recent papers have developed symmetric cross entropy measures of single valued neutrosophic sets for fault diagnosis in bearings [49, 50].

Many filter based and Stochastic feature extraction techniques like synchronous averaging, resonance demodulation, adaptive filtering have also been researched and utilized to diagnose faults in rolling bearings and gears successfully [51–53].

2.3.2 Vibration Signature Analysis in Frequency Domain

Features providing frequency information of vibration signals such as frequency domain features and time frequency domain features have been widely investigated by researchers. An overview of developed frequency domain and time-frequency domain feature extraction techniques is given in Table.2.2.

Table 2.2: Overview of Frequency Domain and Joint Time-Frequency Domain Vibration Feature Extraction techniques.

First Order	Spectrum (FFT), Correlation of Spectrum, Short Time Fourier Transform (STFT), Continuous Wavelet Transform (CWT), Discrete Wavelet, Transform (DWT), Discrete Wavelet Packet Analysis (DWPA)
Second Order	Power Spectrum, Power Cepstrum, Cyclostationarity, Spectrogram, Wigner Distribution, Scalogram
Third Order	Bicoherence Spectrum, Bilinearity, Wigner Bi-Spectra
Fourth Order	Wigner Tri-Spectra

Spectral analysis of the vibration signals is perhaps the most widely used approach to bearing fault detection [54]. The Fast Fourier Transform (FFT) is the most conventional diagnosis technique and has been widely used to identify the frequency features of signals. These signals can be raw signals or pre-processed signals. The power spectrum, whose amplitude is the square of the amplitude of the spectrum is an effective method to diagnose machinery faults. The power cepstrum, which is a logarithm of the power spectrum was also applied to machinery fault diagnosis.

As shown in Table 2.2, frequency analysis techniques are being researched to effectively extract coefficients by increasing the order of frequency or time frequency transformation parameters. Higher Order Spectra like Bispectrum and Trispectrum were used for fault diagnosis in motor bearings [55–57]. Bispectrum along with Empirical Mode Decomposition for detecting outer race bearing defects has been utilized in [58]. A novel approach for bearing fault diagnostics, based on bispectrum features and support vector machine classifier (SVM) is presented in [59], whereas an approach that combines Bispectrum and fuzzy clustering method for rolling element bearing fault recognition is presented in [60] . A feature extraction method for bearing fault detection based on bispectrum contour maps involving dimensionality reduction by space alignment and identification by SVM and genetic optimisation algorithm is proposed in [61].

Researchers also proposed the use of composite spectrum for vibration-based condition monitoring [62–64]. Cyclostationarity is the second order of a frequency domain synchronised averaging method. The spectral correlation function derived from second-order Cyclostationarity is an efficient parameter for the early diagnosis of faults. The application of Cyclostationarity to early diagnosis of spalling in gear teeth demonstrated the power of this new parameter [65–67]. A comparison between Cyclostationarity and bilinearity was researched and presented for early diagnosis in gearboxes. The Bicoherence spectrum, which is a third-order spectrum, has also been used to monitor bearing condition [68].

2.3.3 Vibration Signature Analysis in Time-Frequency Domain

Time frequency analysis techniques have been studied and applied to machinery fault diagnosis due to their capability of representing signals in both the time and

frequency domains [69, 70]. This characteristic of time frequency analysis technique meets the requirements for analysing vibration signals which are non-stationary.

Preliminary time frequency analysis techniques, windowed Fourier transform and Short time Fourier transform (STFT) were applied to monitoring the condition of machinery. The Wigner distribution and the spectrogram are the most well-known quadratic time frequency representations belonging to the Cohen class which were applied to diagnosing gear faults. The use of the third- and fourth-order Wigner moment spectra, called the Wigner bi- and tri-spectra respectively was used to analyse the signals of rotating machinery [71, 72].

The continuous wavelet transform (CWT) has been developed based on the STFT with better time frequency resolution and applied to rotating machinery fault diagnosis. The Scalogram, the squared modulus of the CWT was applied in diagnosing gear faults. The Discrete wavelet transform (DWT) was also used to diagnose rolling bearing defects. Discrete wavelet packet analysis (DWPA) and discrete wavelet analysis also showed their potential in fault diagnosis [73–77].

2.4 VIBRATION SIGNATURE ANALYSIS USING ARTIFICIAL INTELLIGENCE TECHNIQUES

There are a variety of machine learning algorithms that have been used for intelligent fault diagnosis for many years. Reviews summarizing the use of artificial intelligence techniques on motor fault diagnostics can be found in [34], [78]. Liu et al. gave a comprehensive review of "Artificial algorithms" used in rotating machinery fault diagnosis [78]. They have discussed advantages and limitations of KNN, Naïve Bayes, SVM, ANN and Deep learning algorithms. Zhang et al. have presented a systematic review of "Machine learning and Deep Learning algorithms" for bearing fault diagnostics [79]. A review on "Meta Classification Algorithms" using WEKA is presented by Bal and Sharma [80].

Researchers were able to utilize the power of popular supervised machine learning techniques like Random Forest [32], k-nearest neighbour classifier and its variants [81–84], Support Vector Machine [39], [46] in the domain of fault diagnosis. The use of artificial neural networks (ANN) in fault detection was also explored widely [37], [85, 86].

The potential of Rule based classifiers, in the domain of bearing fault diagnosis, has also been shown to be promising [87, 88]. Application of Rule based classifiers has also been successfully extended for severity detection. Li et al. proposed an association rule mining method based on equal probability discretization to perform a 10 class prediction on select time-domain and frequency domain features for fault classification and severity prediction [88].

Researchers have also explored the use of Rough Set Theory and Analogy Based Reasoning for classification algorithms [89–91]. In machine learning, Analogy-based reasoning methods allow us to reason out and take decision about objects based on the similarities between them. Using analogy-based reasoning paradigm, A. Wojna modified distance metrics and used them to develop improved k-nearest neighbour classifiers [90]. Rule based and k-nearest neighbours classification algorithms based on rough set theory have been developed. C. Xin et al. performed bearing fault diagnosis by Kurtosis computation in time-domain and wavelet analysis in frequency domain [91]. The fault diagnosis algorithm was built by attribute reduction based on rough set. X. Zhu, Y. Zhang and Y. Zhu used statistical features and combined the kernel method and neighborhood rough sets to design the wrapper feature selection algorithm in their proposed methodology of bearing fault diagnosis [92]. J. Yan et al. reduced the fault data of rolling bearing by the application of greedy algorithm of rough set and used neural network for classification [93]. V. Muralidharan and V. Sugumaran used wavelet features with rough set and fuzzy logic for fault diagnosis of centrifugal pumps using vibration signals [94]. Rules were generated by applying Rough set theory and the faults are identified based on the strength of the rules. W. Li, W. Pan and S. Zhang proposed a rough set and back propagation neural network based algorithm for rolling bearings fault diagnosis [95]. It was seen that rough set was able to reduce the dimensionality of the raw feature set and rough sets combined with BP neural network effectively classified bearing faults. Hence studies proved that rough set based classifiers can be effectively utilized for bearing vibration signature analysis. [92–96].

As enormous amounts of vibration data are collected for bearing health monitoring, it becomes important to integrate different concepts for intelligent decision making for accurate fault diagnosis. Literature shows a growing research interest in combining a set

of learning algorithms to generate ensembles for investigating complex problems. These ensemble algorithms tend to exploit the strengths of the base classifiers to enhance the overall accuracy There have been numerous studies in the past decade for combining machine learning classifiers into an ensemble [97–100].

Kotsiantis et al. described ensembles of classifiers for improving classifier accuracy in their paper titled "Machine learning: A review of classification and combining techniques" [98]. Dietterich T. reviewed Ensemble Methods in Machine Learning and explained why ensemble classifiers often perform better than any single individual classifier [99]. An ensemble of rule-based classifiers for fault diagnosis of rotating machinery was proposed by Dou et al. to predict potential faults and subsequent breakdown of rotating machinery by using 6 time- domain and 5 frequency-domain features. A classifier ensemble was constructed and validated on the vibration data of two types of bearings; SKF6203 and NU205 [87]. Zio et al. proposed feature-based classifier ensembles for multiple fault diagnosis in rotating machinery. In their work, a multi-objective genetic algorithm is used for feature selection and ensembles of classifiers is developed to achieve higher accuracies. A voting technique is used to effectively combine the predictions of the base classifiers [97] . Sikder et al. pre-processed vibration data using FFT and then applied an ensemble learning method, Random Forest for bearing fault diagnosis [101]. The validation for the proposed scheme was done on the Case Western Reserve University (CWRU) dataset. Sharma, Amarnath and Kankar used 15 time domain, frequency domain and wavelet-based features in feature vector and applied ensemble techniques namely rotation forest and random subspace for fault diagnosis [100]. G. Xu et al. proposed a bearing fault diagnosis method based on deep convolutional neural network (CNN) and random forest (RF) ensemble learning. They generated two dimensional gray-scale images from one dimensional time domain vibration signals, extracted multi-level features using convolution neural network and used ensemble of multiple Random Forest classifiers for classification of faults [102]. Karimi and Jazayeri-Rad applied boosting methods to compare the fault diagnosis performances of single neural networks with two ensemble neural networks [103]. A detailed description of ensemble classifiers is given in the book titled "Combining Pattern Classifiers: Methods and Algorithms" by L. Kuncheva [104].

Researchers have also explored analysis of different types of vibration images for efficient fault diagnosis. In this approach, first, the 1-D vibration signal is transformed into a 2-D image which provides a basis for the image-based feature extraction and fault diagnosis. Research in this domain can be divided in two major parts:

1. Methodology to convert 1-D vibration signal into 2-D vibration image
2. Methodology to extract fault-sensitive features from vibration images and to use them for fault classification

Some of the studies that used different types of vibration images and classification techniques are as follows: Vibration Images and Scale Invariant Feature Transform (SIFT) [105], Vibration Spectrogram and Quaternion Invariable Moment [106], Spectrum images of vibration signals and 2D PCA [107], 2D image of vibration signal and Micro texture analysis plus KNN [108], 2D image of vibration signal based on recurrence plot (RP) technique and SIFT plus Probabilistic neural network (PNN) [109], Bispectrum of vibration signal and speeded up robust features (SURF) [110, 111], vibration images and Convolutional Neural Networks plus transfer learning [112], fractal features of bispectrum images using digital image processing and fractal theory in [113].

In the initial years, machine learning techniques were used to analyse vibration images for the purpose of fault diagnosis. With the advancements in deep learning, research in the field of fault diagnosis using vibration images has progressed tremendously. Conventional machine-learning techniques have a drawback of not being able to process signals in their raw form, thereby creating a need for manually extracting features from the signals. This requires domain expertise. On the contrary, in the case of Deep Learning methods (deep neural networks), feature spaces are automatically extracted. Figure 2.3. depicts the difference between conventional machine learning and deep learning process flow. Deep learning enabled fault diagnosis using images of vibration signals has been researched in recent years [114–116]. DL-based intelligent fault diagnosis techniques have been reviewed and summarized in [117–119]. These reviews play an emphasis on intelligent fault diagnosis of typical rotating machinery, including bearing, gear etc. Like in other domains, Deep Learning has invaded the domain of machine health monitoring too and has established its effectiveness [120–123] .

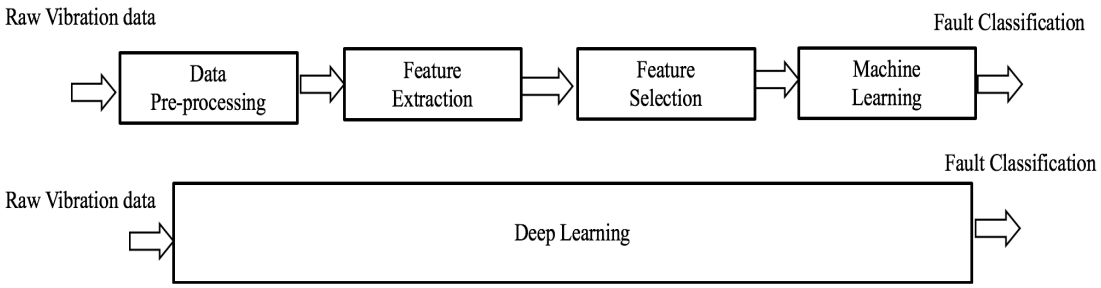


Figure 2.3: Machine Learning Vs Deep Learning techniques for Vibration Signature Analysis.

Deep learning can find complicated hidden patterns in large data sets by using the backpropagation algorithm to calculate its internal parameters that are used to compute the representation in each layer from the representation in the previous layer [124].

G. Xu et al. proposed an online fault diagnosis method, taking time-domain signal data from bearing and centrifugal pump, converting it to images and feeding them to a deep transfer CNN framework consisting of an online LeNet-5 and various offline CNNs [112]. Y.M. Hsueh et al. proposed a novel methodology for classifying a three-phase induction motor as faulty or healthy, involving transformation of the raw signal to grayscale images using pattern recognition followed by extraction of significant features using deep CNN [125]. Y. Xie and T. Zhang combined compressed features extracted from a CNN and EMD energy entropy extracted from first few Intrinsic Mode Functions, to train classification models for bearing fault detection [126]. D. Verstraete et al. generated image representations of bearing time-frequency data and used a deep CNN architecture for automatically learning the features for classification of faults [115].

Specifically, with bispectrum, Deep learning has shown its superiority in various domains other than condition monitoring. I. Mitiche et al. presented an efficient method for High Voltage power plant equipment monitoring through the application of a Deep Residual Neural Network (ResNet) to the Bispectrum images generated from discharge signals [127]. M. Sohaib and J.M. Kim used CNN based on AdaMa optimisation for bispectrum analysis, in order to detect bearing faults at different shaft speed variations with cracks of various scales [128]. The tremendous progress in deep learning and images have popularized the use of Bispectrum images in other domains as well. The application of deep neural networks in the enhancement of a bispectrum

phase retrieval algorithm for astronomical imaging has been explored in [129]. Bispectrum analysis to forecast epileptic seizures has been explored in [130, 131], wherein they have applied Long Short-Term Memory (LSTM) and 5-layer multilayer perceptron classifier, respectively.

After the detailed literature survey on condition monitoring and fault diagnosis using vibration signature analysis, it can be concluded that a vibration-based condition monitoring system fundamentally includes data acquisition, signal processing and decision making modules. Standard vibration data acquisition systems consist of pairs of accelerometers (piezo-electric transducers) mounted on each bearing of the rotating shaft of the equipment, at right angles from each other. The signal processing module includes data pre-processing methods and algorithms for feature extraction. The decision making module includes ML/DL based classification for fault detection. These basic building blocks of vibration based condition monitoring are shown below in the Figure 2.4.

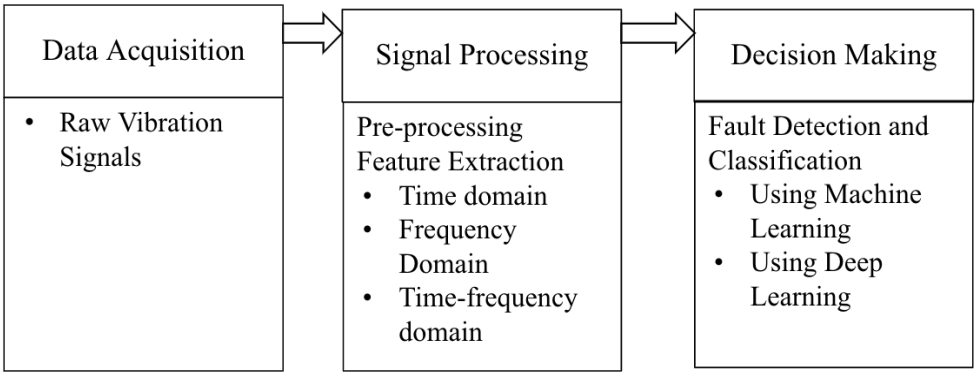


Figure 2.4: Block diagram of Vibration Signature Analysis for Fault Diagnosis

2.5 CHAPTER SUMMARY

This chapter presented a review of condition monitoring of rotating machinery of power plant such as steam turbine, gas turbine, generator and major auxiliaries like pumps. Fault sensitive features of vibration signals from different domains have been reviewed and research gaps have been identified. Review of machine learning techniques give a direction for designing of Ensemble Classifiers.

Furthermore vibration signature analysis using image processing have also been reviewed to understand a rapidly growing research direction of fault diagnosis using deep learning.

The detailed review of existing body of literature leads to following conclusions:

1. From the literature review of condition monitoring and fault diagnosis of Power-plant components, it is concluded that one of the key and universal component of a rotating machinery is a bearing which is usually the cause of failure of the machinery and hence will be a key research area for fault diagnosis in this research work.
2. The study on the research done on fault sensitive features, helped in understanding that although all of the studies cited above have applied various time-domain parameters, whether conventional or novel and dimensional or dimensionless, for bearing fault diagnosis, either exclusively or in addition to other frequency and time-frequency domain features, none of the studies discusses and provides the reason behind the use of a set of particular statistical features. Majority of the previous studies, till now, have focused more on optimising fault diagnosis classifiers, very few have compared established features with novel ones. Since, fault diagnostic accuracy depends on the choice of both features and classifier, it is important to emphasise on features too by proposing novel statistical features and/or developing a strategy to conduct an extensive analysis of new and old statistical features, so as to derive the most optimal set of fault-sensitive features.
3. Furthermore, it is also learnt that frequency-domain or spectral analysis of the vibration signals is becoming a highly effective approach for bearing defect detection. Higher order spectral analysis methods like Bispectrum analysis are becoming very popular, over traditional methods like power spectrum, in fault diagnosis of rolling element bearings since higher order spectrum can provide more diagnostic information than power spectrum for vibration signals. In particular, Bispectrum, the 3rd order cumulant, helps retain phase and amplitude information of vibration signals.
4. Literature review of machine learning techniques show that Ensembles methods have been applied to a wide range of industrial problems in the area of condition monitoring. There is empirical evidence of the effectiveness of this approach in fault diagnosis of rotating machines.
5. From the literature review of vibration signature analysis using image processing techniques, it is learnt that intelligent fault diagnosis based on deep learning has become a rapidly growing research domain, and is redefining state-of-the-art

algorithms in the area of machine health monitoring. Deep learning methods involve multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level. With the generation of enough such transformations, very complex features can be learned. This is a key advantage of deep learning.

6. Even though considerable prior research has been conducted on vibration-based fault diagnosis of rotating machinery components, most of the research conducted so far has been tested on simulated data from laboratory setups and not on real industrial system data. As laboratory setups are different from industrial equipment with regards to scale and operating conditions, there is no guarantee that research solutions validated with simulated data will perform equally well on actual systems. So while conducting experiments the validity of the proposed solution or algorithm is best confirmed when it's tested on data from real industrial systems.

CHAPTER III
VIBRATION DATA ACQUISITION

3.1 INTRODUCTION

Data is the foundation for artificial intelligence based fault diagnosis of rotating machines. Since the degradation of a rotating machine may take many years, the real data collection can be time consuming. Hence sometimes researchers generate artificial vibration signals to test their vibration signature analysis methodologies. Researchers working in the field of Machine Condition Monitoring conduct experiments in the vibration research labs and collect data using components with artificially seeded faults. Researchers from few organizations have put efforts in vibration data collection and made the fault dataset repositories publically available. These datasets also serve as standards for the comparison and validation of different algorithms. Neupane, D. and Seok, J. have discussed many open source and closed source bearing datasets in their review paper of Fault Detection and Diagnosis using Deep Learning and Case Western Reserve University dataset [132].

This chapter explains Case Western Reserve University (CWRU) dataset [133] and Machinery Failure Prevention Technology (MFPT) dataset [134] used in this research work along with their experimental setups for data acquisition.

3.2 FAULT DATA ACQUISITION FROM VIBRATION RESEARCH LAB

This data was acquired by collecting Vibration Signatures of a faulty machine by controlled experiments on a test rig in the "Vibration Research Lab" of the "Department of Mechanical Engineering, IIT Delhi" under **"Summer Faculty Research Fellowship Programme-2013"**.

3.2.1 Experimental Setup for Data Acquisition

To gain comprehensive knowledge of vibration signatures related to different types of faults, experiments in a controlled environment were performed on a test rig that mirrors real world machinery. SpectraQuest's Machinery Fault Simulator (MFS) shown in Figure 3.1 is the test rig used for performing these experiments for learning

the vibration signatures of different types of machine faults [135].

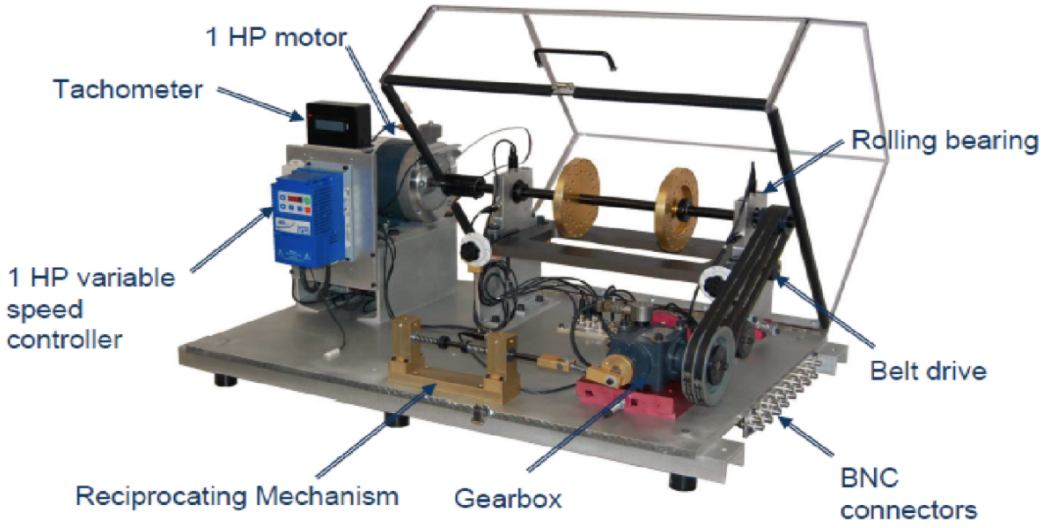


Figure 3.1: SpectraQuest's Machinery Fault Simulator (MFS)

MFS is a bench-top test rig whose electrical and mechanical specifications are given in Table 3.1. The experiments were performed on MFS by mounting two Brüel Kjær Piezoelectric Charge Accelerometer (Type 4371), one at each end. The two channels' acceleration data was collected simultaneously using a FFT analyzer.

3.2.2 Dataset and Experiment Description

This dataset is composed of multivariate time-series vibration data acquired by sensors hooked onto the MFS Rig. The dataset comprises following four types of vibration signals:

1. **Normal (no fault):** The MFS was carefully aligned to its baseline configuration and the baseline data was collected at 25 Hz.
2. **Misalignment faults (Angular and Parallel):** After collecting baseline data, the inboard bearing house was misaligned intentionally by 300 microns in the vertical direction to all six locations by inserting 300 micron shims. Thus a parallel misalignment was introduced. Vibration signatures were recorded for 30 Hz, 35 Hz, 40 Hz and 45 Hz in both the channels. After the acquisition of parallel misalignment, the outboard bearing house was misaligned by 200 microns in the vertical direction to the operator's side. Therefore, an angular misalignment was introduced. Again, the data was collected under the same rotating speeds. The

Table 3.1: Specifications of Machinery Fault Simulator

Motor	3 Phase, 1 HP motor, pre-wired self-aligning mounting system for easy installation/removal
Tachometer	Built-in tachometer with LCD display and one pulse per revolution
Drive	1 HP variable frequency AC drive with multi-featured front panel programmable controller
RPM range	0 to 6000 rpm (short duration) variable speed
Voltage	115/230 VAC, Single phase, 60/50 Hz
Current Measurement	Power leads accessible for current measurements
Shaft Diameter	3/4" diameter; Turned, Ground, & Polished (TGP) steel
Bearing	Two rolling element ball bearings with squeeze lock type. Split bracket bearing housings with five mounting positions for shaft span reconfiguration
Rotor Base	18" long, completely movable using jack bolts for easy horizontal misalignment and standard shims for vertical misalignment. Pinned for easy realignment.
Rotors	Two 6" balance rotors with two rows of holes
Belt Mechanism	Two double groove "V" belt with one set screw mounting and one bush/key mounting
Reciprocating Mechanism	Adjustable spring engagement timing and two stroke settings.
Gearbox and Brake	Accessible three-way straight cut bevel gearbox with 1.5:1 ratio (20 gear input).
Centrifugal Pump	1⁄2 HP, 27psi at 0gpm, 25gpm at 0psi with water at 4000rpm
Instrumentation Connectors	16 BNC connector plate under the rotor base linked to BNC connector panel mounted on the edge for the base plate for direct connection to data collectors
Safety Cover	Lockable clear, impact resistant hinged plastic cover with motor interlock switch to shut down motor when cover is raised
Foundation	1/2" die cast aluminium base, base stiffener and eight rubber Vibration isolators

acceleration data was exported to a FFT analyzer to determine the effect of misalignment on Spectrum. The raw data was exported to MATLAB for further processing. The change in vibration levels with machinery alignment condition was investigated and it was observed that misalignment produces several harmonics of the running speed.

3. **Imbalance fault:** Seeded fault data with load values within the range from 5g to 25g was recorded. The number of sequences acquired per weight were 10. The effect of mechanical imbalance on the machine vibration was studied and it was noticed that the level of vibration increased significantly in an unbalanced machine.
4. **Bearing faults (inner race, outer race and ball-bearing):** Rolling element bearings are the most susceptible elements to fault occurrence in rotating machines. To investigate the vibration signature of faulty bearings, three defective bearings with bearing faults (inner race, outer race and ball-bearing) were placed one at a time in two different positions: a) bearing between the rotor and the motor and b) rotor between the bearing and the motor. Vibration signals were measured for each case and their Kurtosis values were calculated. Bearing faults were practically undetectable in the absence of imbalance. So, small masses were added to induce a detectable effect, with different rotation frequencies as before.

3.3 FAULT DATA ACQUISITION FROM CWRU AND MFPT REPOSITORIES

3.3.1 CWRU Experimental Setup and Dataset Description

Case Western Reserve University (CWRU) Bearing Data Set [133] has been used in this research. The setup for vibration data acquisition, shown in Figure 3.2. and Figure 3.3., consists of a 2 HP Electric motor, transducer/encoder, dynamometer and accelerometers. Specifications of CWRU Test rig and experimental setup is listed in Table 3.2. Data for normal bearings is recorded using the drive-end bearing SKF6205 for motor loads of 0, 1, 2, 3 horsepower (motor speeds of 1797, 1772, 1750, 1730 RPM respectively). Then faults are introduced in the bearings using Electro-discharge machining. Faults of diameters: 0.007 inches, 0.014 inches, 0.021 inches are induced in the SKF6205 bearing and 0.028 inches in its NTN equivalent. Different parts of deep groove ball bearing used in the rig are shown in Figure 3.4.

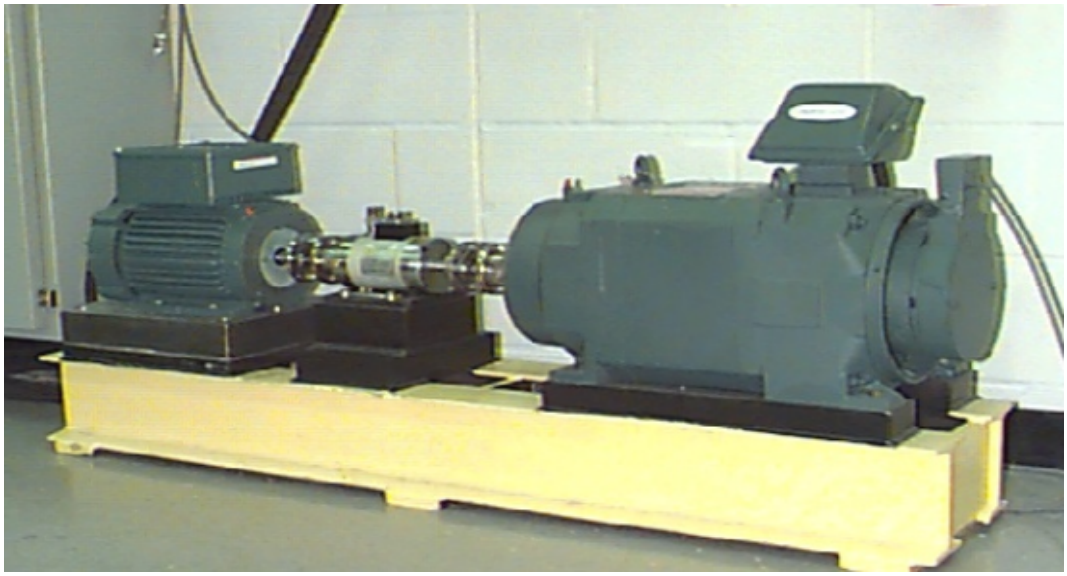


Figure 3.2: CWRU test rig

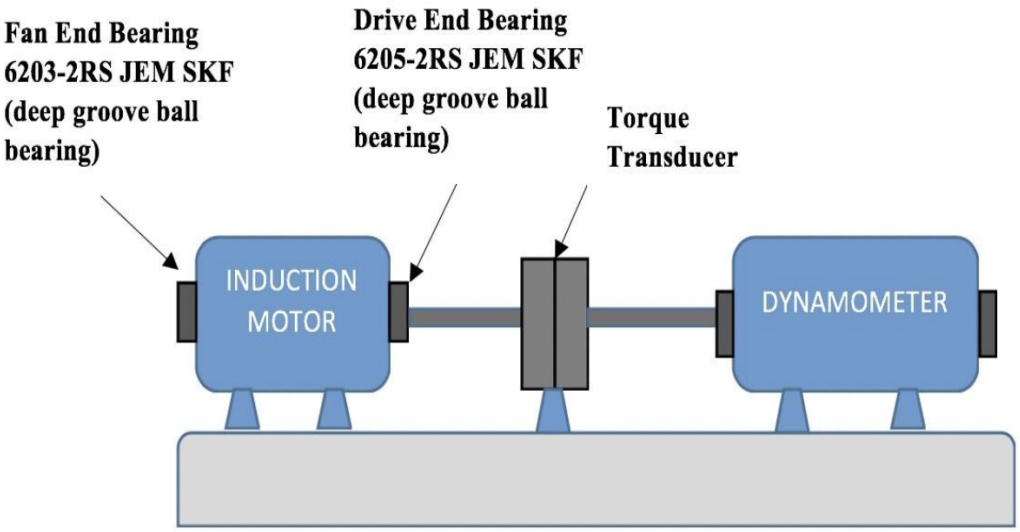


Figure 3.3: Schematic diagram of CWRU experimental setup

Figure 3.4: Parts of 6205 SKF deep groove ball bearing

Table 3.2: Specifications of CWRU test rig and experimental setup.

Bearing specification							
Position	Number	Type	Inside Diameter	Outside Diameter	Thickness	Ball Diameter	Pitch Diameter
Drive end bearing	6205-2RS JEM SKF	deep groove ball bearing	0.9843 inch	2.0472 inch	0.5906 inch	0.3126 inch	1.537 inch
Fan end bearing	6203-2RS JEM SKF	deep groove ball bearing	0.6693 inch	1.5748 inch	0.4724 inch	0.2656 inch	1.122 inch
Sampling Frequency							
Drive end	12000 samples/sec and 48000 samples/sec						
Fan end	48000 samples/sec						
Motor Speed							
1797 RPM, 1772 RPM, 1750 RPM, 1730 RPM							
Motor Load							
0 HP, 1 HP, 2 HP, 3 HP							
Fault Types							
Inner Race, Outer Race, Ball Bearing							
Fault Size Diameter							
0.007 inch, 0.014 inch, 0.021 inch, 0.028 inch							
Data Files Used							
Training	13648						
Test	3412						
Total	17060						

After reinstalling these faulted bearings into the test motor, vibration data is again recorded for same motor loads as in the case of normal bearings. Four cases are considered: outer raceway fault, inner raceway fault, rolling element (ball) fault, and the normal (no fault). The defect of outer raceway fault is located at 6 o'clock (orthogonal to the load zone). Vibration data is collected using a 16 channel DAT recorder for single-point drive end defects at 12,000 samples/second. Rotation rate and horsepower data are gathered through a torque transducer/encoder.

Data samples of size 1024 points were extracted from these vibration signals in a linear non-overlapping fashion, i.e. for each signal, the first 1024 points were extracted as its first data sample, the next 1024 points as its second data sample and so on. A total of 17060 samples consisting of 3314 normal, 4735 inner race, 4272 outer race and 4739 ball bearing samples were extracted. Using these samples 3 datasets were prepared.

Table 3.3: CWRU dataset A for Fault Type Detection.

Class	**No of Train samples**	**No of Test samples**	**Total**
Normal	2649	665	3314
Inner Race	3774	961	4735
Outer Race	3473	799	4272
Ball Bearing	3752	987	4739
Total	13648	3412	17060

Table 3.4: CWRU dataset B for Fault Severity Detection.

Class	**Fault Type**	**Fault size**	**No of Train Samples**	**No of Test Samples**	**Total**
Normal	–	–	2663	651	3314
IR 007	Inner Race	0.007”	1135	292	1427
IR 014	Inner Race	0.014”	1139	277	1416
IR 021	Inner Race	0.021”	1152	269	1421
IR 028	Inner Race	0.028”	373	98	471
OR 007	Outer Race	0.007”	1158	267	1425
OR 014	Outer Race	0.014”	1126	296	1422
OR 021	Outer Race	0.021”	1134	291	1425
BB 007	Ball Bearing	0.007”	1131	288	1419
BB 014	Ball Bearing	0.014”	1136	289	1425
BB 021	Ball Bearing	0.021”	1125	299	1424
BB 028	Ball Bearing	0.028”	376	95	471
Total			13648	3412	17060

Table 3.5: CWRU dataset C for Fault Diagnosis using Bispectrum Images.

Class (**Fault Diameter in inches**)	**Ball Bearing**		**Inner Race**		**Outer Race**		**Normal**	**Total**
	0.007	**0.014**	**0.007**	**0.014**	**0.007**	**0.014**		
Train set	331	332	333	330	332	332	1160	3150
Validation set	71	72	72	71	72	71	249	678
Test set	71	71	71	71	71	71	248	674

The three datasets are: **Dataset A** for Fault Type Detection listed in Table 3.3, **Dataset B** for Fault Severity Detection listed in Table 3.4 and **Dataset C** listed in Table 3.5 for Fault Diagnosis using Bispectrum images. Dataset A has data labelled into 4 classes- outer raceway defect, inner raceway defect, ball bearing defect, and normal whereas Dataset B, has the data labelled into 12 classes defined by fault types and fault sizes (in inches). Dataset C has 7 classes - Normal class, Inner Race with 0.007 inch and 0.014 inch, Outer Race with 0.007 inch and 0.014 inch, and finally Ball Bearing with 0.007 inch and 0.014 inch.

3.3.2 MFPT Experimental Setup and Dataset Description

The second repository dataset used in this research is the dataset provided by the Machinery Failure Prevention Technology (MFPT) Society [134]. The MFPT data was acquired from a NICE bearing [136] as shown in Figure 3.5. It has roller diameter of 5.97 mm, pitch diameter of 31.62 mm, contact angle of 0 degree and an input shaft rate of 25 Hz. A single radial accelerometer has been used to obtain the data. The acquired data is stored in a MATLAB® double-precision, binary format *.mat file. In addition to acceleration data, the data files also include sampling rate, Motor Speed and Motor Load as shown in Table 3.6.

The dataset consists of data files collected from a bearing test rig and from real machines. Data collected from test rig includes 3 files of healthy bearing under fixed load, 3 files with vibration signals from bearing having outer race faults under fixed load, 7 files with vibration signals from bearing having outer race faults under seven types of loads and 7 files with vibration signals from bearing having inner race faults under seven types of loads.

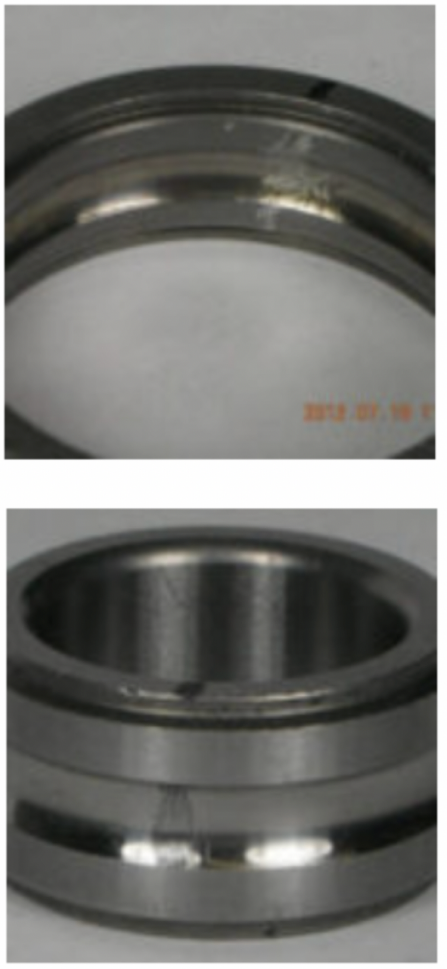


Figure 3.5: Bearing used in MFPT experimental setup

Data related to following three conditions has been used for this research work:

- Baseline (No Fault): 3 baseline or healthy conditions with a sample rate of 97,656 Hz and 270 lbs of load recorded for 6 sec.
- Fault in Outer Race: 7 outer race fault conditions with a sample rate of 48,828 Hz and various loads of 25, 50, 100, 150, 200, 250, 300 lbs recorded for 3 sec.
- Fault in Inner Race: 7 inner race fault conditions with sample rate of 48,828 Hz. and various loads of 0, 50, 100, 150, 200, 250, 300 lbs recorded for 3 sec.

In order to match the sample rate of other fault sets, the baseline data set was down-sampled to 48,828 Hz. The original vibration signals were split into pieces such that each file contains 2048 points. Vibration data after segmentation include 498 files of Inner Race Faults, 498 files of Outer Race Faults and 430 files of Normal bearing signals as listed in Table 3.7.

Table 3.6: Specifications of MFPT experimental setup.

Bearing Specification	
Bearing type	NICE Ball bearing
Roller diameter: rd	0.235
Pitch diameter: pd	1.245
Number of elements: ne	8
Contact angle: ca	0
Motor Speed	
25 Hz or 1500 RPM	
Motor Load	
0, 25, 50, 100, 150, 200, 250 , 270 and 300 pounds	
Sampling Frequency	
48,828 Hz, 97,656 Hz	
Fault Types	
Inner Race, Outer Race	
Data file format	
MATLAB double-precision, binary format *.mat file	
Data Files Used	
Baseline or healthy condition	3 with a sample rate of 97,656 Hz, 270 lbs of load, 6 sec. of duration
Fault in Outer Race	7 with a sample rate of 48,828 Hz, various loads of 25, 50, 100, 150, 200, 250, 300 lbs, 3 sec. of duration
Fault in Inner Race	7 with a sample rate of 48,828 Hz, various loads of 0, 50, 100, 150, 200, 250, 300 lbs, 3 sec. of duration

Table 3.7: MFPT dataset

Class	**Bearing Condition**	**Number of Train Samples**	**Number of Test Samples**	**Total**
Normal	No fault	343	86	429
IR	Inner Race fault	398	99	497
OR	Outer Race fault	398	99	497
Total		1139	284	1423

3.4 FAULT DATA ACQUISITION FROM POWER PLANT

The vibration data for validation of the proposed methodology was taken from a motor of an operational power plant. Data was acquired from Rolling Element Bearing of an electric motor of service water pump of 500 MW Kosti Thermal Power Plant commissioned by Bharat Heavy Electricals Limited, India.

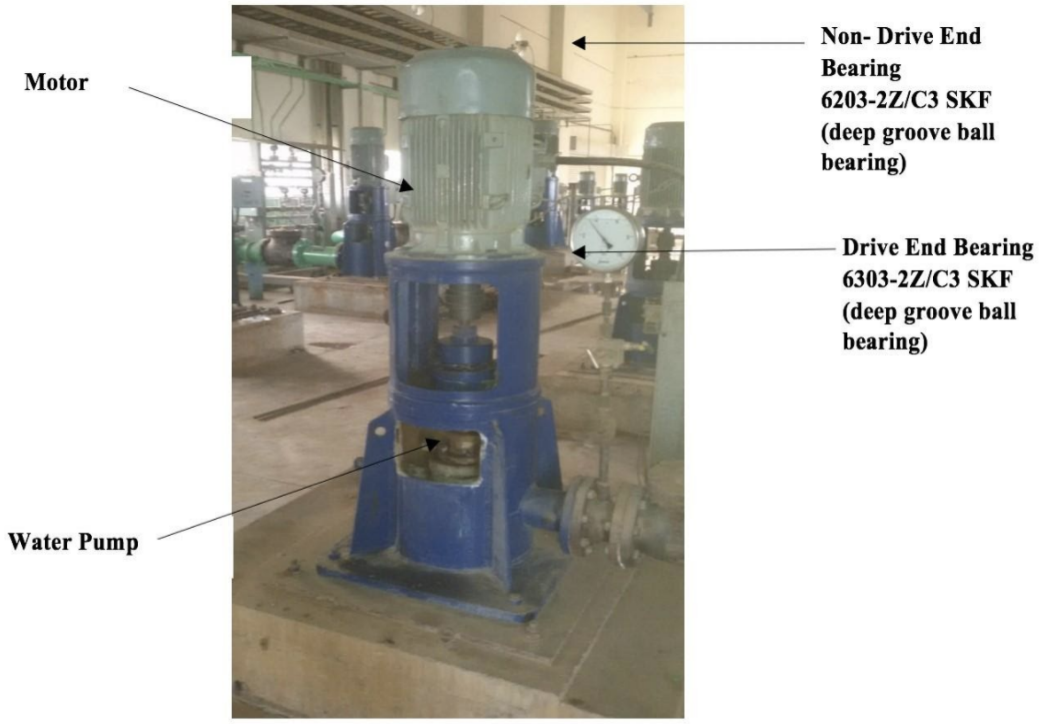


Figure 3.6: Experimental setup consisting of vertical electric motor of service water pump at Kosti Thermal Power Plant.

The setup, shown in Figure 3.6., consists of a 7 HP vertical electric motor of service water pump of Power-plant. Data for normal baseline bearings was recorded using the deep groove ball bearing SKF6303-2Z/C3 with dimensions 17mm X 47mm X 14mm at drive-end with motor speed of 3000 rpm. Vibration data was collected at 12,000 samples/second. Vibration data was taken for 2 different timestamps and for both vertical and horizontal positions, leading to a total of 4 vibration signals.

3.5 CHAPTER SUMMARY

This chapter discusses the acquisition of vibration data used in this research. Data acquired from Vibration labs and real Power-plant is explained. Specifications of the experimental setup for each experimental study is elaborated.

CHAPTER IV

PRE-PROCESSING OF VIBRATION SIGNALS

4.1 INTRODUCTION

Vibration signals collected from sensors are often contaminated by noise and it is difficult to diagnose machine faults, if these signals are used in raw form. Faults can go undetected without pre-processing of vibration signals, hence effective pre-processing techniques are needed to increase the fault diagnosis accuracy at later stages. For example, when a rolling bearing operates, vibration signals captured from it exhibit modulation. Therefore, different methods for analyzing the demodulation of rolling bearing signals have been established, envelope demodulation method being a popular choice. The limitation of the conventional envelope analysis is that it requires a proper filtering band to be chosen in order to obtain reliable results since the selection of the filtering band has a great influence on the accuracy of the analysis result.

To overcome this limitation, many researchers have adopted the Empirical Mode Decomposition (EMD) method that treats the signals as "fast oscillations superimposed on slow oscillations" and decomposes the parent signal into a number of Intrinsic Mode Functions (IMFs). These IMFs are time-varying mono-component (single frequency) functions, which can be regarded as a self-adaptive filter whose bandwidth and central frequency change with the signal itself. Since the decomposition is based on the local characteristic time scale of the data, EMD is an adaptive data-driven method and therefore it is especially suitable for the analysis of non-linear and non-stationary signals such as bearing vibration signals.

Empirical mode decomposition (EMD) has been extensively studied and widely applied in fault diagnosis of rotating machinery [31], [137–139]. An extensive review on empirical mode decomposition in fault diagnosis of rotating machinery is available in [140]. Variants of EMD, like Bi-variate EMD and Ensemble EMD for multi-fault diagnosis are also found in the literature [141–144]. It is important to note that not all the constituent signals, that is IMFs, generated by EMD, bear fault information. Different criteria, such as - correlation coefficient and kurtosis - have been used to select significant IMFs [48], [35].

4.2 EMPIRICAL MODE DECOMPOSITION

Empirical Mode Decomposition (EMD) was proposed by Huang et al. as a mathematical tool to analyze a non-stationary and non-linear signal by decomposing it into different IMFs [137].

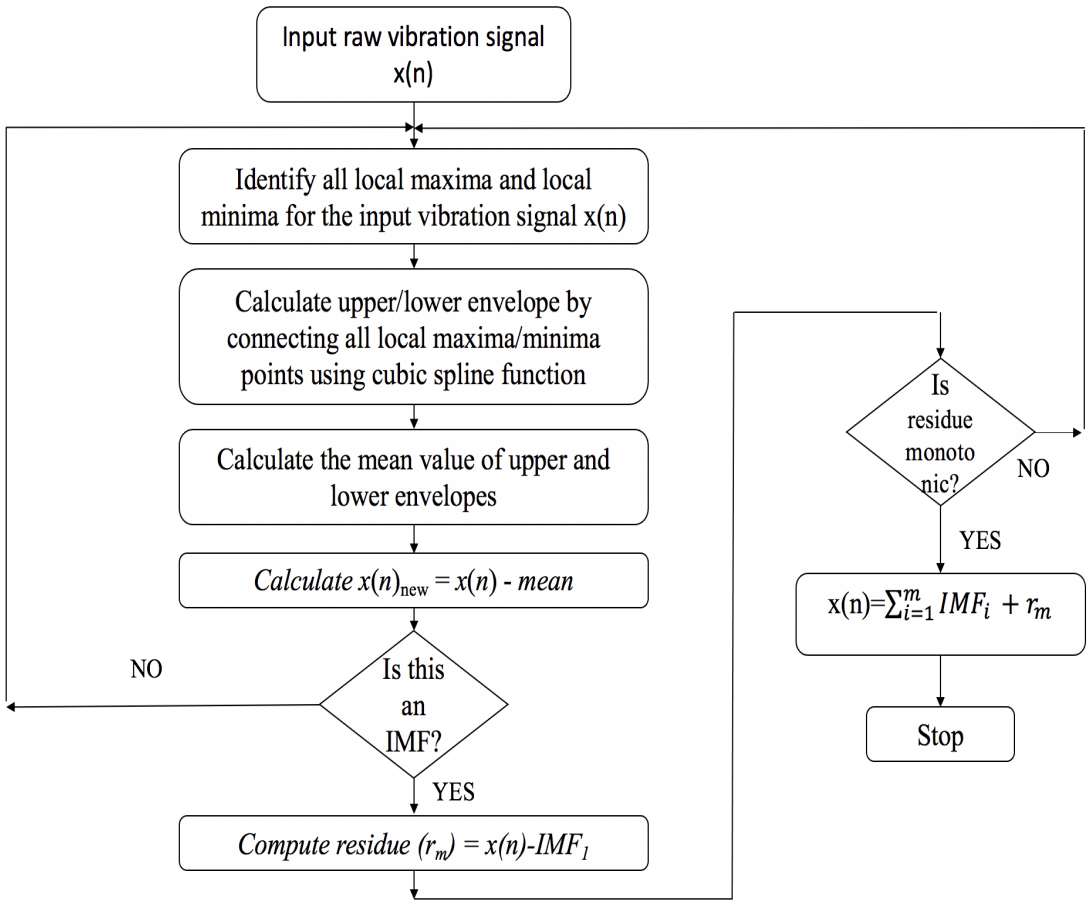


Figure 4.1: Flowchart of Empirical Mode Decomposition (EMD)

The EMD Algorithm is shown as a flow-chart in Figure 4.1. and its steps are explained below:

Step 1: Identification of all local maxima and local minima for the input vibration $x(n)$ signal.

Step 2: Connecting all local maxima points using cubic spline function and calculating upper envelope.

Step 3: Connecting all local minima points using cubic spline function and calculating lower envelope.

Step 4: Calculating the mean value of upper and lower envelopes.

Step 5: Updating the signal by subtracting the mean calculated in step 4 by using the

formula $x(n)_{new} = x(n) - mean$ and repeating steps 1 to 5 on the updated signal until it can be considered as an IMF i.e. the difference between the number of extrema and the number of zero crossings is at most one and the mean value of the both of the envelopes is zero.

Step 6: Subtracting first IMF from $x(n)$ to obtain residue (r_m) i.e. $residue = x(n) - IMF_1$

Step 7: Iterating the above steps on the residue to find all the IMFs of the vibration signal; the algorithm will terminate when the residue becomes a monotonic function. Thus, after Empirical mode decomposition, the vibration signal x(n) can be represented as a sum of IMFs and residue:

$$x(n) = \sum_{i=1}^{m} IMF_i + r_m \tag{4.1}$$

4.3 BEST INTRINSIC MODE FUNCTION SELECTION

Although the EMD process decomposes raw vibration signal into a number of IMFs, the main fault information is often contained in only a few IMFs. If all the IMFs are considered, the dataset will increase considerably and some IMFs might serve as noise affecting the classification analysis. Research studies also show that IMF selection alleviates the problems of mode mixing and spurious modes that might arise because of working with insignificant IMFs. That's why researchers have used different criteria to select IMFs in their proposed methodologies for designing an efficient fault diagnosis system [145, 146]. Some of them are: cross correlation, kurtosis, energy, power-harmonic ratio etc. [145, 147, 148].

In this research, the best IMFs were selected based on Pearson correlation coefficient. The IMF having the largest correlation coefficient with the original signal is selected as the best IMF since it retains most of the information of the original signal, thereby ensuring that the features extracted from the selected IMF accurately represent the parent signal. The Pearson correlation coefficient $\rho_{X,Y}$ was computed as:

$$\rho_{X,Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \tag{4.2}$$

where E is the expectation, σ_X and σ_Y are standard deviations of X and Y, μ_X and μ_Y are means of X and Y respectively.

4.4 RESULTS AND DISCUSSION

Vibration signals were decomposed by EMD resulting in 12-17 IMFs along with residue for each signal. Eight out of seventeen IMFs generated by Empirical Mode Decomposition for sample vibration signal with Inner Race Fault of 0.007 inches are shown in the Figure 4.2.

The IMF having the highest Pearson correlation $\rho_{X,Y}$ with the parent signal was taken as its representative signal. The larger the value of $\rho_{X,Y}$, the stronger the correlation between two signals. Table 4.1 shows the calculated values of $\rho_{X,Y}$ for 12 IMFs of parent signals taken from the classes: Inner race 0.007 inch fault (IR007), Inner race 0.021 inch fault (IR021), Ball bearing 0.007 inch fault (BB007) and Ball bearing 0.021 inch fault (BB021). Plots of the best IMFs selected on the basis of Pearson correlation coefficient for each of the 12 classes from CWRU Dataset B are shown in Figure 4.3. These selected best IMFs will be used later for fault sensitive feature extraction.

Table 4.1: Pearson Correlation Coefficient between a raw vibration signal and each of its IMFs.

IMF No.	**Pearson Correlation Coefficient**			
	IR 007	**IR 021**	**BB 007**	**BB 021**
1	0.7732	0.909	0.9018	0.8925
2	0.679	0.7286	0.6647	0.6375
3	0.5711	0.5134	0.5282	0.5286
4	0.6099	0.4344	0.5131	0.5046
5	0.6174	0.674	0.6324	0.6102
6	0.6927	0.8245	0.7028	0.6432
7	0.722	0.7871	0.6702	0.5587
8	0.6656	0.6234	0.6691	0.5329
9	0.6563	0.0898	0.7705	0.5329
10	0.677	0.5651	0.5731	0.392
11	0.5092	0.7159	0.6447	0.247
12	0.2464	0.5056	0.6099	0.1647

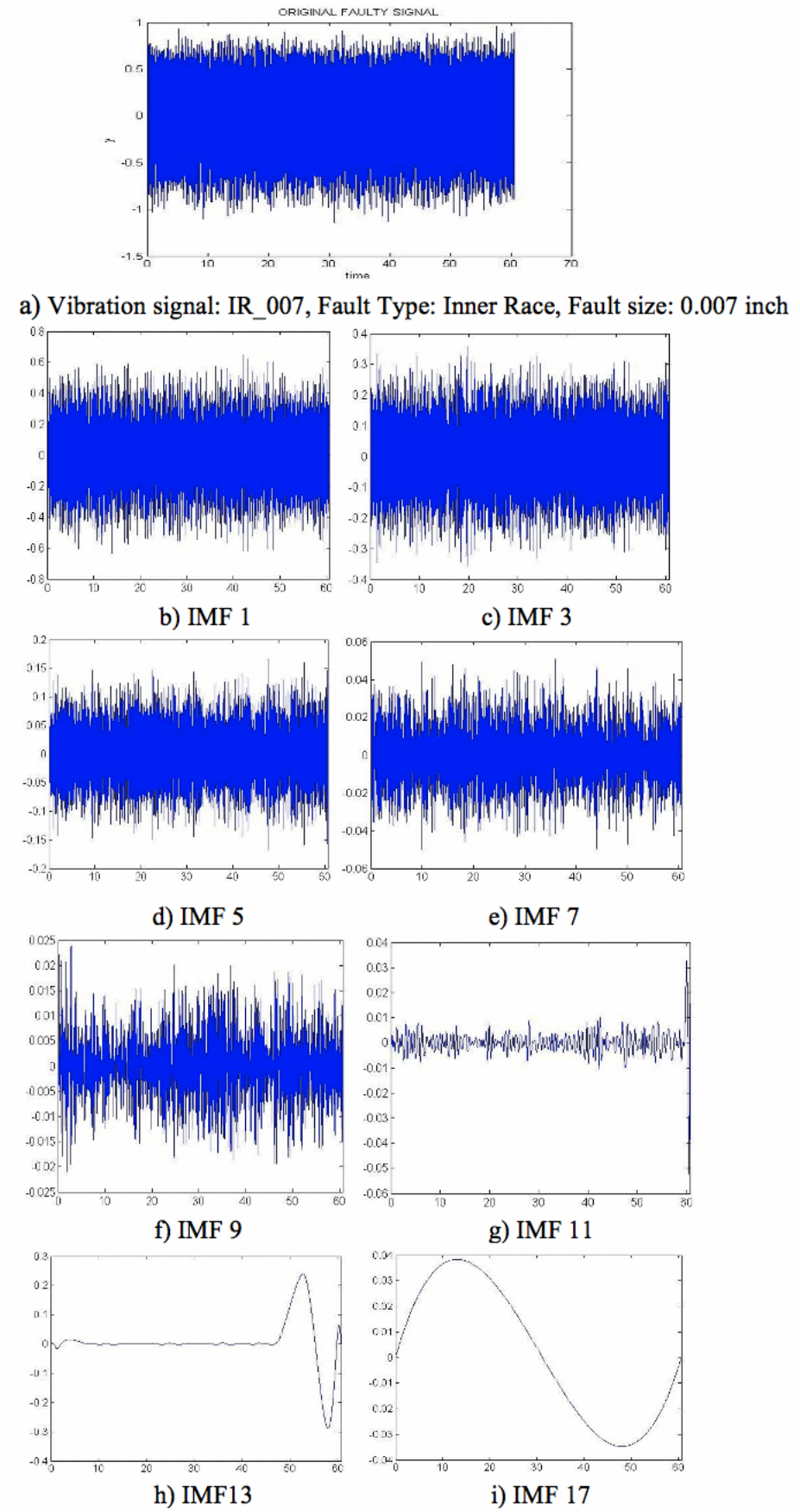
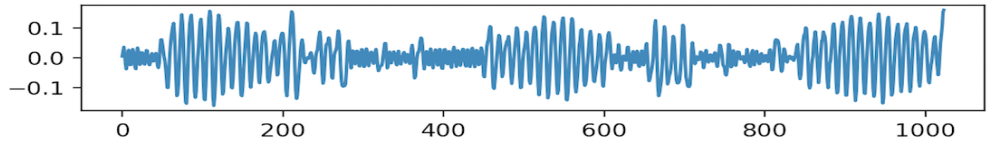
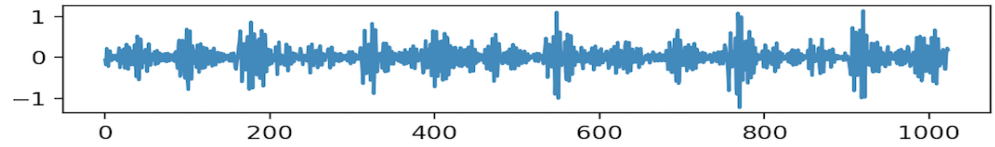


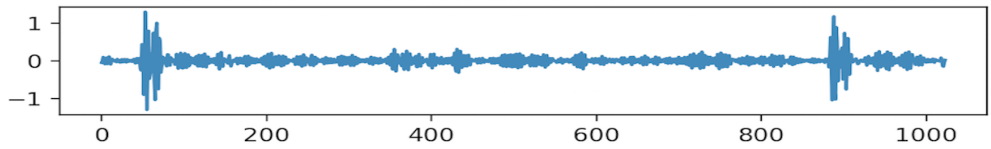
Figure 4.2: Plots of vibration signal with 0.007" Inner Race fault (x axis is time & y axis is voltage), shown in a); and its Intrinsic Mode Functions obtained after EMD, shown in b) to i).



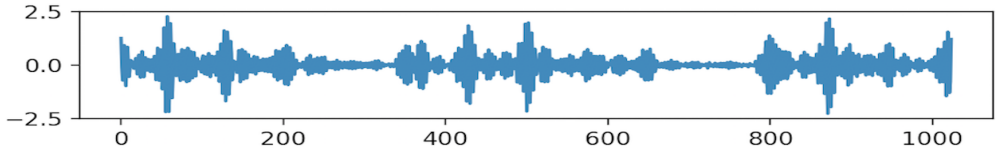
(a) Normal



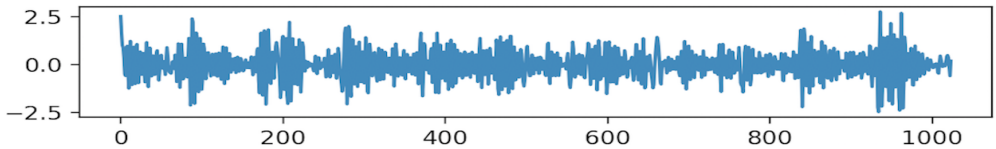
(b) IR 007



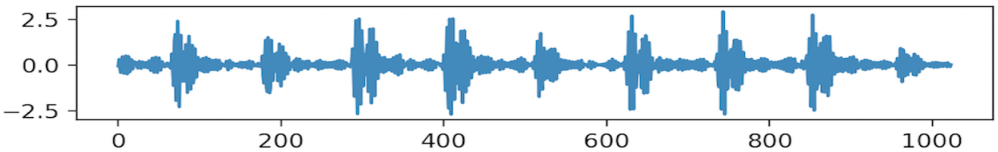
(c) IR 014



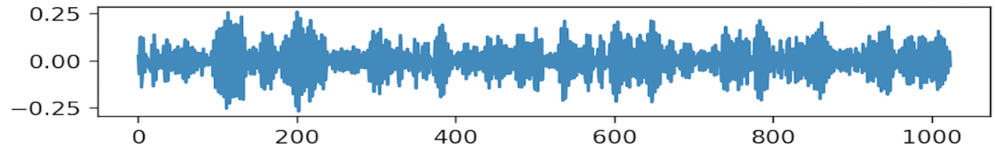
(d) IR 021



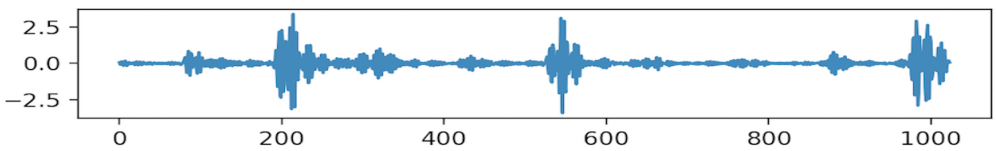
(e) IR 028



(f) OR 007

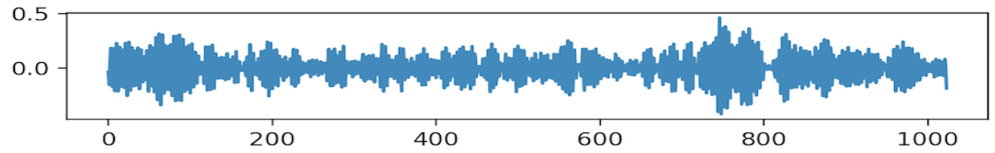


(g) OR 014

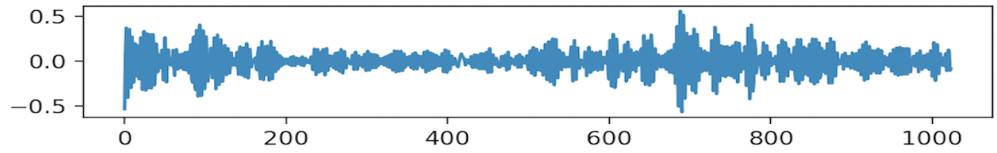


(h) OR 021

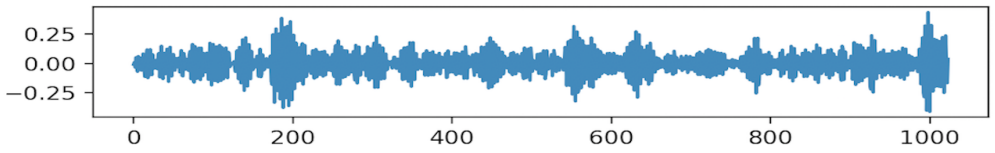
Figure 4.3: Plots of the best IMFs selected on the basis of Pearson Correlation Coefficient for each of the 12 classes from CWRU Dataset B.



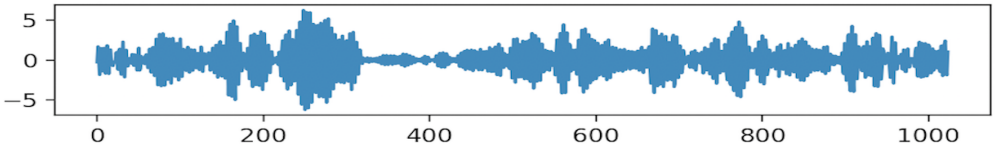
(i) BB 007



(j) BB 014



(k) BB 021



(l) BB 028

Figure 4.3: Plots of the best IMFs selected on the basis of Pearson Correlation Coefficient for each of the 12 classes from CWRU Dataset B. (Contd.)

4.5 CHAPTER SUMMARY

This chapter deals with the vibration signal pre-processing techniques. Empirical mode decomposition (EMD) is one of the most powerful signal processing techniques and has been extensively studied and widely applied in fault diagnosis of rotating machinery. The results of EMD and best Intrinsic Mode Function (IMF) selection out of the many IMFs generated by EMD process are presented in this chapter.

CHAPTER V

FEATURE EXTRACTION FROM VIBRATION SIGNALS

5.1 INTRODUCTION

In vibration based condition monitoring of power plant components, a variety of vibration feature extraction techniques (in time domain, frequency domain, and time-frequency domain) have been proposed during the past few decades to extract useful diagnostic information from raw or pre-processed vibration signals. Effective and efficient feature extraction techniques are critical for reliable fault diagnosis using vibration signature analysis. These techniques can locate certain components in signals to assist detection of machine faults. The features extracted from these techniques can be useful in pinpointing the source of abnormalities in vibration and thus these techniques have been successfully applied to turbine-generators, pumps, and fans in power-plants to prevent trip-out and fatigue induced damage.

Literature review in Chapter 2 proves that statistical time domain features have been effectively used by many researchers in their extracted feature set along with various other features [34], [37], [149]. This chapter discusses definition, significance and values of 31 statistical time domain features extracted from best Intrinsic Mode Function obtained by Empirical Mode Decomposition of bearing vibration signals for various fault conditions in Chapter 4. Out of these 31 features, 4 features are novel features proposed in this research for effective fault diagnosis. This chapter also discusses concept and results of Higher Order Spectral Feature "Bispectrum", extracted from frequency domain and stored in the form of images.

5.2 TIME DOMAIN FEATURES

Time domain statistical features extracted from pre-processed vibration signal in this research can be grouped as:

1. Conventional Statistical Features
2. Proposed Novel Statistical Features

In this research work a comprehensive set of 31 statistical features have been used. Table 5.1 lists the notations of all features as T1, T2, T3, T4, T5,..... T30, T30*.

Table 5.1: Notations of Statistical Time Domain Features

T1	**T2**	**T3**	**T4**	**T5**
Arithmetic Mean	RMS	Root	Max	Peak -to- Peak
T6	**T7**	**T8**	**T9**	**T10**
Standard deviation	Median	25^{th} Percentile	75^{th} Percentile	Skewness
T11	**T12**	**T13**	**T14**	**T15**
Kurtosis	Crest factor	Shape factor	Impulse factor	Clearance /Margin Factor
T16	**T17**	**T18**	**T19**	**T20**
Skewness Factor	Kurtosis Factor	Geometric Mean	RSSQ	Mean Absolute Deviation
T21	**T22**	**T23**	**T24**	**T25**
Median Absolute Deviation	Zero crossing rate	Entropy	Histogram upper bound	Histogram lower bound
T26	**T27**	**T28**	**T29**	**T30/T30***
First Hjorth parameter	Second Hjorth parameter	Third Hjorth parameter	Weibull Negative log likelihood	Normal Negative log likelihood for single Gaussian/ GMM

5.2.1 Conventional Statistical Features

Table 5.2 shows the mathematical expression and significance of Statistical Time Domain Features for a sampled vibration signal , where x_i are individual data points of sampled vibrational signal and i=1 to N is the length of vibration signal i.e.

$$x[n] = x_i = x_1, x_2, x_3 ..... x_N. \tag{5.1}$$

The **dimensional** statistical features are T1, T2, T3, T4, T5, T6, T7, T8. T9. T10, T11, T18 and T19 and **dimensionless** statistical features are T12, T13, T14, T15, T16, T17. They are self-explanatory from the mathematical expressions given in Table 5.2.

Table 5.2: Mathematical expression and significance of Statistical Time Domain Features

Notation	**Feature**	**Mathematical Expression**	**Significance**
T1	Arithmetic Mean (AM)	$\frac{1}{N}\sum_{i=1}^{N} x_i$	Average value of a signal is termed as mean value. This is the DC value of the vibration signal and is the measure of central tendency.
T2	Root Mean Square (RMS)	$\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$	The Root Mean Square value is generally the most useful because it is directly related to the energy content of the vibration signal. As faults get developed, the Root Mean Square value increases progressively.
T3	Root	$(\frac{1}{N}\sum_{i=1}^{N} \sqrt{\lvert x_i \rvert})^2$	This is computed as the square of mean of the root of the signal.
T4	Max	$max(\lvert x \rvert)$	This is the maximum value of the signal. The peak value is valuable for shock events.
T5	Peak–to-Peak	$max(\lvert x \rvert) - min(\lvert x \rvert)$	Peak–to-Peak value of a vibration signal is the difference between maximum and minimum values in a signal, that provides the maximum excursion of the wave. This is useful when looking at the displacement information.

T6	Standard deviation (SD)	$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu)^2}$	The standard deviation is the average deviation of the signal if the averaging is done with power instead of amplitude ($power \propto voltage^2$). It is the measure of energy content in the vibration signal. The standard deviation only measures the AC component of a signal, while the rms value measures both the AC and DC components.
T7	Median	$\frac{50(N+1)}{100}th\,observation$	The median is the same as the 50th percentile or second quartile (Q2), since it is that number such that 50% observations are found less than or equal to it.
T8	25^{th} Percentile	$\frac{25(N+1)}{100}th\,observation$	This is the lower or first quartile (Q1). It is a number such that 25% of observations are less than or equal to it.
T9	75^{th} Percentile	$\frac{75(N+1)}{100}th\,observation$	This is the upper or third quartile (Q3). It is a number such that 75% of observations are less than or equal to it.
T10	Skewness (SK)	$\frac{1}{\sigma^3}(\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu)^3)$	Skewness measures the asymmetry of probability density function of vibration signal.
T11	Kurtosis (KU)	$\frac{1}{\sigma^4}(\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu)^4)$	Kurtosis measures the degree of flatness of the pdf near its center. It is a measure of whether the data is peaked or flat relative to a normal distribution.

T12	Crest Factor (CF)	$\frac{max(\|x\|)}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}}$	Crest Factor is defined as the max value of a waveform divided by its RMS value (T4/T2). CF is a measure of how much impact occurs during rolling element and raceway contact in the bearing and hence is appropriate for "spiky signals".
T13	Shape Factor (SHF)	$\frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}}{\frac{1}{N}\sum_{i=1}^{N} \|x_i\|}$	Shape factor is defined as Root Mean Square value divided by mean value of a signal (T2/T1).
T14	Impulse Factor (IF)	$\frac{max(\|x\|)}{\frac{1}{N}\sum_{i=1}^{N} \|x_i\|}$	Impulse factor is defined as max value divided by mean value of a signal (T4/T1).
T15	Margin Factor (MF)	$\frac{max(\|x\|)}{(\frac{1}{N}\sum_{i=1}^{N} \sqrt{\|x_i\|})^2}$	Margin or Clearance factor is known as the ratio of max value and root value of a signal (T4/T3).
T16	Skewness Factor (SF)	$\frac{1}{\sigma^3}\frac{\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu)^3}{(\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2})^3}$	Skewness Factor is defined as skewness divided by $(rms)^3$ of a signal.
T17	Kurtosis Factor (KF)	$\frac{\frac{1}{\sigma^4}(\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu)^4)}{(\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2})^4}$	Kurtosis Factor is defined as Kurtosis divided by $(rms)^4$ of a signal.
T18	Geometric Mean (G-Mean)	$(\prod_{i=1}^{N} x_i)^{1/N}$	Geometric-mean is the Nth root of the product of N numbers. N is the total number of samples of the vibration signal.

T19	Root-Sum-of-squares (RSSQ)	$\sqrt{\sum_{=1}^{N} \lvert x_i \rvert^2}$	RSSQ level of x(n) is defined as square root of sum of square values of x(n). It is also referred to as L2 norm.
T20	Mean Absolute Deviation (MAD)	$mean(\lvert x_i - \mu \rvert)$	It is defined as the mean of the absolute deviations from the data's mean.
T21	Median Absolute Deviation (MeAD)	$median(\lvert x_i - median(x_i) \rvert)$	It is defined as the median of the absolute deviations from the data's median. MeAD is a measure of statistical dispersion. As compared to standard deviation, its more resilient to outliers in a data set.
T22	Zero crossing rate (ZCR)	$\frac{number\ of\ zero\ crossings}{total\ number\ of\ points}$	A point where a signal changes sign from positive to negative or negative to positive is called a zero-crossing point. Zero crossing rate denotes the number of times the signal level crosses zero point during a constant period of time. It hence provides a good understanding of the frequency of the signal and is very useful for separating signals with different frequencies.
T23	Entropy	$-\sum_{i=1}^{N} h(x_i) log_2 h(x_i)$	Entropy, e(X), is a statistical measure of the uncertainty and randomness of a sampled vibration data. Entropy estimation is a two stage process: first, a histogram is estimated, and then the entropy is calculated.

T24	Histogram upper bound (HU)	$max(x_i) + \frac{max(x_i) - min(\frac{x_i}{N-1})}{2}$	A histogram is a discrete PDF that is used to characterize the amplitude of the vibrational data. Histogram Upper Bound is defined as shown in the formula.
T25	Histogram lower bound (HL)	$max(x_i) - \frac{max(x_i) - min(\frac{x_i}{N-1})}{2}$	The lower bound of Histogram is defined as shown in the formula.
T26	Hjorth parameter-Activity (HA)	$\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2$	Hjorth Parameter 1 i.e. activity parameter represents the variance of a time function. This indicates the surface of the power spectrum in the frequency domain.
T27	Hjorth parameter-Mobility (HM)	$\frac{standard\ deviation\ of\ x^{'}(n)}{standard\ deviation\ of\ x(n)}$	Mobility represents the mean frequency or the proportion of standard deviation of the power spectrum. It is defined as the ratio of standard deviation of the first derivative of the vibration signal and the standard deviation of the vibration signal.
T28	Hjorth parameter-Complexity (HC)	$\frac{Mobility\ of\ x^{'}(n)}{Mobility\ of\ x(n)}$	Complexity is defined as the ratio of mobility of the first derivative of vibration signal to the mobility of the vibration signal. It compares the signal's similarity to a pure sine wave, where the value converges to 1 if the signal is more similar.

T29	Weibull Negative log likelihood (WNLL)	$-\sum_{i=1}^{N} logf(a,b\|x_i)$	This gives the negative log likelihood that weibull distribution fits the sample data. Here, f is the weibull distribution, and a and b are the scale and shape parameters of the distribution respectively.
T30	Normal Negative log likelihood for single Gaussian (NNLL)	$-\sum_{i=1}^{N} log[\frac{1}{\sigma\sqrt{2\pi}} exp\frac{-(x_i-\mu)^2}{2\sigma^2}]$	It gives the negative of the likelihood that a normal distribution of given mean and standard deviation fits the sample data.
T30*	Normal Negative log likelihood for Gaussian mixture model	$-\sum_{i=1}^{N} log(\sum_{k=1}^{K} \pi_k N(x_i\|\mu_k,\sum_k))$	It gives the negative of the likelihood that a Gaussian mixture model of k means and k standard deviations fits the sample data. It is an improvement over T30.

T20 and T21 are the arithmetic means of the absolute deviations of each x_i from mean and median respectively. T22 is the rate of zero crossing (ZCR) and can be defined as the ratio of number of zero crossings to the total number of points in the signal. In other words, ZCR is the rate at which a signal changes sign from negative to positive or vice versa, thus providing a good understanding of the frequency of the signal and hence very useful for separating signals with different frequencies. T23 or entropy, is a measure of randomness of the values of a vibration signal.

The two histogram features T24 and T25 are Histogram upper and lower bound respectively. T29 or Weibull Negative log likelihood, is computed as below:

$$T29 = -log\prod_{i=1} f(a,b|x_i) = -\sum_{i=1}^{N} logf(a,b|x_i) \tag{5.2}$$

T30 or Normal Negative log likelihood for single Gaussian is computed as:

$$T30 = -\sum_{i=1}^{N} log[\frac{1}{\sigma\sqrt{2\pi}} exp\frac{-(x_i-\mu)^2}{2\sigma^2}] \tag{5.3}$$

5.2.2 Proposed Novel Statistical Features

Finding fault sensitive features has been a hot topic of research for analysis of not just bearing vibration signals but other types of signals too. Hjorth derived three statistical features: Activity, Mobility and Complexity, based on the concept of variance in order to characterize Electroencephalogram (EEG) pattern [150]. Since then, these Hjorth Parameters have been frequently used by researchers to study electrical activity of the human brain for neurological disorders through electroencephalography (EEG) signals dataset acquired from EEG readings collected from patient cohorts.

Some of the studies that used these parameters to analyse EEG signals are: Hjorth Mobility and Complexity [151], Hjorth Activity and Mobility [152], Hjorth Activity and Complexity along with fuzzy detection [153], Hjorth parameters in association with Power Spectral Density and Adaptive Autoregressive coefficients as feature vector for Extreme Learning Machine [154], Hjorth Parameters combined with Genetic Algorithm along with Principal Component analysis followed by k-Nearest Neighbour classification [155], Hjorth Parameters and four classifiers: Support Vector Machine, k-Nearest Neighbour, Classification Trees and Artificial Neural Network [156].

In addition to EEG analysis, Hjorth Parameters have been found to be useful in other domains such as Lung sound analysis [157], In-hand object recognition via surface textures [158] etc.

This work does the unprecedented research of evaluating the effectiveness of Hjorth parameters for rolling element bearing fault diagnosis. Furthermore, it also introduces and explores another entirely unconnected and new statistical feature Normal Negative log likelihood for Gaussian mixture model distribution.

As per the nomenclature in this research work, T26, T27 & T28 statistical parameters are the Hjorth parameters in the list of 31 statistical parameters and the fourth proposed novel feature known as Normal Negative log likelihood for Gaussian mixture model distribution is T30*.

The mathematical expressions and the explanation of these parameters is given below:

1. **Hjorth Activity (HA) :** The first Hjorth parameter, Activity, represents signal power and can be computed by calculating the variance of the vibration signal amplitude. i.e.

$$Activity(x(t)) = variance(x(t)) \tag{5.4}$$

where x(t) is the vibration signal.

If $x(n) = x_i = \{x_1, x_2, x_3 \ldots\ldots\ldots x_N\}$ is the sampled vibration signal, then the standard deviation(σ) of $x(n)$ is :

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(x(n)-\mu)^2} \tag{5.5}$$

where $\mu = \frac{1}{N}\sum_{n=1}^{N}|x(n)|$ is the mean of x(n).

So, activity of x(n) is defined as:

$$Activity(x(n)) = \sigma^2 \tag{5.6}$$

2. **Hjorth Mobility (HM):** The second Hjorth parameter, Mobility, is defined as the square root of the ratio of variance of the first derivative of the vibration signal $\frac{d(x(t))}{dt}$ to the variance of the vibration signal $x(t)$ or as the ratio of the standard deviation of the first derivative of the vibration signal $\sigma_{x'}$ to the standard deviation σ_x of the vibration signal.

$$Mobility(x(t)) = \sqrt{\frac{Variance(\frac{dx(t)}{dt})}{Variance(x(t))}} = \frac{\sigma_{x'}}{\sigma_x} \tag{5.7}$$

If $x(n) = x_i = \{x_1, x_2, x_3 \ldots\ldots\ldots x_N\}$ is sampled vibration signal,

then the first difference is defined as:

$$x'(n) = x(n+1) - x(n) \qquad for\ n = 1, 2, \ldots N \tag{5.8}$$

If σ_1 = standard deviation of the first difference of the vibration signal $x'(n)$,

then Mobility of x(n) is defined as:

$$Mobility(x(n)) = \frac{\sigma_1}{\sigma} \tag{5.9}$$

3. **Hjorth Complexity (HC):** The third Hjorth parameter, Complexity, is defined as the ratio of mobility of the first derivative of vibration signal to the mobility of the vibration signal. It gives an estimate of the bandwidth of the signal. It indicates how closely the vibration signal resembles a pure sine wave.

$$Complexity(x(t)) = \frac{Mobility(x^{'}(t))}{Mobility(x(t))} = \frac{Mobility(\frac{dx(t)}{dt})}{Mobility(x(t))} = \sqrt{\frac{\frac{\sigma_{x''}}{\sigma_{x'}}}{\frac{\sigma_{x'}}{\sigma_x}}} \tag{5.10}$$

Where $\sigma_{x''}$ is the standard deviation of the second derivative of the vibration signal $x''(t)$.

For sampled vibration signal $x(n) = x_i = \{x_1, x_2, x_3 \ldots\ldots x_N\}$, the Complexity of x(n) is defined as:

$$Complexity(x(n)) = \sqrt{\frac{\frac{\sigma_2}{\sigma_1}}{\frac{\sigma_1}{\sigma}}} \tag{5.11}$$

where

σ_1 : standard deviation of the first difference of the vibration signal $x'(n)$

σ_2 : standard deviation of the second difference of the vibration signal $x''(n)$.

4. **Normal Negative log likelihood for Gaussian mixture model distribution:** The fourth novel statistical feature proposed in this research work is **T30***, which is Normal Negative log likelihood for Gaussian Mixture Model distribution. It is an improvement over T30 (Normal Negative Log-Likelihood for Single Gaussian) which gives the negative of the likelihood that a normal distribution of given mean and standard deviation fits the sample data, in this case, the sample data being the representative IMF of bearing vibration signal.

The Normal Negative Log Likelihood for Gaussian Mixture model (T30*) gives the negative of the likelihood that a GMM of k means and k standard deviations fits the sample data.

Mathematical formula for Normal Negative log likelihood of GMM can be defined as:

$$-log\,Pr(\pi, \mu, \sum) = -\sum_{i=1}^{N} log(\sum_{k=1}^{K} \pi_k N(x_i|\mu_k, \sum_k)) \tag{5.12}$$

where π_k is the mixing coefficient of the kth Gaussian.

Here k was instantiated as the number of local maxima in the probability distribution of the sample vibration data.

5.2.3 Results and Discussion

The results of feature extraction and effectiveness of proposed novel features are as below:

1. Numerical values of all statistical features extracted from the best IMF of a sample of vibration signals taken from CWRU Dataset are listed in Table 5.3. The first column in this table contains the names of the Matlab files used for storing the signals. The second column represents fault class and the remaining columns represent the statistical features. These features will be ranked in the next chapter in order to investigate their effectiveness and also to find an optimal feature subset for fault type diagnosis and severity detection.

2. Numerical values of three Hjorth Parameters (HA, HM, HC) obtained from the best IMF of few sample vibration signals taken from CWRU experimental rig have been recorded in Table 5.4.

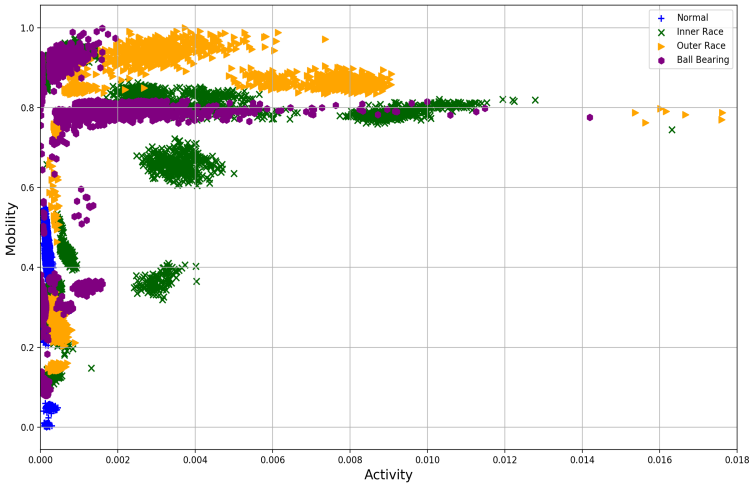
   To investigate the effectiveness of Hjorth parameters in the prediction of bearing faults, scatter plots for the pairs of the three parameters have been studied, particularly the layout and structure of the clusters obtained in these plots. These plots are reported in Figure 5.1. The data points are coloured according to their true labels.

   From Figure 5.1. a), it can be seen that the plot of Hjorth Activity and Hjorth Mobility forms a number of distinct, well defined clusters with slight overlapping in few clusters.

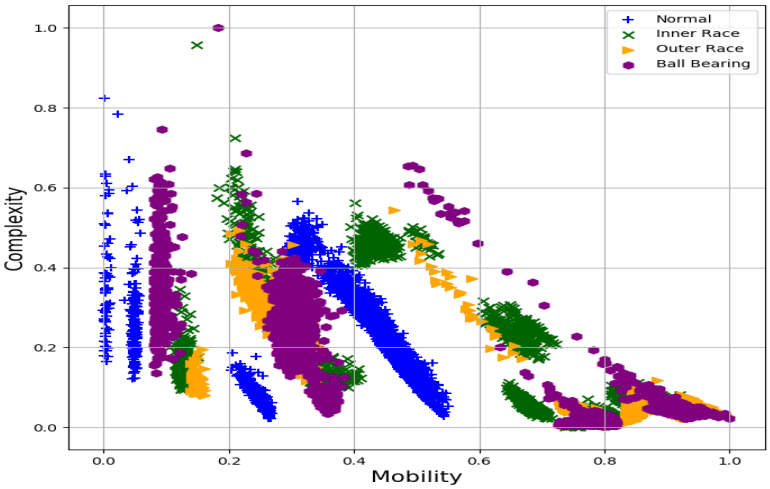
   From Figure 5.1. b), it is seen that the plot of Hjorth Mobility and Hjorth Complexity too has a number of distinct clusters though those are less well defined and show relatively more overlapping.

   From Figure 5.1. c), it is seen that the clusters are not distinct in the plot of Hjorth Activity and Hjorth Complexity. They are not well defined and show a lot of overlapping.

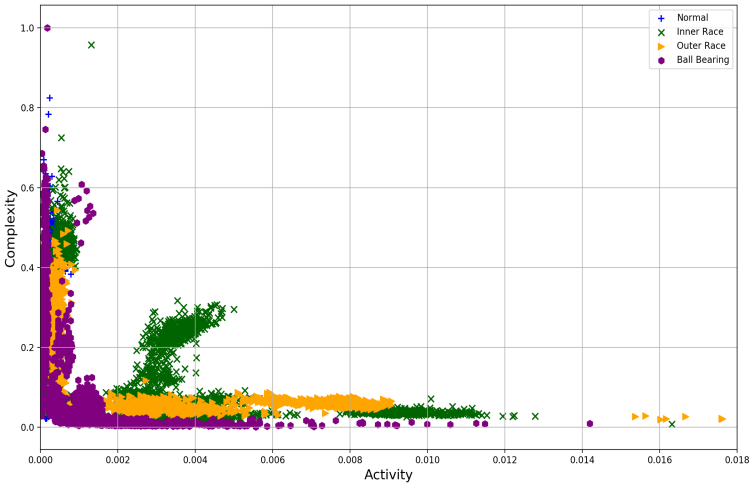
   Hence, it can be concluded that the pair of Activity and Mobility is capable of comfortably distributing the bearing vibration data into clusters. These parameters can be useful for machine learning classifiers in performing accurate fault diagnosis.



(a) Activity and Mobility



(b) Mobility and Complexity



(c) Activity and Complexity

Figure 5.1: Scatter plots for each of the three pairs of Hjorth parameters.

Table 5.3: Numerical values of Statistical parameters for sample vibration signals. The first column contains the names of the Matlab files used for storing the signals.

Filename	**Class**	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**
		-0.0005	0.0659	0.0445	0.2559	0.256	0.066
		T7	**T8**	**T9**	**T10**	**T11**	**T12**
05_BA_time	Inner	-0.0010	-0.0436	0.0443	-0.106	3.099	3.882
_100_best_imf	Race	**T13**	**T14**	**T15**	**T16**	**T17**	**T18**
.mat		-128.2	-497.685	5.7556	-369.738	164000.6	0.035
		T19	**T20**	**T21**	**T22**	**T23**	**T24**
		2.1098	0.0525	0.0442	0.2148	-1.396	-0.234
		T25	**T26**	**T27**	**T28**	**T29**	**T30**
		-0.1904	0.0044	0.6449	1.1506	-2032.33	-1331.46
		T1	**T2**	**T3**	**T4**	**T5**	**T6**
		0.0003	0.0599	0.0412	0.1826	0.182	0.06
105_BA_time	Inner	**T7**	**T8**	**T9**	**T10**	**T11**	**T12**
_101_best_imf	Race	0.0016	-0.0421	0.044	-0.0126	2.805	3.048
.mat		**T13**	**T14**	**T15**	**T16**	**T17**	**T18**
		223.2820	680.5035	4.432	-58.3245	217539.5	0.033
		T19	**T20**	**T21**	**T22**	**T23**	**T24**
		1.9176	0.0483	0.0435	0.2314	-1.49	-0.164
		T25	**T26**	**T27**	**T28**	**T29**	**T30**
		-0.1271	0.0036	0.6936	1.1212	-2126.73	-1429.21
		T1	**T2**	**T3**	**T4**	**T5**	**T6**
		0.0006	0.2735	0.1616	1.1908	1.191	0.274
105_DE_time	Inner	**T7**	**T8**	**T9**	**T10**	**T11**	**T12**
_100_best_imf	Race	-0.0068	-0.14	0.1465	0.0341	5.305	4.353
.mat		**T13**	**T14**	**T15**	**T16**	**T17**	**T18**
		481.2894	2095.182	7.3669	1.6653	947.622	0.123
		T19	**T20**	**T21**	**T22**	**T23**	**T24**
		8.7531	0.1989	0.1437	0.5283	0.010	-1.062
		T25	**T26**	**T27**	**T28**	**T29**	**T30**
		-0.8253	0.0749	1.4574	1.0405	-636.742	125.548

105_DE_time_101_best_imf.mat	Inner Race	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**
		0.003	-0.1296	0.134	-0.0274	5.592	4.609
		T7	**T8**	**T9**	**T10**	**T11**	**T12**
		496.6332	2288.8202	8.0478	-1.4562	1119.748	0.115
		T13	**T14**	**T15**	**T16**	**T17**	**T18**
		-1.1073	2095.182	7.3669	1.6653	947.622	0.123
		T19	**T20**	**T21**	**T22**	**T23**	**T24**
		8.5066	0.1898	0.132	0.5225	-0.046	-1.062
		T25	**T26**	**T27**	**T28**	**T29**	**T30**
		-0.8716	0.0707	1.4378	1.0443	-680.718	96.303
105_FE_time_100_best_imf.mat	Inner Race	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**
		0.0082	0.1679	0.1137	0.6734	0.673	0.168
		T7	**T8**	**T9**	**T10**	**T11**	**T12**
		0.0104	-0.1036	0.1179	-0.0102	4.134	4.011
		T13	**T14**	**T15**	**T16**	**T17**	**T18**
		20.5513	82.4341	5.9238	-2.1558	5204.191	0.093
		T19	**T20**	**T21**	**T22**	**T23**	**T24**
		5.3724	0.1319	0.1114	0.5879	-0.469	-0.536
		T25	**T26**	**T27**	**T28**	**T29**	**T30**
		-0.4089	0.0281	1.3253	1.19	-1105.545	-375.517
105_FE_time_101_best_imf.mat	Inner Race	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**
		0.0086	0.1696	0.1151	0.7632	0.763	0.169
		T7	**T8**	**T9**	**T10**	**T11**	**T12**
		0.0121	-0.1039	0.1163	0.1066	4.390	4.499
		T13	**T14**	**T15**	**T16**	**T17**	**T18**
		19.628	88.3155	6.6296	21.837	5303.249	0.095
		T19	**T20**	**T21**	**T22**	**T23**	**T24**
		5.428	0.1331	0.1099	0.5723	-0.466	-0.688
		T25	**T26**	**T27**	**T28**	**T29**	**T30**
		-0.5386	0.0287	1.2723	1.2223	-1099.272	-365.088

Table 5.4: Numerical values of Hjorth parameters for sample vibration signals. The first column contains the names of the Matlab files used for storing the signals.

Filename	**Class**	**HA**	**HM**	**HC**
Normal_3100_DE_time _154_best_imf.mat	Normal	0.001198587	0.47914082	1.078444013
Normal_2 99_DE_time _201_best_imf.mat	Normal	0.001014132	0.973725602	1.07729559
Normal_0 97_FE_time _93_best_imf.mat	Normal	0.00320465	0.624952755	1.397140811
211_BA_time _111_best_imf.mat	Inner Race	0.002050358	0.278848771	1.148809846
211_DE_time _42_best_imf.mat	Inner Race	0.206956722	1.383336326	1.005922226
198_DE_time _88_best_imf.mat	Outer Race	0.007234727	1.563411856	1.029569729
131_FE_time _64_best_imf.mat	Outer Race	0.063591347	1.615458314	1.073613976
170_DE_time _22_best_imf.mat	Inner Race	0.035018519	1.463248107	1.029223682
118_DE_time _74_best_imf.mat	Ball Bearing	0.013313157	1.468992945	1.015068866
237_FE_time _10_best_imf.mat	Outer Race	0.02815248	1.760488596	1.048879752
Normal_299_FE_time _208_best_imf.mat	Normal	0.001682639	0.876762229	1.171696348
197_DE_time _109_best_imf.mat	Outer Race	0.008830986	1.566142927	1.028934363
234_BA_time _76_best_imf.mat	Outer Race	0.004193641	0.495338199	1.299229055
200_DE_time _53_best_imf.mat	Outer Race	0.013163485	1.545402627	1.034146996
172_BA_time _67_best_imf.mat	Inner Race	0.001758115	0.640674039	1.123165077
Normal_2 99_FE_time _272_best_imf.mat	Normal	0.001561567	0.87454923	1.177806602
188_DE_time _51_best_imf.mat	Ball Bearing	0.043607085	1.45481856	1.010113852
212_FE_time _31_best_imf.mat	Inner Race	0.016988761	1.549122899	1.073225708
199_FE_time _82_best_imf.mat	Outer Race	0.007167253	1.683912138	1.040435214
131_DE_time _8_best_imf.mat	Outer Race	0.309513337	1.499372493	1.02141219

3. Feature T30 was replaced with its Gaussian Mixture Model variant, labelled as T30*. This was done because on selecting some random samples from the set of representative IMF signals and plotting their frequency distribution histograms with fitting curves (shown in Figure 5.2.), it was found that the idea of fitting a single Gaussian curve is not working well with some of the signals. Some signals, like samples 1 and 2 in Figure 5.2.a) and Figure 5.2.b) respectively, have more than one distinct maxima and a single Gaussian curve was not be able to accurately model these signals.

   This trend was mostly seen for samples belonging to faulty classes. The Normal Negative Log Likelihood for GMM values obtained for the 3 samples are plotted in Figure 5.2. The values shown are taken before Min-Max scaling was applied to the data.

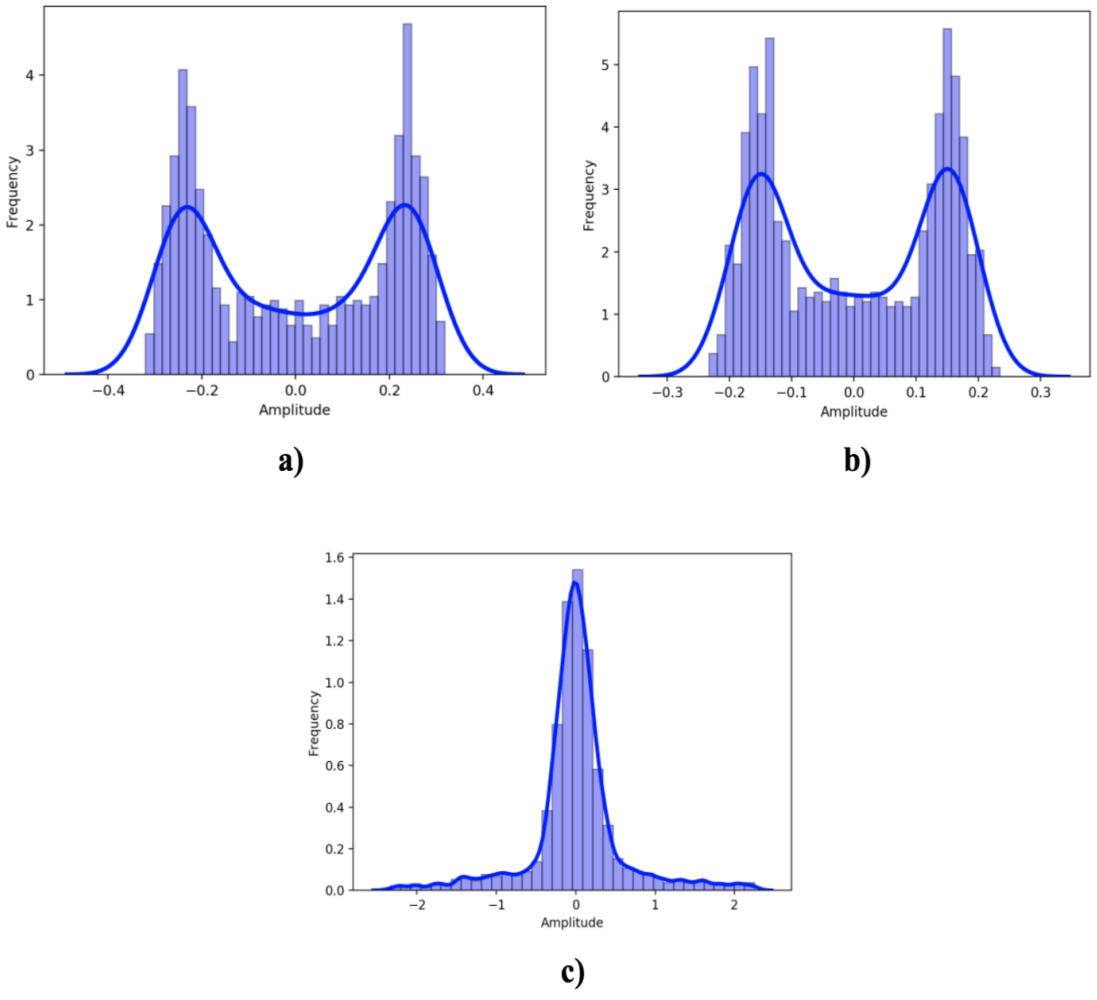


Figure 5.2: Frequency Distribution Histograms with fitting curves for selected samples. (a) Sample 1. (b) Sample 2. (c) Sample 3.

Table 5.5: Comparison of Normal Negative Log Likelihood for Gaussian Mixture Models with Normal Negative Log Likelihood for Single Gaussian for selected samples.

Sample No.	**Normal Negative Log Likelihood for Single Gaussian (T30)**	**Normal Negative Log Likelihood for GMM (T30*)**
1	-143.347	-466.000
2	-567.550	-987.950
3	901.479	901.479

The Normal Negative Log Likelihood for GMM values obtained for the 3 samples are also recorded in Table 5.5. The Normal Negative Log Likelihood for GMM for Sample 1 is -466, which is far better than its Normal Negative Log Likelihood for Single Gaussian equal to -143.347, in terms of fitting of the model. Same is the case with Sample 2. Sample 3 shows no change in value because it can be fitted well with a Single Gaussian. In general, it is concluded that proposed feature T30* has better fault sensitivity as compared to T30.

5.3 FREQUENCY DOMAIN FEATURES

The first and second order statistics completely describe the properties of a stationary signal with a Gaussian probability density function. But occurrence of faults causes non-linearities in the machine vibration signals and that's why signal processing techniques based on Higher Order Spectra (HOS) for vibration signature analysis, are used as they provide more diagnostic information even more than that provided by the power spectrum. Hence higher order statistical measures are used for vibration based condition monitoring to get details about the signal which the conventional second order statistics cannot provide. In this research, Bispectrum (the 3rd order cumulant, or 2-D Fourier transform of Triple auto-correlation function) of the vibration signals is generated. Bispectrum helps retain phase and amplitude information of the signals and thus helps in accurate fault detection.

5.3.1 Higher Order Spectral Feature: Bispectrum

Bispectrum analysis examines the relationship between the underlying sinusoidal components of the vibration signal at two primary frequencies, f1 and f2, and a

modulation component at the frequency (f1 + f2). Compared to power spectrum analysis, this analysis is superior because due to correlation calculation between the frequency components, the phase coupling between vibration signals can be identified.

For each triplet (f1, f2, and f1 + f2), the Bispectrum, B (f1, f2), a quantity incorporating both phase and power information, can be calculated as given in Equation 5.13 [159, 160].

$$B(f_1, f_2) = |X(f_1)X(f_2)X^*(f_1 + f_2)| \tag{5.13}$$

Where B (f1, f2) is the Bispectrum at frequencies f1 and f2, $X(f_1)$ and $X(f_2)$ are discrete Fourier transform coefficients at frequencies f_l and f_2 respectively. X^* is the complex conjugate of X.

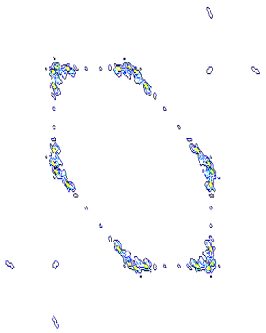
5.3.2 Results and Discussion

In this research, bispectrum arrays were estimated for each of the data samples and were stored as RGB Bispectrum images i.e. Bispectral colour contour maps. The dataset derived from original CWRU dataset has four fault types (Normal, Inner race, Outer race, Ball bearing) with two fault size diameters (0.007 inch, 0.014 inch). Hence, total seven classes (Normal, IR_07, IR_14, OR_07, OR_14, BB_07, BB_14) were considered.

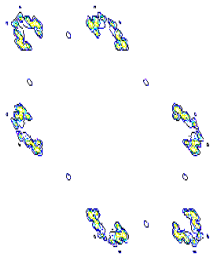
First, data samples of size 1024 points each were extracted from four types of CWRU vibration signals, that are Normal, Inner Race, Outer Race and Ball Bearing. These signals are further divided according to the fault sizes, resulting in 7 classes - Normal class, Inner race with 0.007 inch and 0.014 inch, Outer race with 0.007 inch and 0.014 inch, and finally Ball bearing with 0.007 inch and 0.014 inch.

Bispectrum arrays were then estimated for each of the data samples using the MATLAB HOSA toolbox for direct fast fourier transform (FFT) based approach with origin at the centre of the array and axis pointing down and towards right. FFT length and overlap were taken as 128 and 0 respectively. The Bispectrum arrays were stored as RGB images. Bispectrum Images obtained for Ball Bearing and Normal classes are 875x656 in size and that for Inner Race and Outer Race are 1200 x 900 in size.

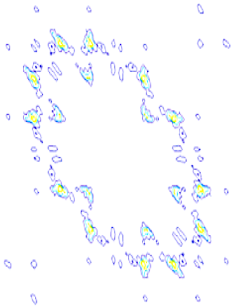
Sample Bispectrum images for seven cases of bearing faults mentioned above are shown in Figure 5.3.



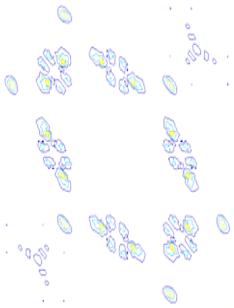
(a) Ball bearing with 0.007” fault



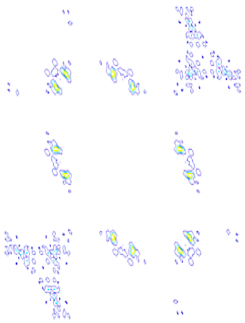
(b) Ball bearing with 0.014” fault



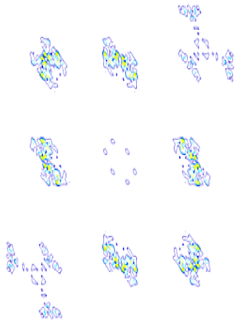
(c) Inner Race with 0.007” fault



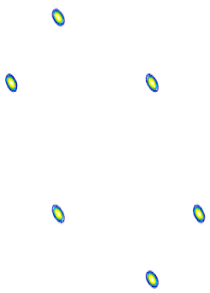
(d) Inner Race with 0.014” fault



(e) Outer Race with 0.007” fault



(f) Outer Race with 0.014” fault



(g) Normal class

Figure 5.3: Bispectrum Images of Bearing Fault Signals.

5.4 CHAPTER SUMMARY

This chapter discusses time and frequency domain fault-sensitive features extracted from best IMF obtained from pre-processing of vibration signals. Since each IMF represents the natural oscillatory mode embedded in the raw vibration signal, the features extracted from such components are more effective for fault diagnosis than those extracted directly from the raw signal. Hence to acquire more accurate fault information, fault sensitive features were extracted not directly from raw vibration signals but from the best IMF produced by EMD of vibration signal. Definition, mathematical formula and significance of conventional statistical features and proposed novel statistical features from time domain has been explained. The values of these statistical features for some sample vibration signals representing different types of faults are presented in tabular form. Effectiveness of proposed novel features is justified with the help of graphical plots. From frequency domain, framework of higher order spectral feature, bispectrum, has been discussed and sample bispectrums have been shown in image form for various types and severity of bearing faults.

CHAPTER VI
FEATURE SELECTION FOR DIMENSIONALITY REDUCTION

6.1 INTRODUCTION

The performance of machine learning algorithms, in terms of accuracy and efficiency, largely depends on the number of features used and their discriminating capability. For this reason, Researchers have proposed many dimensionality reduction techniques in the past. Selecting the right set of features has been shown to improve the performance of machine learning classifiers, in the case of high-dimensional input data, to mitigate the curse of dimensionality. This chapter proposes a methodology for "Optimal feature subset selection" and discusses its results when applied on 31 extracted statistical time domain features.

There are three types of feature selection methods as shown in Figure 6.1.

1. **Filter methods** that apply a statistical measure for assigning a score to each feature.
2. **Wrapper methods** that consider the selection of a set of features as a search problem, where different combinations are prepared, evaluated and compared to other combinations.
3. **Embedded methods** that learn which features best contribute to the accuracy of the model while the model is being created.

Filter based feature ranking and feature subset selection techniques have been used in this research work. The proposed methodology for "Optimal feature subset selection" based on filter based feature ranking and feature subset selection techniques is explained in detail in the following sections.

6.2 FEATURE RANKING

Feature-ranking techniques rank every individual feature of the given set according to some mathematical criteria. The reduced set of features can then be selected manually as per the ranks generated. These feature ranking techniques use different statistical ranking metrics like Chi-square, Information Gain, Gain ratio etc.

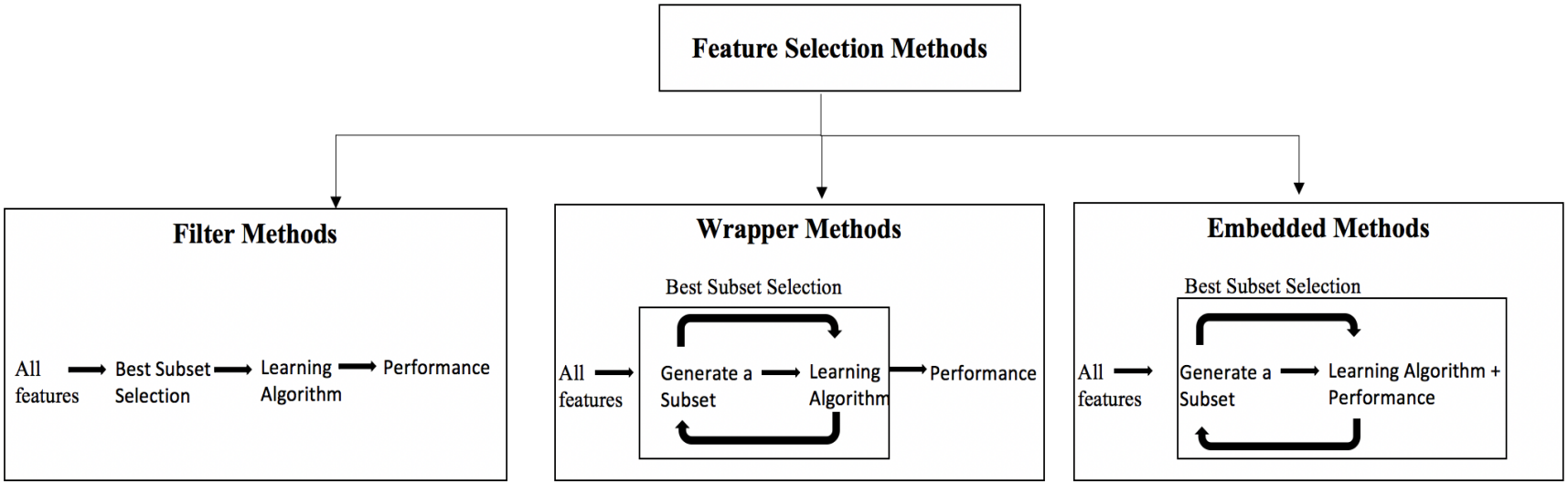


Figure 6.1: Feature Selection methods

6.2.1 Filter Based Feature Ranking

Filter-based feature selection algorithms have been used effectively in fault diagnosis [161, 162]. Filter based techniques work well with any classification model producing accurate results with shorter run times [161]. Moreover, these techniques perform well for identification of dominant features for both fault type and severity detection [162].

In this research, following three filter-based ranking metrics have been used to rank the features for identifying important features for bearing fault diagnosis:

1. **Chi-square:** In Chi-square feature selection, the value of a statistic called chi-square (χ_c^2), is calculated for the feature and response variables. Those features that have high chi-square values are then selected. The chi-square value is calculated as:

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{6.1}$$

   where c = degrees of freedom, O_i = observed value, E_i = expected value.

2. **Information Gain:** Information Gain test uses the value of a statistic called information gain, that is computed for each feature (L) in the context of a response variable (R), ultimately selecting features with high information gain values. Higher the information gain score for an attribute, more is the information about the response variable that the attribute can contribute.

   Information gain can be understood as the reduction in entropy (H) and is calculated as:

$$Information\ Gain = \Delta H = H - \frac{m_L}{m} H_L - \frac{m_R}{m} H_R \tag{6.2}$$

   where,

$$Entropy(H) = -\sum_{i=1}^{K} p_k log_2 p_k \tag{6.3}$$

   and m = total number of instances, m_k = number of instances belonging to class k, where i = 1, . . . , K.

3. **Gain Ratio:** The third metric of Gain Ratio is a ratio of information gain to the intrinsic information as given below:

$$GainRatio(S,A) = \frac{Gain(S,A)}{SplitEntropy(S,A)} \tag{6.4}$$

where

$$SplitEntropy(S,A) = -\sum_{V \in Values(A)} \frac{|S_Y|}{|S|} log \frac{|S_Y|}{|S|} \tag{6.5}$$

where A = candidate attribute, V = possible values of A, S = set of all training examples X, S_V = subset where X_A = V.

In this method, the attribute with the highest gain ratio is selected as the splitting attribute.

6.2.2 Experimental Description

Three feature ranking techniques, based on the statistical metrics of ChiSquared, Info gain and Gain Ratio, were applied to the data. These techniques ranked the features according to their importance in classifying the samples into 4 classes for Dataset A [Table 3.3] and likewise into 12 classes for Dataset B [Table 3.4]. The threshold for the search method used by the feature ranking techniques was set in a way to ensure none of the features gets discarded.

6.2.3 Results and Discussion

The ranking of the features is shown in Table 6.1 and Table 6.2. The 3rd, 4th and 5th columns list the rank obtained by ranker methods ChiSquared, Info gain and Gain Ratio respectively, for the feature in the corresponding row. The last column shows the average rank score calculated by averaging out the ranks of features over all the three ranking metrics.

6.3 FEATURE SUBSET SELECTION

Feature subset selection process identifies a subset of features that will provide the same results as the original full set of features. The process checks the relative importance of the features so that fault-sensitive features can be retained and irrelevant or redundant features can be removed. Selecting the optimal set of features reduces the complexity and improves the accuracy of a model. It also enables the model to train

Table 6.1: Feature ranking of statistical features for CWRU dataset A.

Feature	**Feature Description**	**Chi-Squared Statistics**	**Info Gain**	**Gain Ratio Score**	**Average Rank**
T22	Zero crossing rate	1	1	1	1
T27	Hjorth Parameter - Mobility	2	2	3	2.33
T30	Normal Neg Log Likelihood	3	3	4	3.33
T26	Hjorth Parameter - Activity	4	4	7	5
T2	Root Mean Square	6	6	5	5.67
T6	Standard Deviation	5	5	8	6
T19	Root-sum-of-squares	7	7	6	6.67
T18	Geometric Mean	12	12	2	8.67
T20	Mean Absolute Deviation	9	9	9	9
T23	Entropy	8	8	14	10
T29	Weibull Neg Log Likelihood	10	11	12	11
T3	Root	11	10	13	11.33
T21	Median Absolute Deviation	13	13	11	12.33
T9	75th Percentile	16	15	15	15.33
T17	Kurtosis Factor	15	14	17	15.33
T24	Histogram upper bound	14	16	16	15.33
T8	25th Percentile	20	19	10	16.33
T5	Peak-to-Peak	17	18	19	18
T4	Max	18	17	20	18.33
T25	Histogram lower bound	19	20	18	19
T28	Hjorth-Complexity	21	21	21	21
T16	Skewness Factor	22	22	22	22
T15	Clearance factor	23	23	23	23
T11	Kurtosis	24	24	24	24
T10	Skewness	25	26	25	25.33
T12	Crest Factor	27	25	26	26
T13	Shape Factor	26	27	27	26.67
T14	Impulse Factor	28	28	28	28
T7	Median	29	29	29	29
T1	Mean	30	30	30	30

Table 6.2: Feature ranking of statistical features for CWRU dataset B.

Feature	**Feature Description**	**Chi-Squared Statistics**	**Info Gain**	**Gain Ratio Score**	**Average Rank**
T2	Root Mean Square	3	4	4	3.67
T18	Geometric Mean	4	6	2	4.00
T20	Mean Absolute Deviation	1	3	8	4.00
T30	Normal Neg Log Likelihood	2	5	9	5.33
T22	Zero crossing rate	15	2	1	6.00
T27	Hjorth Parameter - Mobility	16	1	3	6.67
T17	Kurtosis Factor	5	13	5	7.67
T19	Root-sum-of-squares	6	14	6	8.67
T9	75th Percentile	7	15	7	9.67
T23	Entropy	8	12	11	10.33
T26	Hjorth Parameter - Activity	10	7	16	11.00
T29	Weibull Neg Log Likelihood	11	8	15	11.33
T6	Standard deviation	12	9	14	11.67
T21	Median Absolute Deviation	13	10	12	11.67
T8	25th Percentile	9	16	10	11.67
T3	Root	14	11	13	12.67
T24	Histogram upper bound	17	17	19	17.67
T4	Max	18	18	18	18.00
T5	Peak-to-peak	19	19	17	18.33
T25	Histogram lower bound	20	20	20	20.00
T28	Hjorth-Complexity	22	21	22	21.67
T11	Kurtosis	23	22	21	22.00
T16	Skewness Factor	21	23	24	22.67
T15	Clearance Factor	24	24	23	23.67
T12	Crest factor	25	25	25	25.00
T7	Median	26	26	27	26.33
T1	Mean	27	27	26	26.67
T10	Skewness	28	28	28	28.00
T14	Impulse Factor	29	29	30	29.33
T13	Shape factor	30	30	29	29.67

faster and reduces overfitting. Filter based feature subset selection techniques use conventional search techniques (Best first, Greedy search etc.) and the newly introduced swarm search techniques (Ant, Bat, Bee, Cuckoo, Elephant etc.) [163].

6.3.1 Conventional Search Techniques

In this research, two conventional search techniques have been performed:

1. **Best first search**: This is a conventional greedy search technique that finds subsets by backtracking and provides flexibility of searching in forward, backward and bi-directions.
2. **Greedy stepwise search**: This is similar to best first search as it also searches for subsets in a greedy manner but instead of backtracking, it searches in a stepwise fashion and stops when a decrease of evaluation occurs after an addition or deletion of an attribute.

6.3.2 Swarm Search Techniques

Swarm search methods are metaheuristics-based search methods that are nature-inspired optimization algorithms. Nine Swarm search techniques have been performed in this research:

1. **Ant search** is a stochastic combinatorial optimization technique that is modelled on how the ant colonies work. It finds solutions in the early stages of the search process [164].
2. **Bat search** is based on how bats determine object location by using sound waves [165].
3. **Bee search** is inspired by the foraging behaviour of honeybees. It solves optimization problems by doing exploitative neighbourhood search and random explorative search together [166].
4. **Cuckoo search** algorithm models the parasitic behaviour of cuckoos and the Levy flight behaviour of fruit flies [167].
5. **Elephant search** models the behavioural characteristics of elephant herds [168].
6. **Firefly search** is inspired by the flashing patterns of fireflies and offers the advantage of automatic subdivision and the ability to deal with multi-modality [169] .
7. **Flower Search** follows the survival of the fittest rule based on the process of

pollination of flowers [170].

8. **Wolf search** models how wolves hunt for prey and survive by avoiding their enemies. It is known to give good results when used on large search spaces [171].
9. **Rhinoceros search** models the herding behaviour of rhinos and can be used for solving global continuous optimization problems [172].

6.3.3 Experimental Description

Correlation based feature subset evaluator (CfsSubsetEval) is used to obtain a subset of features which show high correlation with ground truth labels but low correlation to each other. 11 search techniques, including both conventional search techniques (best first search and greedy stepwise search) and the newly introduced swarm search techniques were performed. The controlling parameters for search methods are as follows: for implementing Best first search, bidirectional search was used while in Greedy stepwise, forward search was used. The number of particles in the swarm, chaotic coefficient and mutation probability were set as 20, 4.0, 0.01 respectively for all searches. In Bee search, the discount factor for crossover was set as 0.8 and in Cuckoo search, the sigma rate was kept as 0.69. For Firefly search, zero distance attractiveness for the firefly population was set as 0.33. The same value was taken for wolf search. In Flower search, a pollination rate of 0.33 was used.

Table 6.3 and Table 6.4 show the set of features selected with these techniques for Dataset A and Dataset B.

6.3.4 Results and Discussion

The results can be concluded as below:

1. It is observed that T22 (Zero Crossing Rate) has highest rank followed by T27 (Hjorth parameter - Mobility), T30 (Normal Negative log likelihood), T26 (Hjorth parameter - Activity), T2 (RMS), T6 (Standard deviation), T19 (RSSQ), T18 (Geometric Mean), T20 (Mean Absolute Deviation) and T23 (Entropy) among the top 10 rank features for fault type detection. [Table 6.1].
2. Top 10 rank features for fault severity detection are T2 (RMS), T18 (Geometric Mean), T20 (Mean Absolute Deviation) , T30 (Normal Negative log likelihood), T22 (Zero Crossing Rate) , T17 (Kurtosis factor), T19 (RSSQ), T9 (75th percentile)

and T23 (Entropy). [Table 6.2]

3. It is observed that the subset of T2 (Root Mean Square), T18 (Geometric Mean), T22 (Zero Crossing Rate), T27 (Hjorth parameter - Mobility), T30 (Normal Negative log likelihood for Single Gaussian) is selected by majority of search techniques [Table 6.3]. It is also observed that these five features in the subset are among the top ten

Table 6.3: Feature subset selection results for CWRU dataset A.

Technique	**Set of features selected**
Best First	T2, T18, T22, T27, T30
Greedy Stepwise Search	T2, T18, T22, T27, T30
Ant Search	T2, T18, T22, T27, T30
Bat Search	T18, T22, T26, T27, T30
Bee Search	T18, T19, T22, T27, T30
Cuckoo Search	T2, T18, T22, T27, T30
Elephant Search	T18, T22, T26, T27, T30
Firefly Search	T6, T22, T26, T27, T30
Flower Search	T18, T19, T22, T27, T30
Rhinoceros Search	T2, T18, T22, T27, T30
Wolf Search	T2, T18, T22, T27, T30

Table 6.4: Feature subset selection results for CWRU dataset B.

Technique	**Set of features selected**
Best First	T2, T9, T17, T18, T19, T22, T27, T30
Greedy stepwise	T2, T9, T17, T18, T19, T22, T27, T30
Ant Search	T2, T9, T17, T18, T19, T22, T27, T30
Bat Search	T9, T17, T18, T22, T26, T27, T30
Bee Search	T2, T17, T18, T21, T22, T27, T30
Cuckoo Search	T9, T19, T18, T22, T26, T27, T29, T30
Elephant Search	T2, T9, T17, T19, T22, T27, T30
Firefly Search	T17, T18, T20, T21, T22, T26, T27
Flower Search	T2, T15, T17, T18, T22, T27, T19
Rhinoceros Search	T2, T3, T9, T19, T22, T27, T30
Wolf Search	T9, T17, T18, T20, T22, T27, T30

average rank scorers as shown in Table 6.1, hence validating the importance of these 5 features for fault type diagnosis. After replacing T30 by T30*, the selected optimal feature subset becomes **[T2, T18, T22, T27, T30*]**. It is noted that the features T6 (Standard Deviation),T19 (RSSQ) and T26 (Hjorth Parameter - Activity) are ranked well by the ranking metrics but get selected by only a few subset search techniques. This could be due to the way subset selection works. Since one of the criteria for a good subset is that the subset elements should have very less correlation among each other, features like T6, T19 and T26 got rejected by most of the search techniques as they may be showing some correlation with an already chosen member of the set. Since the aim is to select the best performing subset for fault diagnosis, the set of only the 5 features stated above was selected.

4. For fault severity detection, from Table 6.4, it is observed that the subset of features - T2 (RMS), T19 (RSSQ), T17 (Kurtosis Factor), T18 (Geometric Mean), T22 (Rate of zero crossing), T27 (Hjorth parameter - Mobility), T30 (Normal Negative log likelihood for Single Gaussian) and T9 (75th Percentile) - are selected by majority of search techniques, where these 8 features are also among the top 10 average rank scorers as shown in Table 6.2. After replacing T30 by T30*, the selected optimal feature subset becomes **[T2, T9, T17, T18, T19, T22, T27, T30*]**.

6.4 CHAPTER SUMMARY

In this chapter, the proposed methodology for "Optimal feature subset" selection based on filter based feature ranking and feature subset selection techniques is explained and implemented on two types of dataset for fault type and fault severity detection.

An in-depth two-step approach, based on Filter-based Ranking with three statistical ranking metrics (Chi-square Information Gain and Gain ratio) and Feature Subset selection with 11 search techniques (two conventional search techniques and nine swarm search techniques) was developed to select an optimal feature subset for bearing fault type detection and severity estimation. Filter based selection techniques being model-agnostic, make sure that the selected features work efficiently with any machine learning classifier.

Obtained results have been tabulated and discussed to obtain an optimal feature subset. Results show that features Root Mean Square, Geometric Mean, Rate of zero

crossing, Hjorth parameter Mobility, and Normal Negative log likelihood are important statistical features for bearing fault diagnosis. For fault severity detection, these five parameters and three additional parameters - 75th Percentile, Kurtosis Factor and RSSQ, collectively form an important and useful feature subset.

Improvement in one of the selected high ranked features, Normal Negative log likelihood for single Gaussian is also proposed by suggesting its replacement with its Gaussian Mixture Model (GMM) variant.

The performance of the proposed optimal feature subset with the Machine Learning classifiers has been validated and discussed in detail in the next chapter.

CHAPTER VII
FAULT DIAGNOSIS USING MACHINE LEARNING CLASSIFIERS

7.1 INTRODUCTION

Fault diagnosis and prognosis in power-plant components is often a labour-intensive and time-consuming practice and conducting effective and efficient fault diagnosis remains a challenge for plant maintenance technicians. Hence researchers are increasingly interested in automating the diagnosis procedure using artificial intelligence techniques along with advanced signal processing techniques. Tools and techniques from machine learning and deep learning domains, that can be used in conjunction with effective signal processing techniques, are being researched for effective decision making.

Automatic fault diagnosis using machine learning is conducted typically in the following order: Data Acquisition, Pre-processing, Feature Extraction, Feature Selection for Dimensionality Reduction and Fault Classification. These are discussed in Chapter 3, Chapter 4, Chapter 5, Chapter 6 & Chapter 7 respectively.

There are a variety of machine learning algorithms that have been used for intelligent fault diagnosis for many years. Literature review summarizing the use of various artificial intelligence techniques on fault diagnostics has been covered in depth in Chapter 2. The potential of Rule based classifiers in the domain of bearing fault diagnosis has been proved promising according to the literature review. Literature shows a growing research interest in combining a set of learning algorithms to generate ensembles for investigating complex problems. These ensemble algorithms tend to exploit the strengths of the base classifiers to enhance the overall accuracy. Hence in this research work a methodology for vibration signature analysis for fault type fault severity classification using **Ensemble Rule Base Classifiers** is proposed.

This chapter discusses fault classification using optimal feature subset and ensemble rule based classifiers. A feature vector is created using derived optimal feature subset, which is sent as an input to ensemble rule based Classifiers for

achieving fault classification. All steps of the proposed methodology (pre-processing, feature extraction, feature subset selection, along with fault classification using ensemble classifiers are shown in Figure 7.1.

7.2 FAULT CLASSIFICATION

This section discusses the classifiers used in this research as well the metrics used for comparing their performance.

7.2.1 Rule Based Classifiers

Rule based classifiers use "if...then...else rules" to do classification. If the set of feature values in a test sample conforms to a set of rules, then the sample is classified as belonging to the class associated with that set of rules.

The classification rules can be built either by directly extracting them from the data (Direct method) or from other classification models (Indirect Method). These classifiers are very expressive in their results as the rules learnt by them can help understand the underlying mechanics of the classification. Some of the rules generated by one of the classifiers in Exp-1 "Fault Type Detection with Feature Selection" are tabulated in Table 7.4. Four Rule-based classifiers used in this work are explained below:

1. **Decision Table**: DecisionTable is a rule-based classifier that evaluates feature subsets using best-first search cross-validation. A decision table maps conditions against actions. Actions are chosen depending on the conditions which are satisfied. The information presented by decision tables can be represented either as decision trees or as if-then-else rules. Decision tables are popularly used for multi-label classification as they are easy to interpret and understand.
2. **JRip**: JRip is an implementation of the RIPPER (Repeated Incremental Pruning to Produce Error Reduction) algorithm that builds association rules using reduced error pruning (REP). First a rule is added with conjuncts to improve the information gain, then the rules in the set are pruned using incremental REP. This is repeated until the criteria for discretion length is achieved. The rule set is then optimized [173] .
3. **OneR**: OneR (one rule) classification algorithm follows the direct method where rules are extracted directly from data. For each predictor in the data it generates one rule.

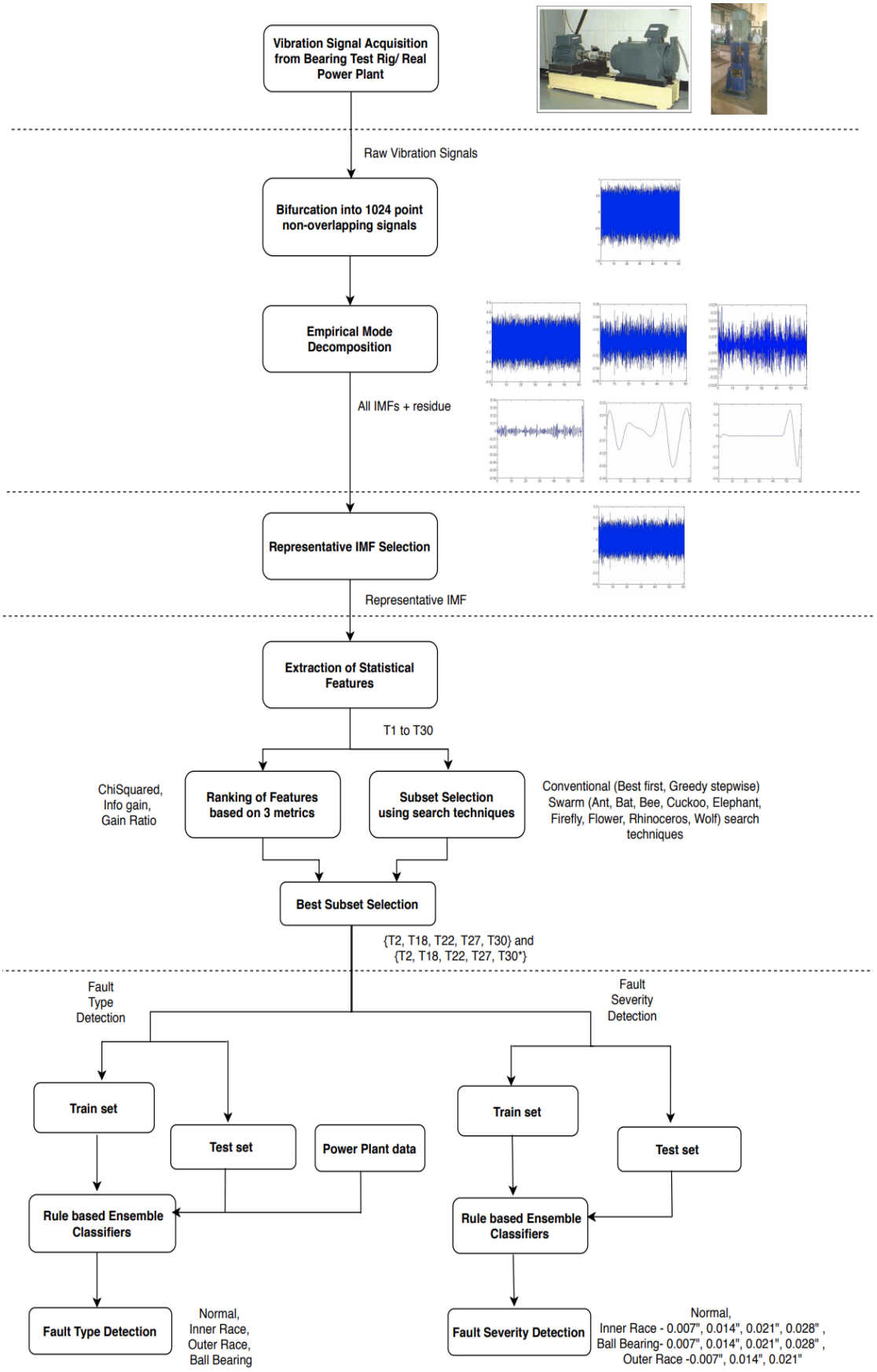


Figure 7.1: Flowchart of proposed Methodology I for vibration signal preprocessing, feature subset selection, fault type & fault severity classification using ML Classifiers

4. **PART**: PART is an efficient and accurate rule based algorithm based on partial decision trees. It follows the indirect method where rules are extracted from other classification models, decision trees in this case. PART builds rules in a recursive fashion by combining the features of divide-and-conquer and rule learning strategies. It forms a rule, then removes the instances that rule covers and again creates a rule for remaining instances, and these steps repeat till the number of remaining samples becomes zero [174] .

7.2.2 Rough-Set Based Classifiers

The Rough set is composed of two crisp sets, out of which one represents the lower boundary and second represents the upper boundary of target set. Following three classification algorithms based on rough set theory are used in this research:

1. **RoughSet Classifier**: This is a rule based classifier whose rule induction is based on rough sets theory. Its classification algorithm is based on the principles of discernibility matrix, reducts and rules generated from reducts. A discretization method (e.g. equal width, equal frequency static or dynamic entropy minimization), a type of discernibility matrix and an algorithm generating reducts (e.g. local/global and all/partial) is selected and then the classifier computes a set of decision rules.

   If "O" is the object to be classified, then the classifier calculates the vote of each decision class and the decision with the greatest vote is assigned to "O"

$$vote_i(O) = \sum_{\{q \to (s1,s2...,sm) \in Rules: O \text{ matches } q\}} s_i.support(q \to (s1,s2...,sm)) \qquad (7.1)$$

   The decision with greatest vote is assigned to the object "O" [96], [175].

2. **LocalKnn Classifier**: The second classifier LocalKnn is the extended version of the KNN classifier that utilizes local metric induction for each classified object. It applies a global metric to find a large set of neighbours and then generates a local metric to select k nearest neighbours. Then it generates a new, local metric from this large set of neighbours.The selected k nearest neighbours are used to vote for the decision. This classifier can handle large data sets reasonably well and gives better accuracy if data contains nominal attributes [96], [175].

3. **RseslibKnn Classifier**: This is a modified K nearest neighbours classifier that provides variety of distance measures and has built-in feature to compute optimal value of k. It can classify large datasets because of its very fast neighbour search algorithm. In the learning phase a distance measure is induced from a training set and an indexing tree is built to achieve fast neighbour search. The algorithm can also learn the optimal value of k from the training set. The classifier provides two distance metrics for nominal attributes: Hamming distance (HD) and Value Difference Metric (VDM). There are three metrics for numerical attributes: The City-block distance (CBD), Density-Based Value (DBVDM) and Interpolated Value Difference Metric (IVDM). Weights computation in the distance measure is done by three methods: method using perceptron, method based on distance and method based on accuracy. The classifier applies induced distance measure to find k nearest neighbours in the training set and applies one of the three methods of voting (equally weighted, with inverse square distance weights or with inverse distance weights) to get the decision by the neighbours. [96], [175].

A research involving rough set based classifiers with CWRU dataset was performed. Though the research proved that rough set based classifiers are not bad in performing fault diagnosis, the results derived from the experiments were not promising enough in the presence of other more convincing classifiers and hence these classifiers were dropped from the case studies.

7.2.3 Ensemble of Classifiers

There are many ways by which base classifiers can be combined together to generate Ensemble Classifiers, which are proven to outperform any single classifier within the ensemble. Ensemble methods combine the predictions from multiple models and have emerged as a powerful machine learning technique for improving the accuracy and robustness of classification.

In this research work five ensemble machine learning algorithms are used:

1. **Bagging**: Bagging (Bootstrap Aggregation) is based on estimation of a statistical parameter like mean from numerous random samples of data. From the training data set, numerous random samples are drawn (and later replaced) to train multiple machine learning models. Prediction is made by each model and the results are

averaged to reduce the variance of predictions [176].

2. **Random Forest**: It is an extension of bagging ensemble classifier [177]. Bagged decision trees have a shortcoming that greedy algorithm is used to select the best split point for building trees. Because of this generated trees look quite similar and the variance of the predictions from different bags gets reduced, which ultimately effects the robustness of the predictions. Random Forest decreases the similarity between the bagged trees by disrupting the greedy algorithm during tree generation. This leads to the use of random subset of the input attributes to generate split points.

3. **Boosting**: Boosting ensemble method uses machine learning models in succession to boost the prediction outcomes by removing the errors in predicted outcomes by previous models.

   (a) **AdaBoost**: It uses decision tree models having a single decision point. The construction of first model done by weighing each instance in the training dataset and continuously updating the weights based on the overall accuracy of the model. Next models are trained and added in succession. The process continues until no further improvements are possible [178].

   (b) **LogiBoost**: It performs classification using additive logistic regression scheme and can handle multi-class problems.

   (c) **MultiBoost**: It is a combination of AdaBoost and an improved version of bagging called wagging. MultiBoost uses base learning algorithm C4.5 to generate decision tree. It combines the high bias and variance reduction properties of AdaBoost with excellent variance reduction property of wagging [179].

4. **Voting**: Voting works by taking two or more sub-models for making predictions. Finally, the predictions are combined based on some criteria e.g. by taking the average of the predictions. J. Kittler and L. Kuncheva have discussed Voting algorithm in detail in their works [180, 181] respectively.

5. **Stacking**: An extension to voting ensembles is stacking. In this ensemble method multiple sub-models are selected and instead of taking the average of predictions another supervisor model is trained to combine the predictions from the sub-models to give best outcome.

7.2.4 Evaluation Metrics

Following evaluation metrics are used to compare the performance of various classifiers:

1. **Accuracy**: It is defined as the ratio of correctly classified instances (i.e. sum of True Positive and True Negative instances) to the total number of instances.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7.2}$$

   where, TP - True Positive, FP - False Positive, FN - False Negative and TN - True Negative

2. **Mean Absolute Error(MAE)**: Sum of absolute errors for all instances divided by the number of instances is called mean absolute error.

$$MAE = \frac{1}{N}\sum_{i} |\hat{x}_i - x_i| \tag{7.3}$$

   where $\hat{x}_i$: predicted label, x_i: true label, N: number of instances.

3. **Matthews Correlation Coefficient (MCC)**: It is a correlation coefficient that indicates the correlation between predicted class and actual class. It can be calculated mathematically using TP, FP, TN and FN values as:

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TN + FN)(TN + FP)(TP + FN)}} \tag{7.4}$$

4. **F1 Score**: It is calculated by taking the harmonic mean of the precision and recall. Given a threshold value the F1 score provides a measure for goodness of classifier.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{7.5}$$

5. **Mean Scheme Entropy**: This is the entropy per instance for the classification scheme. The cross-entropy for the classification model across the entire training dataset is required to be minimized. So this is calculated by calculating the average cross-entropy across all training examples.

Cross-entropy for classification of multi class problem can be calculated as:

$$-\sum_{l=1}^{C} y_{o,l} log(p_{o,l}) \tag{7.6}$$

Where C: number of classes, l: class label, o: observation and p: predicted probability for observation o of class l and y: binary indicator

6. **Area under the ROC curve (AUROC)**: This metric calculates the area under the ROC curve, which is a graph plotted between sensitivity and (1-specificity) of a classifier.

7. **Time taken to build model**: This is the time elapsed since the training of the classifier model started and before it ended.

7.3 FAULT DIAGNOSIS USING MFPT BEARING DATA

The aim of this experimental study is to investigate Ensemble classification approach for rolling element bearing fault diagnosis. The proposed approach has been validated on Machinery Failure Prevention Technology (MFPT) dataset using six simple conventional statistical features. These include two Histogram features: Histogram Upper Bound (HU) and Histogram Lower Bound (HL); and four Moments: 1st moment (arithmetic mean, AM), 2nd moment (variance, VAR), 3rd moment (skewness, SK) and 4th moment (kurtosis, KU). Description of these features is given in Table 5.2.

7.3.1 Experimental Description

For validation and comparative evaluation of various ensemble classifiers, data samples of size 2048 points are extracted from the vibration signals obtained from MFPT dataset [134], giving a total of 1423 samples consisting of 429 normal (healthy) and 994 (faulty). Out of 994 samples obtained through faulty bearing, 497 samples are with fault in inner race and 497 samples are with fault in outer race of bearing. This distribution is shown in Table 3.7. The 1423 signals are first individually decomposed into a set of IMFs by applying EMD algorithm. The best IMF, i.e. the IMF having the highest correlation with the parent signal, is referred to as the representative signal. Only representative signals are considered for the remaining steps.

To increase the number of instances in the dataset in a well-balanced manner, Synthetic Minority Oversampling Technique (SMOTE), developed by N. Chawla, Bowyer K.W. et al. is used [182]. This statistical technique increased the number of instances in the dataset in a balanced way by generating new instances from originally existing minority cases in the training dataset. The amount of SMOTE percentage taken is 25% and number of nearest neighbours is taken as 5. The algorithm took samples and its 5 nearest neighbours for each fault class, and combined features of the target case with features of its neighbours to generate new cases.

Proposed ensemble classifiers are used to diagnose three types of conditions in rolling element bearings: normal (faultless), fault in inner race and fault in outer race and performance evaluation is done using 7 evaluation metrics; Training accuracy, Test accuracy, Mean Absolute Error (MAE), Matthews Correlation Coefficient (MCC), F1-Score, Mean Entropy and Area under the ROC curve (AUROC).

The training data is fed to Weka's package for Meta classifiers. In bagging, a standard decision tree based on reduced error pruning (REPTree) is configured as the model being bagged. REPTree builds a decision tree using either information gain or variance and prunes it by reduced-error pruning. The size of each bag is taken as 100% to generate a new sample of different composition but of same size as that of the training dataset. The 100 number of iterations are performed on the dataset. Keeping all these parameters same for Random Forest classifier, the model is trained on training data set. In all of three boosting models i.e. AdaBoost, Logiboost and Multiboost, weak learner is chosen as REPTree algorithm and total 10 number of iterations are performed.

In Vote ensemble four classification model that can make uncorrelated predictions are selected. The selected sub models are KNN. PART, Logistic Regression and Random Forest. For combining the predictions of the sub models the parameter chosen is average of probabilities. In the proposed stacked ensemble classifier, four sub models are chosen; one lazy classifier- KNN, one rule based classifier- PART, one function classifier - Logistic Regression and Random Forest which is a tree based classifier. The supervisor model taken is Multilayer Perceptron which is trained to combine the predictions from the sub models in the best possible way. It uses backpropagation to classify instances.

7.3.2 Results and Discussion

In this experimental study, the statistical significance of ensemble classifiers is validated on MFPT dataset. Seven metrics are used to compare the efficacy of the models. The classification accuracy on Training dataset is estimated by stratified 10-fold cross validation. In every trial, a classifier is trained on any 9 folds and validated on the remaining fold. For each classifier, Training & Testing accuracy and time required to build classifier model are recorded in Table 7.1. and graphically plotted in Figure 7.2.

Table 7.1: Accuracy and time taken for model building for the different classifiers.

Classifier	**Accuracy**		**Time taken for model building**
	Training	**Testing**	
Random Forest	94.9378%	91.5789%	0.15 seconds
Bagging	94.9378%	91.9298%	0.13 seconds
AdaBoost	93.7759%	90.8772%	0.06 seconds
Logiboost	94.1909%	90.5263%	0.57 seconds
Multiboost	94.2739%	91.5789%	0.05 seconds
Voting	94.8548%	92.2807%	0.26 seconds
Proposed Stacked Ensemble	95.1867%	92.6316%	4.20 seconds

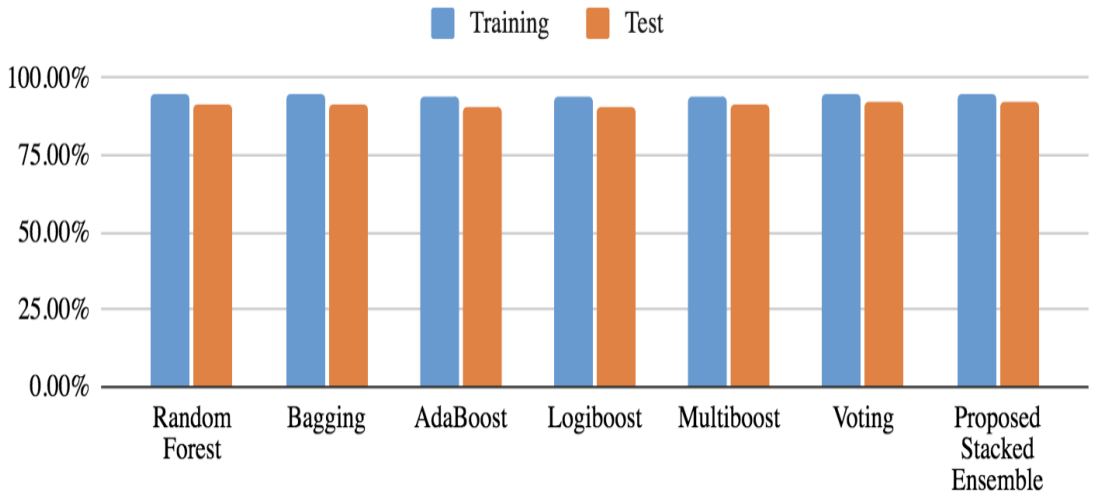


Figure 7.2: Comparison of Training and Testing Accuracies of Ensemble Classifiers

Following observations and conclusions are made:

i) From Table 7.1 and Figure 7.2., it is observed that the performance of proposed stacked ensemble classifier is the best with respect to accuracy although the time taken for building this ensemble classifier model is highest. The 10-fold cross validation accuracy obtained with training dataset is 95.1867% white accuracy obtained with test dataset is 92.6316%. Although with reference to MAE Multiboost gives better performance but its accuracy is very low as compared to proposed classifier.

ii) Five additional evaluation metrics (MAE, MCC, F1-Score, Entropy, AUC) are calculated on Test dataset and are recorded in Table 7.2. From Table 7.2, It is inferred that Proposed ensemble classifier performs decently well in this metric too. It outperforms other ensembles in MCC with the highest value of 0.927. As MCC value "1" indicates a perfect classifier it can be said that proposed classifier is performing very well.

Table 7.2: Metric values obtained for the different classifiers on Test dataset.

Classifier	**MAE**	**MCC**	**F1 Score**	**Mean Entropy**	**AUROC**
Random Forest	0.0704	0.874	0.916	1.53	0.974
Bagging	0.0738	0.879	0.920	0.23	0.968
AdaBoost	0.0645	0.862	0.909	0.42	0.955
Logiboost	0.0817	0.858	0.906	0.27	0.965
Multiboost	0.0557	0.874	0.916	0.84	0.963
Voting	0.0737	0.884	0.916	0.22	0.978
Proposed Stacked Ensemble	0.0656	0.927	0.923	0.25	0.978

iii) As F1-score value 1 indicates perfect precision and recall and the proposed model gives a F1-score of 0.923, it is verified that the model is a good model that correctly distinguishes the three bearing fault classes.

iv) Often, choosing the best model is a trade-off between sensitivity and specificity. Therefore it is best to have some metric that captures both of these aspects. This is

effectively captured by the Receiver operating Characteristics (ROC) curve. The area under the ROC curve (AUROC) is a good indicator of a model's performance. In addition to this, AUROC score also considers the rank of each prediction instead of its absolute value, hence it is independent of the threshold set for classification. The AUROC value for the proposed model is 0.978, highest amongst all classifiers. The proposed classifier also shows remarkably good mean entropy value.

This experimental study gave a research direction that effective fault sensitive features extracted from vibration signals can serve the purpose of vibration signature analysis, if used along with efficient and robust ensemble machine learning classifiers.

7.4 FAULT DIAGNOSIS USING CWRU BEARING DATA

For the experimental validation of proposed methodology of "Optimal Statistical Feature Subset Selection for Bearing Fault Detection and Severity Estimation", Case Western Reserve University (CWRU) dataset [133] was used and following five experiments were performed using Ensemble classification approach. Rule Based Classifiers, PART and JRIP, were used with ensemble techniques boosting (Multiboost in this case) and bagging for performing fault type detection of the 4 classes- Normal, Inner Race, Outer Race and Ball Bearing. The ensemble techniques make use of a single learning algorithm to train multiple models on the dataset and then average out these models to get a final prediction. While bagging helps in getting an ensemble model with less variance by assigning equal weights to individual models for averaging, boosting works by reducing both variance and bias. Both of them help in increasing the stability of the models.

For training PART, the value of pruning confidence factor was taken as 0.25, while in JRIP, the minimum of the instances in a rule was taken as 2 and two optimization runs were performed. These classifiers were trained on the reduced dataset obtained by only retaining those features that were included in the optimal feature subset, as discussed in the previous chapter. To further analyze the efficiency of the proposed methodology, these classifiers were also run on the entire dataset with all the features in order to compare the results with those obtained with reduced dataset. The data augmentation achieved by the initial splitting of vibration signals into signals of 1024 length resulted in a large enough dataset, removing the need to perform cross validation while training

the classifiers. For implementing Bagging and Multiboost, 10 iterations were performed with a batch size of 100.

7.4.1 Experiment-1: Fault Type Detection using Optimal Feature Subset and Ensemble Rule Based Classifiers

In this experiment CWRU Dataset A is used for bearing fault type detection using optimal subset of 5 features (T2, T18, T22, T27, T30) derived in Chapter 6 and Ensemble Rule Based Classifiers. Table 7.3 reports the performance of the classifiers in terms of accuracy obtained on the train and test dataset. Two cases have been considered -

i) When the selected feature set contains T2, T18, T22, T27, T30 corresponding to column 1 of Table 7.3 and
ii) When T30 is replaced with its Gaussian Mixture Model equivalent T30*, the feature set in this case being T2, T18, T22, T27, T30* corresponding to column 2 of Table 7.3.

Table 7.3: Classifier performance in terms of Train accuracy and Test accuracy values for Exp-1.

Classifier	**Train Accuracy**		**Test Accuracy**	
	T30	**T30***	**T30**	**T30***
PART	98.11	98.36	95.43	95.43
PART + Multiboost	**100**	**100**	**96.51**	**96.75**
PART + Bagging	98.72	98.81	96.49	96.72
JRIP	96.15	96.41	94.28	94.37
JRIP + Multiboost	99.37	99.99	96.22	96.37
JRIP + Bagging	98.10	98.21	96.04	96.25

It is observed that PART when used with Multiboost outperforms all the 5 classifiers and gives the best performance with an accuracy of 96.51% on test dataset and 100% on the train dataset. The performance of the features is further studied by inspecting the rules generated by them. Rules are propositional formulae consisting of one or more atomic formulae (enclosed in parenthesis as shown below) that are joined together using logical connectives AND $\wedge$ and IMPLIES $\rightarrow$. The number at the end of a rule indicates the number of instances correctly classified by that rule. PART with

Multiboost generates a total of 87 rules from the 5 important features. With just a small number of rules the entire training data of 13648 samples is fully and accurately classified. Some of the rules generated are listed in Table 7.4.

Table 7.4: Examples of Rules generated by PART+ Multiboost classifier

Rule No.	**Rule**	**No. of Instances classified**
R1	$(T22 > 0.358398)^{\wedge}(T27 \leq 0.318614)^{\wedge}(T18 > 0.025684)^{\wedge}(T2 \leq 0.477539) \rightarrow InnerRace$	660
R2	$(T22 > 0.416992)^{\wedge}(T18 > 0.439639) \rightarrow BallBearing$	375
R3	$(T22 > 0.416992)^{\wedge}(T18 > 0.086578)^{\wedge}(T2 > 0.333767)^{\wedge}(T18 \leq 0.177064) \rightarrow OuterRace$	350
R4	$(T2 \leq 0.050631)^{\wedge}(T22 > 0.291016)^{\wedge}(T22 \leq 0.357422) \rightarrow Normal$	1635

From these rules, it can be observed that a total of 1635 out of 2649 dataset instances were correctly classified as Normal using just one rule R4 generated by the features T2 and T22. Each of the three remaining rules also classify significant portions of samples in the dataset, thus showing the good performance of the features forming these rules. The rules are also very simple in structure, for instance, R2 constructed with the help of just 2 atomic formulae, correctly classifies 375 Ball Bearing instances. The ability to easily distribute the samples into the 4 classes using simple rules shows the efficiency of the selected time domain features.

Since the number of samples belonging to Normal class is 2649, which is less than the number of samples in each of the faulty classes (around 3500 for each), the dataset is slightly class unbalanced. Hence, evaluating the performance of the rule based classifiers on accuracy alone may not be comprehensive and therefore the performance of the classifiers on the test dataset has been further studied by calculating two other metrics: F-measure and Geometric Mean (G-Mean).

Table 7.5 reports the performance of the classifiers on Dataset A using F-measure and G-Mean scores. The F-measure and G-Mean scores for all the classifiers are significantly high, in particular, PART + Multiboost classifier outperforms the other classifiers. Hence, the rule based classifiers are able to classify the slightly unbalanced dataset very well, in turn showing the good performance of the selected features.

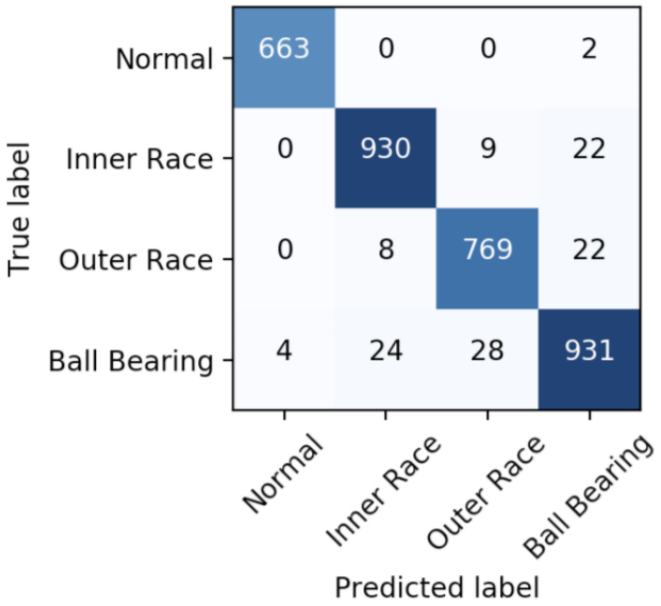
Table 7.5: Classifier performance in terms of F-Measure and G-Mean values for Exp-1.

Classifier	**F-measure**				**Geometric Mean (G-Mean)**			
	Normal	**Inner Race**	**Outer Race**	**Ball Bearing**	**Normal**	**Inner Race**	**Outer Race**	**Ball Bearing**
PART	0.990	0.955	0.949	0.932	0.994	0.970	0.969	0.949
PART + Multiboost	**0.995**	**0.967**	**0.958**	**0.948**	**0.997**	**0.977**	**0.974**	**0.962**
PART + Bagging	0.994	0.966	0.957	0.946	0.995	0.976	0.973	0.961
JRIP	0.987	0.931	0.929	0.934	0.992	0.958	0.947	0.951
JRIP + Multiboost	0.993	0.965	0.955	0.944	0.995	0.972	0.973	0.961
JRIP + Bagging	0.991	0.958	0.957	0.944	0.995	0.971	0.973	0.960

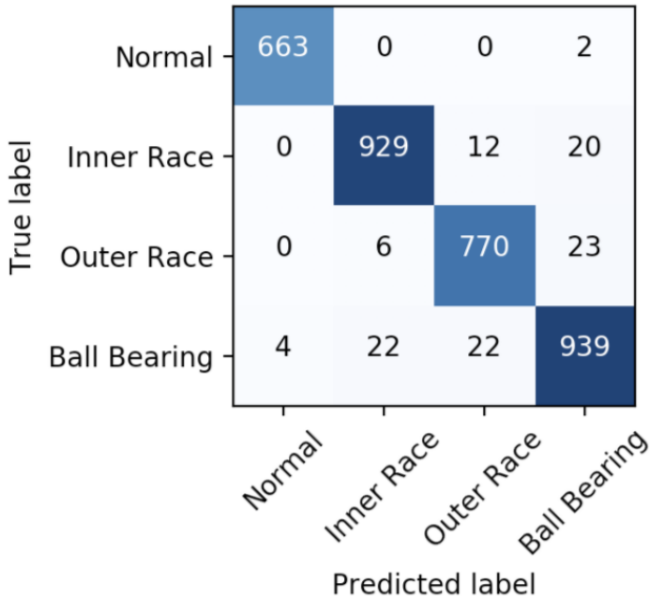
To further validate the good performance of the selected features in fault type detection, confusion matrices were built as shown in Figure 7.3. Figure 7.3.(a) shows the confusion matrix obtained on test dataset with T30 parameter for PART with Multiboost and Figure 7.3.(b) shows the confusion matrix obtained on test dataset with T30* parameter for PART with Multiboost classifier. From the first confusion matrix, it is observed that the Normal instances were accurately classified (with only 2 misclassifications), further confirming they are well classified by the 5 important features selected in this research even though Normal class is slightly underrepresented. Another important observation is that most of the misclassifications involve Ball Bearing class. 22 instances each of Inner Race and Outer Race got misclassified as Ball Bearing. Also, a total of 56 Ball Bearing instances were incorrectly classified.

To validate these observations, 3 features - T22, T27, T30 - were plotted pairwise using scatter plots as shown in Figure 7.4. It can be seen that for Normal class, the plots show mostly non-overlapping clusters, hereby confirming that the selected features are playing a very good role in classifying fault vs. no fault. The plots also show a lot of overlapping between Ball Bearing, Inner Race and Outer Race clusters, confirming the observation about Ball Bearing class. This indicates a lot of diversity in the Ball Bearing instances that is unhandled by the selected features.

Hence, features which can accurately distinguish this class from other classes are needed. A modification in feature T30 (Normal Negative log likelihood for Single Gaussian), resulting in a better classification for ball bearing instances, has been discussed in the results of Experiment 5.

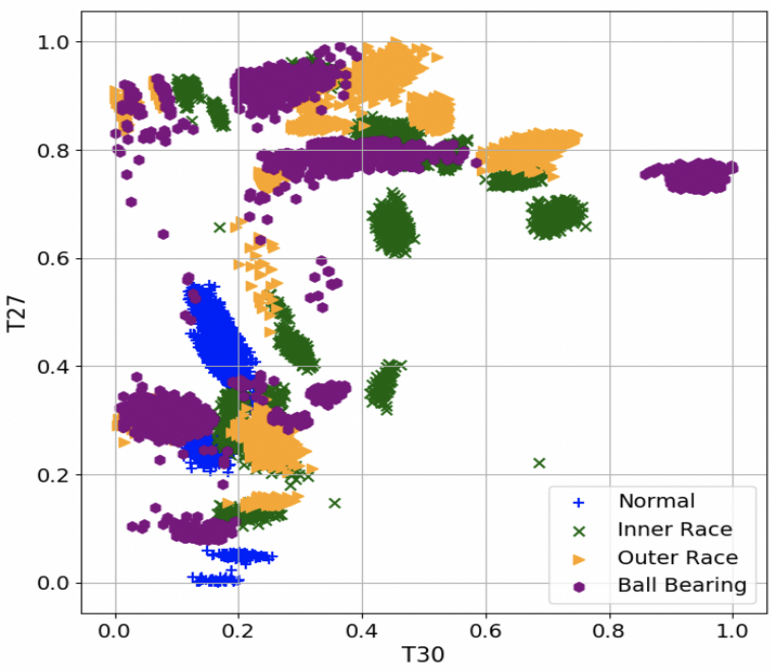


(a) with T30.

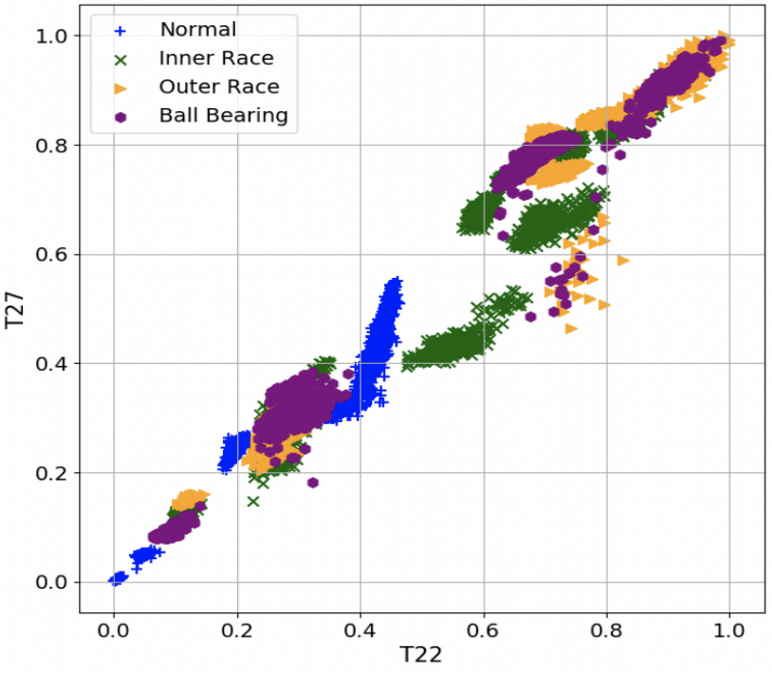


(b) with T30*

Figure 7.3: Confusion matrices for Exp-1 using Multiboost + PART on Test dataset

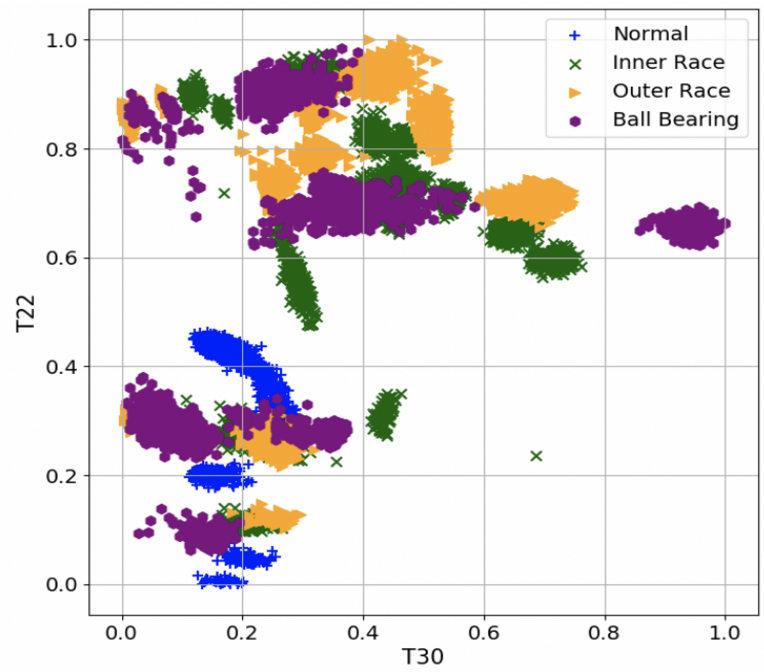


(a) Between T27 and T30.



(b) Between T27 and T22.

Figure 7.4: Pairwise Scatter plots for the top 3 ranked features. Here T22, T27 and T30 are Rate of zero crossing, Hjorth parameter Mobility and Normal Negative log likelihood for Single Gaussian respectively



(c) Between T22 and T30.

Figure 7.4: Pairwise Scatter plots for the top 3 ranked features. Here T22, T27 and T30 are Rate of zero crossing, Hjorth parameter Mobility and Normal Negative log likelihood for Single Gaussian respectively (Contd.)

7.4.2 Experiment-2: Fault Type Detection using 30 Features and Ensemble Rule Based Classifiers

In this experiment CWRU Dataset A is used for bearing fault type detection using all 30 features i.e. feature selection is not performed. Table 7.6 reports the performance of the classifiers in fault type detection in terms of accuracy obtained on Dataset A without feature selection.

Table 7.6: Classifier performance in terms of Train accuracy and Test accuracy values for Exp-2.

Classifier	**Train Accuracy**		**Test Accuracy**	
	T30	**T30***	**T30**	**T30***
PART	98.63	98.82	96.04	96.25
PART + Multiboost	**100**	**100**	**99.13**	**99.63**
PART + Bagging	99.47	99.8	98.6	98.81
JRIP	97.71	97.82	96.6	95.96
JRIP + Multiboost	99.8	99.99	97.04	97.27
JRIP + Bagging	98.82	98.9	97.04	97.25

These classifiers perform better than their feature selected counterparts, in terms of accuracies obtained. Again, in terms of overall performance, PART + Multiboost outperforms all the 5 classifiers, giving the highest train and test accuracies of 100% and 99.13% respectively. On comparing these with the train and test accuracies of 100% and 96.51% obtained with feature selection, it is observed that on the train data the selected features were alone sufficient for attaining a perfect accuracy, while in the case of test data too, a major contribution in achieving the high accuracy can be credited to the selected features.

7.4.3 Experiment-3: Fault Severity Detection using Optimal Feature Subset and Ensemble Rule Based Classifiers

This experiment detects the severity of fault types by performing a 12 class classification using Rule Based Classifiers on the selected 8 features obtained for Dataset B [Table 3.4]. The 12 labels are used for severity detection, where the severity of a fault has been characterized by its size in inches.

Table 7.7 shows the results obtained wherein, once again PART, when used with Multiboost, outperforms all the 5 classifiers and gives the best performance accuracy of 100% on train dataset and 97.92% on test dataset. The results obtained with the parameter T30* are discussed later in Experiment 5 results.

Table 7.7: Classifier performance in terms of Train accuracy and Test accuracy values for Exp-3.

Classifier	**Train Accuracy**		**Test Accuracy**	
	T30	**T30***	**T30**	**T30***
PART	96.1	96.14	94.42	95.18
PART + Multiboost	**100**	**100**	**97.92**	**98.56**
PART + Bagging	97.9	97.96	97.41	97.53
JRIP	91.22	91.83	90.99	91.9
JRIP + Multiboost	96.1	96.29	96.03	96.15
JRIP + Bagging	95.63	95.94	95.35	95.82

Having few IR 028 and BB 028 samples and almost thrice the count for other severity labels, the dataset is unbalanced class wise and hence the F-measure and G-Mean values were obtained to further investigate the performance of the selected

features. Table 7.8 reports the F-measure and G-Mean values obtained with PART + Multiboost for fault severity detection. It is observed that both the underrepresented classes and the remaining classes are accurately classified. Hence, the selected features for fault severity detection are performing well despite the data being class unbalanced.

Table 7.8: PART + Multiboost performance in terms of F-Measure and G-Mean values for Exp-3.

Metric	**Norm -al**	**IR 007**	**IR 014**	**IR 021**	**IR 028**	**OR 007**	**OR 014**	**OR 021**	**BB 007**	**BB 014**	**BB 021**	**BB 028**
F-Meas-ure	0.998	0.988	0.982	0.980	1.00	0.998	0.965	0.981	0.962	0.951	0.953	1.00
G-Mean	0.999	0.994	0.987	0.993	1.00	1.00	0.984	0.992	0.976	0.972	0.968	1.00

7.4.4 Experiment-4: Fault Severity Detection using 30 Features and Ensemble Rule Based Classifiers

In this experiment Dataset B, with all the 30 statistical features, was fed to the rule based classifiers to again investigate fault severity detection. Results for this experimental study are shown in Table 7.9. On comparison with Table 7.7 results, it is observed that the highest train accuracy of 100% remains unaltered while the test accuracy sees a gain of 0.76%. The perfect train accuracy and the small gain in test accuracy shows the superior performance of the selected eight features over the other features, confirming their large contribution towards the accurate classification results. This is in line with the observation for the selected features for fault type detection in Experiment 2.

7.4.5 Experiment-5: Replacing Normal Negative Log-Likelihood for Single Gaussian Parameter with its Gaussian Mixture Model Variant for Experiments 1-4

In feature ranking results, the feature T30: Normal Negative Log-Likelihood for Single Gaussian comes in the set of important features for both fault type and fault severity detection [Section 6.2.2]. For instance, in the case of fault type detection, it occupies rank 3 in the list of average ranks given by feature ranking methods and is

Table 7.9: Classifier performance in terms of Train accuracy and Test accuracy values for Exp-4.

Classifier	Train Accuracy		Test Accuracy	
	T30	**T30***	**T30**	**T30***
PART	98.01	98.15	95.68	95.71
PART + Multiboost	**100**	**100**	**98.68**	**99.12**
PART + Bagging	99.36	99.39	97.5	97.56
JRIP	93.92	94.61	93.77	94.21
JRIP + Multiboost	98.78	98.98	96.32	96.59
JRIP + Bagging	97.73	97.79	96.08	96.35

selected by all the 5 feature subset methods. This feature gives the negative of the likelihood that a normal distribution of given mean and standard deviation fits the sample data, in this case, the sample data being the representative IMF of bearing vibration signal. Experiments 1-4 were repeated by replacing the Normal Negative Log-Likelihood for Single Gaussian parameter (T30) with its proposed Gaussian Mixture Model variant (T30*).

Table 7.3 shows the results obtained for fault type detection when T30 was taken as one of the 5 selected features and when it was replaced by T30*. According to Table 7.3, PART with Multiboost gives an accuracy of 100% on train dataset for both T30 and T30* and shows a slight jump in test accuracy from 96.51% with T30 to 96.75% with T30*. For the rest of the classifiers, a slight improvement in train and test accuracy with T30* over T30 is seen. From the two confusion matrices in Figure 7.3, the number of true positives for the Ball Bearing class has increased from 931 with T30 to 939 with T30*, accompanied by a decrease in the missclassified instances of Ball Bearing into Inner Race and Outer Race classes from 24 to 22 and 28 to 22 respectively. This further confirms the better performance of T30* as compared to T30. Similarly, the accuracy values obtained in Table 7.6 shows better accuracy values with parameter T30* than the values obtained with T30 for all the classifiers. This again shows the effectiveness of T30* over T30.

A similar trend is observed for fault severity detection. From Table 7.7 it is seen that on comparing T30 with T30*, the train accuracy remains at its peak of 100% while test

accuracy increases by 0.64% to become 98.56%. T30* performs better than T30 in the case of without feature selection as well, as can be seen from Table 7.9.

A more clear picture is provided by the confusion matrices in Figure 7.5, where it can be seen that for T30, the misclassifications are more than that for T30*. It is believed that for bearing datasets having a majority of vibration signals with multiple peaks, this feature can prove to be a key performer because of its ability to represent such signals well. Hence, Normal Negative log likelihood of GMMs has the potential of being a more accurate and reasonable parameter for feature selection and bearing fault diagnosis than the Normal Negative log Likelihood of Single Gaussian.

7.4.6 Results and Discussion

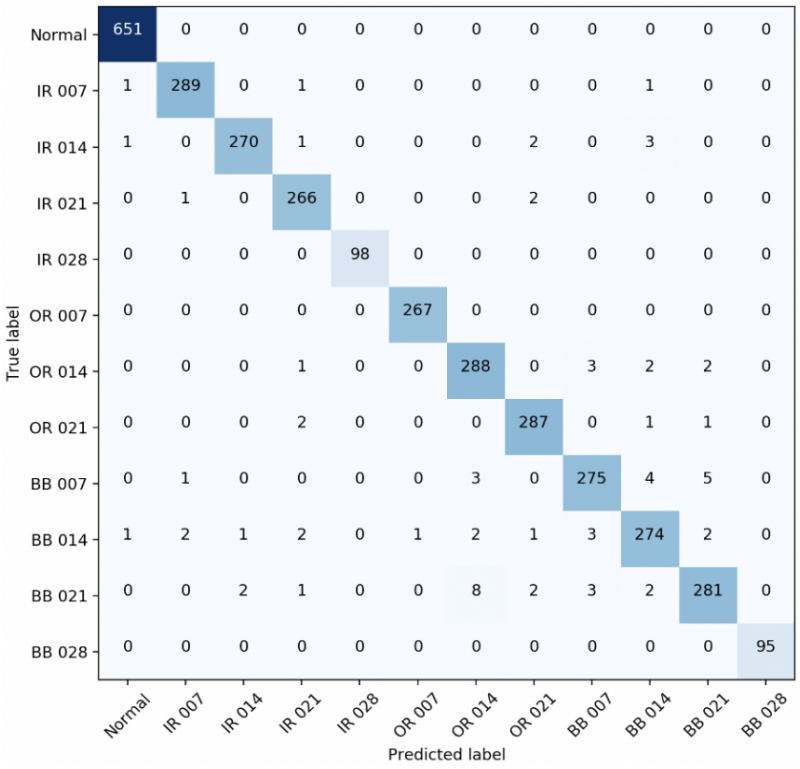
The following points discuss and summarize the obtained results from the five experiments:

i) For the 4-class fault type detection problem, a subset of the features - Root Mean Square (T2), Geometric Mean (T18), Rate of zero crossing (T22), Hjorth parameter Mobility (T27) and Normal Negative log likelihood for Gaussian Mixture Model (T30*) perform well with Ensemble Rule based classifiers, giving the maximum test accuracy of 96.75% obtained with PART + Multiboost.

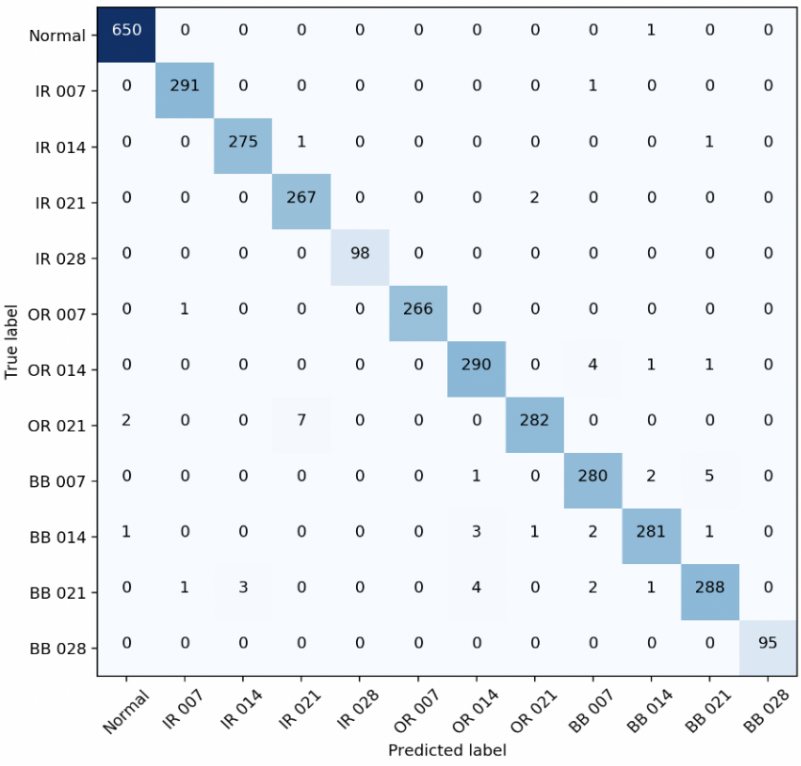
ii) For the 12-class fault severity estimation problem, a subset of the features - Root Mean Square (T2), 75th Percentile (T9), Kurtosis Factor (T17), Geometric Mean(T18), RSSQ (T19), Rate of zero crossing(T22), Hjorth parameter Mobility(T27) and Normal Negative log likelihood for Single Gaussian (T30*) perform very well producing the maximum test accuracy of 98.56% with PART + Multiboost.

iii) An improvement in classification accuracy is observed when Normal Negative log likelihood for Single Gaussian (T30) is replaced by it's GMM variant (T30*). This is because some signals have more than one distinct maxima and a single Gaussian curve cannot model these signals accurately. Thus instead of a single Gaussian curve, it is better to fit these signals with a Gaussian Mixture Model.

iv) The F-measure and G-Mean scores for all the rule based classifiers are significantly high, confirming the potential of the chosen features to accurately classify the unbalanced dataset.



(a) with T30.



(b) with T30*

Figure 7.5: Confusion matrices for Exp-3 using PART+ Multiboost on Test dataset

v) On comparing the results with feature selection vs. without feature selection, the latter performs better for both fault diagnosis and fault severity estimation giving 99.63% and 99.12% test accuracies respectively. Despite this observation, the good accuracy values obtained with the selected features show that a major contribution in achieving the high accuracy in the case of without feature selection can be credited to the selected features. Hence, these selected features play a major role in the diagnostic ability of the classifiers.

7.5 FAULT DIAGNOSIS USING REAL POWER PLANT DATA

In this experiment, bearing data from 500 MW Kosti Thermal Power Plant commissioned by Bharat Heavy Electricals Limited, India [Figure 3.6.] has been used to validate the proposed methodology. Because the bearing configuration of the pump of Kosti Thermal Power Plant matched with that used in CWRU dataset, the test samples acquired from the power-plant are appropriate to validate the trained model.

7.5.1 Experimental Description

Data for normal baseline bearings was recorded using the deep groove ball bearing SKF6303-2Z/C3 with dimensions 17mm X 47mm X 14mm at drive-end with motor speed of 3000 RPM. Vibration data was collected at 12,000 samples/second. Vibration data was taken for 2 different timestamps and for both vertical and horizontal positions, leading to a total of 4 vibration signals. These test samples are mixed with the 20% test set separated from the CWRU Dataset A. The trained model of PART + Multiboost for fault type detection with feature selection, obtained in Experiment 1 of Section 7.4, is chosen for labelling instances as outer raceway defect, inner raceway defect, ball defect, and the normal. The same data pre-processing, including bifurcating the signals to sub signals of length 1024, Empirical Mode Decomposition and feature extraction, is performed on these samples as done for the CWRU dataset.

7.5.2 Results and Discussion

The overall accuracy and average F-measure achieved were 96.75% and 0.972 respectively. Furthermore, per instance classification results were obtained and it was observed that all the normal samples taken from the power plant were correctly

classified. Hence, the results were reasonably good. The mixing of power plant samples with CWRU test samples ensured that the good accuracy was not due to any bias shown by the classifier towards the normal class. This is also confirmed by the good F-measure value. Furthermore, when T30 was replaced by T30* for this dataset, no improvements were seen in the accuracy and F-measure values. This might be due to the mostly Single Gaussian nature of the newly added Normal vibration signals taken from the power plant.

7.6 CONCLUSIONS

So far research in bearing fault diagnosis and severity estimation has majorly focussed on improving the performance of classification algorithms. However, the applicability of these approaches mainly depends upon the quality of features extracted from the bearing signals. In light of this, the research performed in Methodology I focused on optimizing the input features fed to these classification systems with the belief that if new input features are proposed and/or the input features are selected comprehensively using an in-depth analysis, they can contribute more in bearing fault diagnosis, covering all types of fault samples in an exhaustive manner. This can in turn lead to the existing classification methods to perform much better. In an attempt to achieve this, this methodology proposed the usefulness of Hjorth Parameters and Normal negative log likelihood for Gaussian Mixture Model (GMM) for rolling element ball bearing fault detection and severity estimation.

The performance of the feature set was validated by running experiments for different conditions - with and without feature selection - and comparing their results. The dataset being class unbalanced proved that the good performance of the feature set is independent of the class distributions of the dataset. Since in real world scenarios one can never exactly replicate the experimental conditions of a class-balanced dataset, these results guarantee that the selected features are independent of class balanced scenarios.

In addition to CWRU test rig dataset, the performance was further validated on a bearing dataset obtained from an operational power plant. The results show that the methodology is robust, aligns well with the real world situation of class unbalanced datasets and achieves good performance for both fault diagnosis and fault severity

estimation. The results obtained from this methodology can pave the way for researchers to change their current method of developing diagnostic systems by first analyzing the feature set comprehensively and then shifting their focus to the next step of classification.

7.7 CHAPTER SUMMARY

This chapter discusses the last step of the proposed methodology for fault diagnosis using vibration signature analysis using optimal feature subset and ensemble rule-based classifiers. The classifiers used in the proposed methodology are elaborated and the concept of "Ensemble of classifiers" is also presented. The fault classification results obtained for different experimental studies involving different datasets using various classifiers have been compared using different evaluation metrics in order to find the best performing ensemble classifier. The efficacy of the selected time domain statistical features and ensemble classifiers is further confirmed by their good performance in classifying samples taken from an operating power plant.

CHAPTER VIII
FAULT DIAGNOSIS USING DEEP TRANSFER LEARNING

8.1 INTRODUCTION

Researchers have been using different techniques to convert vibration signals into vibration images in order to apply image processing techniques to extract fault sensitive features, which can later be used to efficiently classify these images for detection of faults.

Some of the studies that used different types of vibration images and classification techniques are discussed in literature review in Section 2.4. Literature survey also shows that higher order spectral analysis methods, like Bispectrum analysis, are becoming very popular over traditional methods, like power spectrum, in fault diagnosis of rolling element bearings since higher order spectrum can provide more diagnostic information than power spectrum for vibration signals. In particular, Bispectrum helps retain phase and amplitude information of vibration signals, which is useful for fault detection. The phase retention advantage provided by Bispectrum is used to identify phase coupling effects. Researchers have also explored application of image processing techniques to analyze Bispectrum in the form of images.

With the advancements in deep learning, research in the field of fault diagnosis using vibration images has progressed tremendously. In this chapter, a methodology for bearing fault diagnosis and classification using Bispectrum images of fault signals and Deep Transfer Learning is proposed. This chapter discusses the experimental validation of the proposed methodology. The detailed flowchart of the proposed methodology is given in Figure 8.1.

8.2 IMAGE BASED FAULT DIAGNOSIS USING DEEP LEARNING

Deep Neural Network Architectures, specifically Convolutional Neural Networks (CNN), and Transfer Learning have revolutionized image based fault diagnosis. These concepts along with Bispectral contour maps are explained in the following pages.

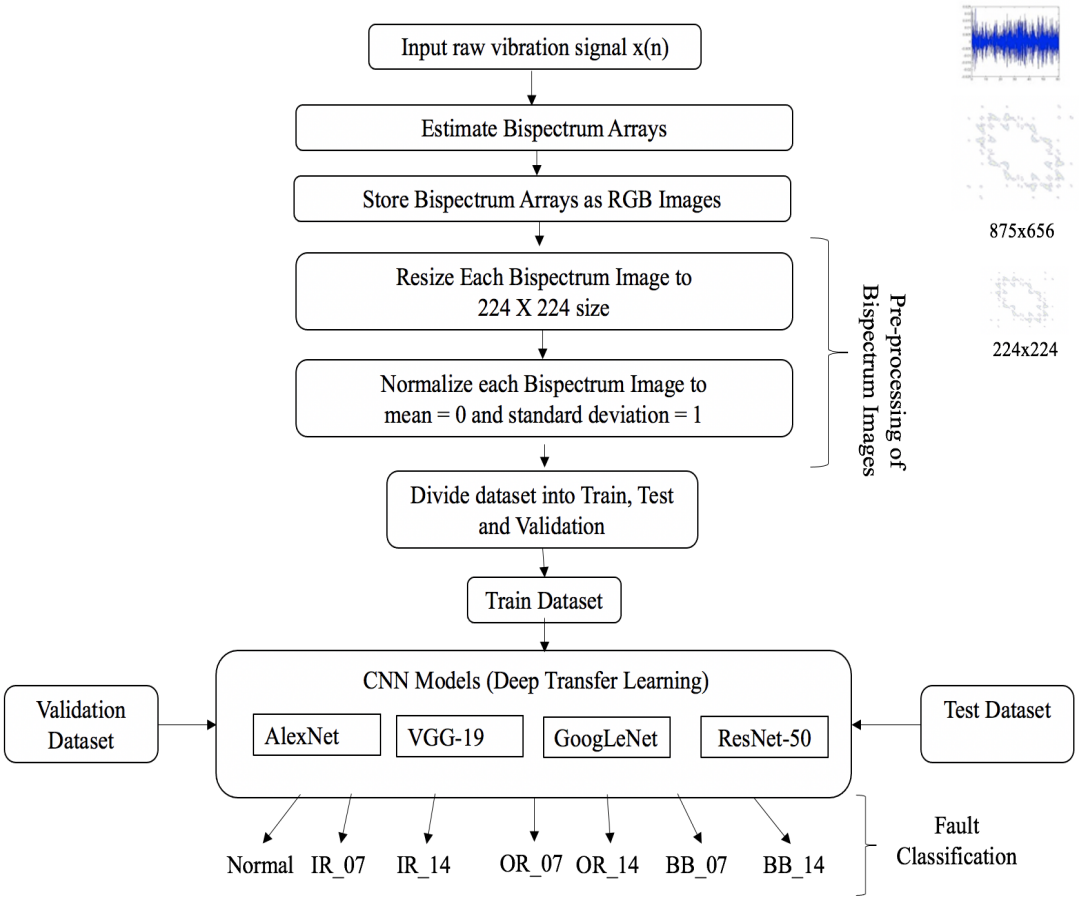


Figure 8.1: Flowchart of Proposed Methodology II for Fault Diagnosis and Classification using Bispectrum Images and Deep Transfer Learning

8.2.1 Bispectral Contour Maps

The first step of the proposed methodology is to generate Bispectral Contour maps i.e. Bispectrum images from the acquired vibration signals. Bispectrum, B(f1, f2), is the 3rd order cumulant and was calculated as given in equation 5.13 in chapter 5. Bispectrum arrays were estimated for each of the data samples using MATLAB HOSA toolbox and stored as RGB images so that deep learning models can learn useful insights from the color intensities. Sample Bispectral Contour Maps for seven classes are shown in Figure 5.3 in chapter 5.

8.2.2 Convolutional Neural Networks and their Architectures

Artificial Neural Networks can be categorized as **Shallow** or **Deep**. Shallow networks have one hidden layer between input layer and output layer, while deep networks have multiple hidden layers between input and output layers. Convolutional Neural Network (CNN), one of the most popular neural network, is a class of Deep

Neural Networks (DNN), mostly used for image classification.

CNNs, at a very high level, are a concatenation of convolutional layers, activation (non-linear) layers, pooling layers and fully connected layers. Different layers of CNN are explained as below:

Convolutional layers – These layers, also the reason why these networks are hence named, are based on the mathematical concept of convolution involving 2- dimensional digital signals or images. Taking the example of 2-D signals, a 2-D mask is run over an image, wherein for every position of the center of the mask, mask weights are multiplied with the corresponding pixel values of image and those products are added to get the result for that position as depicted in Figure 8.2.

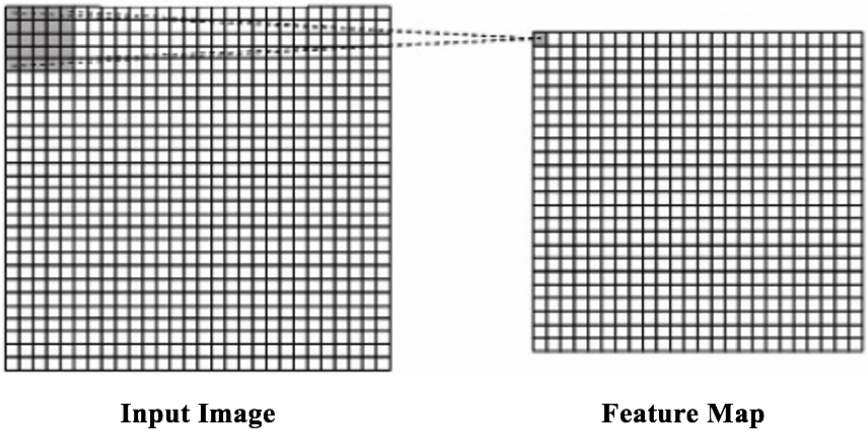


Figure 8.2: Convolution with 5x5 kernel

The output image, also called feature map, is determined by computing these results for all such positions. This simple operation gives the possibility of doing any image processing technique on the input image like image blurring, edge detection and many others, which in turn helps in achieving the final goal of image classification or object detection. The output image, in the case of convolutional neural networks, becomes a part of the input of succeeding hidden layer. The step size by which the filter is moved from one position to the next position is called the stride. The mask is called a kernel or a filter.

Activation layers – These layers are mostly for introducing non-linearity in the network. This is needed for backpropagation of weights during training. In a CNN, activation layers are placed after convolution layers and linear layers. Examples of non-linear activation functions are ReLU, TanH, Sigmoid, Softmax.

Pooling Layers- As their name suggests, these layers pool a set of pixels or data points in the input image to give a single value. This is equivalent to down sampling. Pooling layers perform typical aggregation operations like finding maximum value or average value for a set of data points.

Fully Connected layers – Once the combination of the above 3 layers have achieved the task of finding the important high-level features crucial for classification, the n-dimensional vector generated by these layers needs to be mapped to the most likely class or label. This mapping is performed by fully connected layers. The n-dimensional vector is first flattened to a 1-D vector which is then converted to another 1-D vector of size equal to the number of classes in the dataset. The 1-D vector has score values corresponding to every class. The class having the highest score is the prediction of the network for that input.

Four popular CNN architectures used in this research work are explained below :

i) *AlexNet*

Alex Krizhevsky et al. designed the AlexNet model that won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 by a large margin in classification accuracy. Their work started a new area of hot research focusing on the application of CNNs in computer vision. Alexnet is a deep convolution neural network consisting of 8 layers, of which first five are convolutional layers and last 3 are fully connected layers. The last layer's result is passed through a softmax activation and then finally the output of the model is generated as the probability vector of N classes [183] .

ii) *VGGNet*

VGGNet, introduced by Simonyan et al of the **V**isual **G**eometry **G**roup Lab of Oxford University in 2014. The model was developed out of research for understanding the effect of depth of a CNN in its performance. VGGNet has multiple variants differing in the number of convolution layers in them. All VGG networks have the same configuration of the fully connected layers. In this research work, VGG-19 has been implemented. VGG-19 is 19 layers deep and takes 224x224 RGB input images. The last 3 layers of this model are fully connected layers, with the output of the last layer being fed to a softmax layer at the end. This model has ReLU non-linearity in all hidden layers [184] .

iii) *GoogLeNet*

GoogLeNet is a 22-layer deep architecture based on the deep convolutional neural network architecture codenamed 'Inception' that won the ILSVRC 2014 challenge and set up a new state of the art for classification and detection of ImageNet dataset [185]. This network offers improved utilization of computing resources and was built based on the idea of having multi-sized filters at the same level in a network. In this research work, the pre-trained version for ImageNet has been used to perform transfer learning with an image input size of 224 by 224.

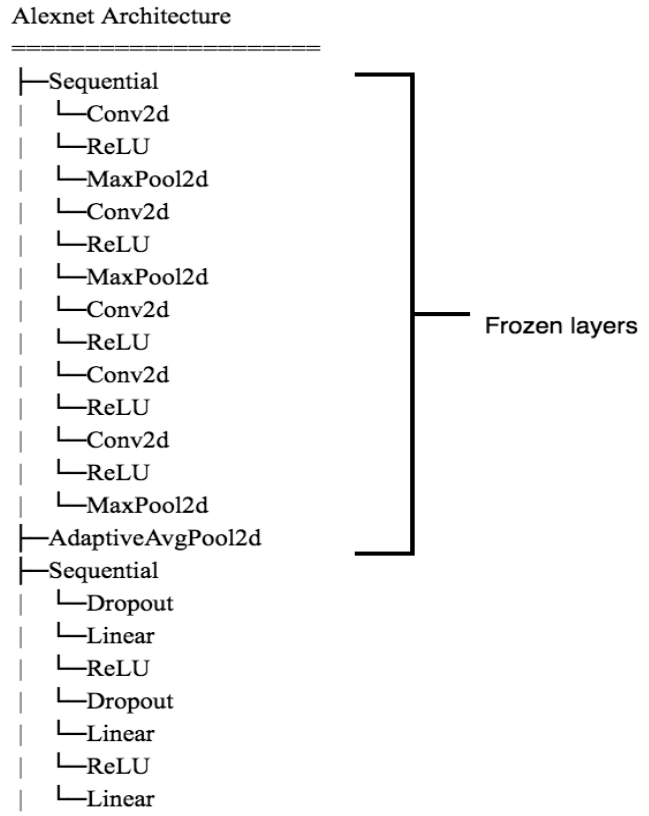
iv) *ResNet*

Residual **net**work (ResNet), introduced by He, Kaiming et al., won the 1st place in the ILSVRC 2015 classification task [186]. ResNet was designed to ease the training process of very deep networks while achieving a lower error rate of classification than state of the art deep CNNs at that time. ResNets were an alternative to the vanishing gradient problem of other deep CNNs. ResNets are different from VGGNet and AlexNet because they have 'identity shortcut connections' that skip one or more layers. In this research work, ResNet-50, which has 50 layers, has been implemented.

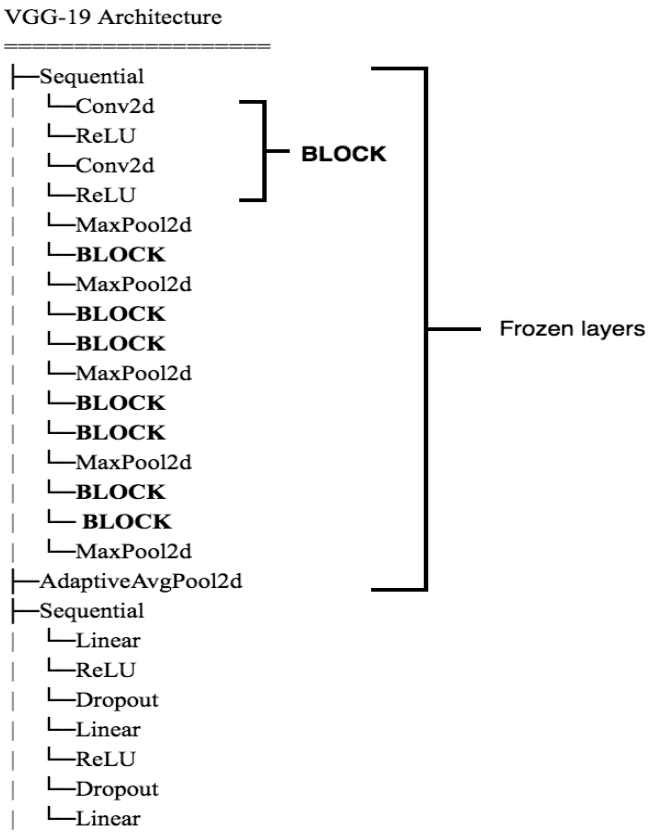
Figure 8.3. shows the architectures of models. The layers that were kept frozen while training have been marked. All of the models were implemented in Python using the Pytorch framework. The architectures have been shown as they are defined in the Python library for easy understanding. In the figure, BLOCK, Inception Block, Bottleneck-1 and Bottleneck-2 represent sets of layers that are repetitive in the model architecture.

8.2.3 Transfer Learning

Transfer learning is a technique that utilises what has been learnt in one setting, called source problem, to improve generalization in another setting, called target problem. The output can be of different nature in the two settings. In the case of CNNs, the network is first trained on a source dataset and task and then the learned weights of the model are transferred to the target network to be trained on a target dataset and task. While dealing with deep convolutional networks that have millions of trainable parameters, this optimization technique works like a shortcut to save time,

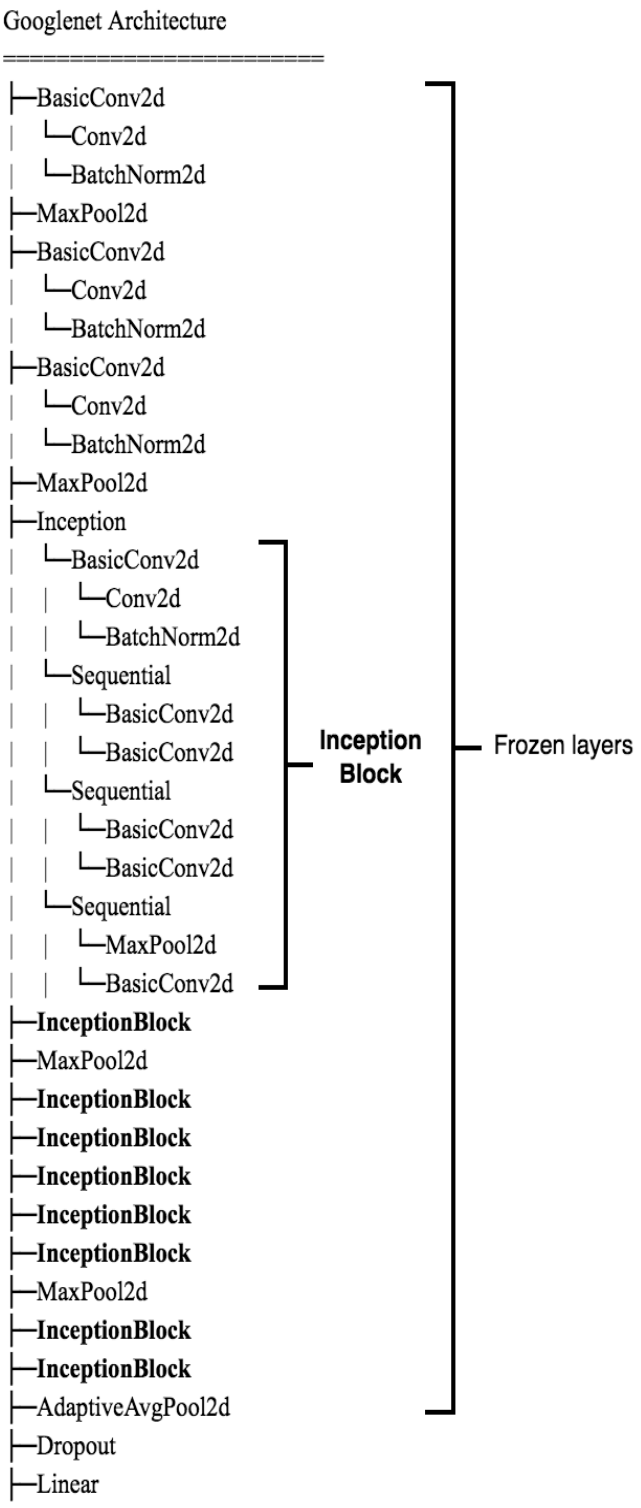


(a) AlexNet



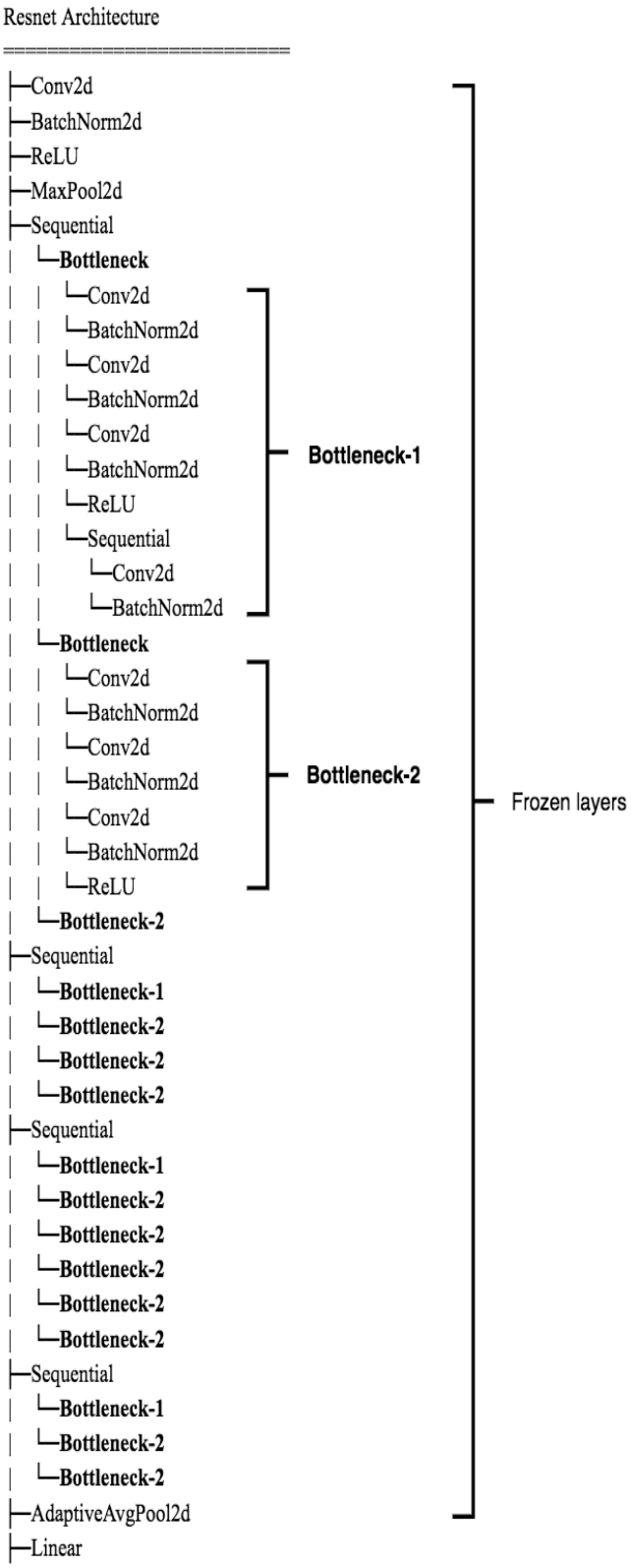
(b) VGG-19

Figure 8.3: Model Architectures



(c) GoogLeNet

Figure 8.3: Model Architectures (Contd.)



(d) ResNet-50

Figure 8.3: Model Architectures (Contd.)

resources and to achieve better performance. Since the technique tries to repurpose the learned weights from the source dataset to fit the target dataset, it can work well on small target datasets. Transfer learning can also speed up the model convergence during training and boost prediction accuracy.

The sub-problem of Transfer Learning within Deep Learning focuses on learning from massive datasets and applies the learned knowledge in other domains. For fault diagnosis of power plant components, Transfer Learning has been applied to datasets of various kinds.

Researchers have worked on vibration signals taken from Case Western Reserve University (CWRU) bearing dataset wherein they have performed transfer learning for bearing fault diagnosis [187, 188]. C. Che et al. have validated through their experiments that deep transfer learning is effective for classifying multi-dimensional graph samples produced from rolling bearing signals [189]. A joint distribution adaptation method based on feature-transfer learning has been verified by Z. Wu. et al.for bearing fault diagnosis on frequency spectra of two types of bearing vibration datasets [190]. Vibration signal images are also highly favoured since image classification with neural networks has seen a lot of advancements.

Researchers have explored transfer learning with vibration signal images generated from CWRU bearing signals [112], [191, 192]. Researchers have also focused on gear fault diagnosis system that uses adaptive feature extractions from images created from a small gear dataset [193] . With only 104 samples per health condition of the gear, they achieved an accuracy of 95.88% with 20% as training data, hence removing the requirement of sufficiently labelled samples. Deep CNNs have revolutionised the state of the art results in many research studies. The four CNNs - AlexNet, VGG-19, GoogLeNet, ResNet-50 - have been applied previously on images in computer vision related research and have been found to be very effective in n-class classification problems. In this research, transfer learning of these four CNNs has been implemented on bispectrum images generated from CWRU bearing fault signals. These networks are first initialized with their corresponding pre-trained weights on the ImageNet dataset and then fine-tuned on bispectrum image dataset. The training details of the models are explained in Section 8.3. Rolling element bearings used in rotating machinery of power plant have 3 common types of faults: Ball bearing, Inner race and

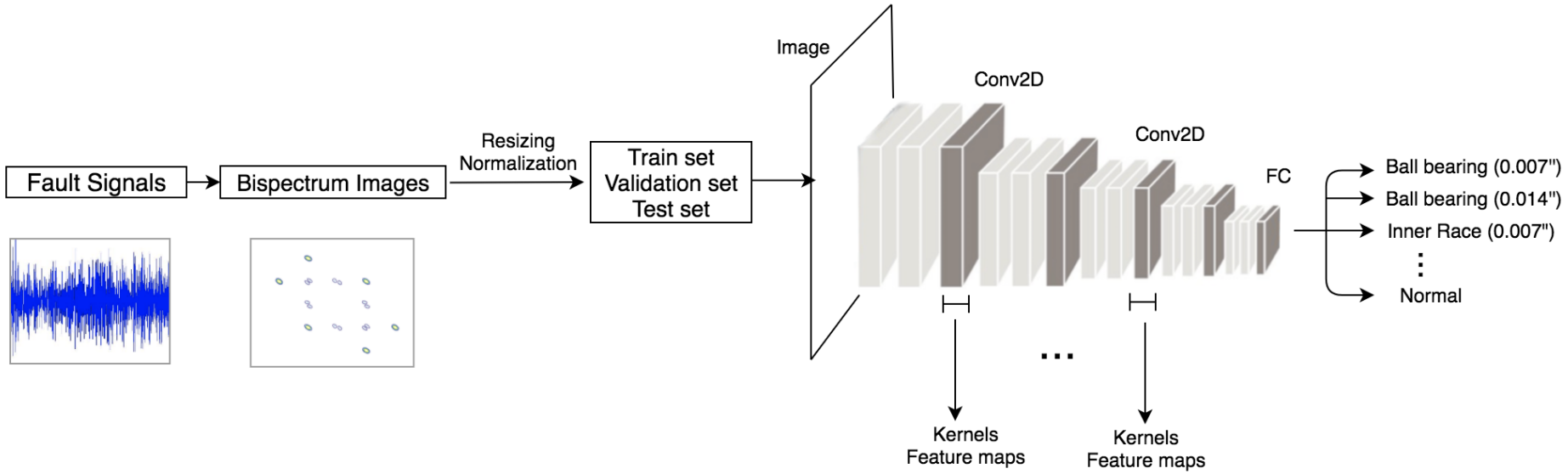


Figure 8.4: Block diagram of Proposed Methodology for Fault Diagnosis and Classification using Bispectrum Images and Deep Transfer Learning.

Outer race. In this research work, bearing faults are diagnosed by taking into account both fault type and fault diameter size. Two fault sizes of 0.007" and 0.014" are considered for each kind of fault, resulting in 6 fault classes. Combining these with the Normal class, a 7-class classification has been performed.

8.3 FAULT DIAGNOSIS USING CWRU BEARING DATA

Bispectrum images were generated from bearing fault signals to create the dataset for the classification models. For the pre-processing phase, resizing and normalization were performed on these images. These pre-processed images were used to form the train, test and validation sets.

The train and validation set images were then passed through the pre-trained CNN models to perform transfer learning based training of these models. Figure 8.4. describes the proposed methodology.

The train, validation and test data are derived from Case Western Reserve University (CWRU) Bearing Dataset [133]. Bispectrum arrays are estimated for each of the data samples using the MATLAB toolbox for direct fast fourier transform (FFT) based approach with origin at the centre of the array and axis pointing down and towards right. FFT length and overlap are taken as 128 and 0 respectively. The bispectrum arrays are stored as RGB images so that deep learning models can learn useful insights from the color intensities. Bispectrum images obtained for Ball Bearing and Normal classes are 875x656 in size and that for Inner Race and Outer Race are 1200x900 in size. Some of these images are shown in Figure 5.3 in chapter 5. These images are divided into train, test and validation sets as depicted in Table 3.5.

Each of the bispectrum images is then preprocessed, as shown in Figure 8.5., before being passed to the networks in order to match the pretrained models' implementation structure. For this, the images are first resized to 224 by 224 size and are then normalised to mean = 0 and standard deviation = 1.

In order to underscore the effectiveness of the proposed approach, 4 popular deep transfer networks namely AlexNet, VGG19, GoogLeNet and ResNet-50 were leveraged. All the models were implemented in Python using the Pytorch framework. They were first initialised with their corresponding pretrained weights on ImageNet dataset and then fine-tuned on the bispectrum dataset for a mini-batch of 128 images

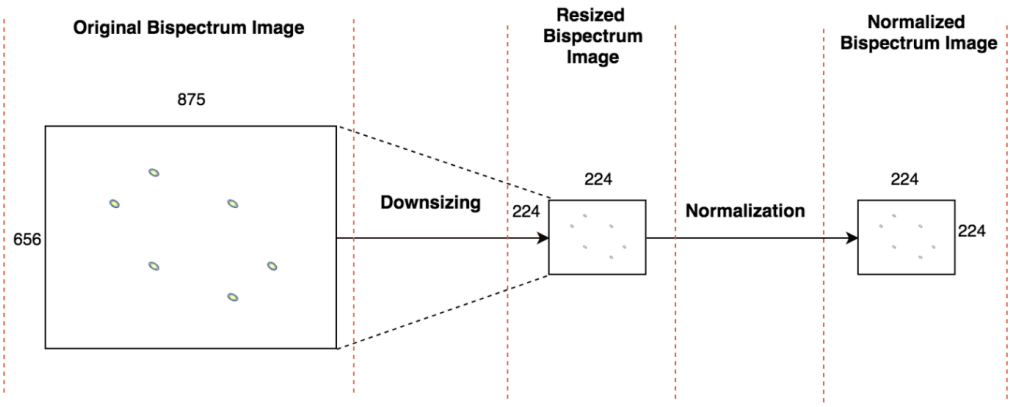


Figure 8.5: Pre-processing of images before feeding them to the Deep CNN model. Here a sample image of size 875 by 656 is first resized to 224 by 224 size and then normalized.

for 30 epochs. This means that for every model, the entire dataset was passed through the network 30 times wherein every pass was further divided into smaller batches of size 128. The last layer of each network was modified to a probability vector of size 7, in order to customize the models to the 7-class bearing dataset. Each model was trained using 3 different optimizers - Stochastic Gradient Descent (SGD), Adaptive Moment Estimation (Adam) and Adamax - making a total of 12 models that were experimented on the dataset. The selected three optimizers are popular gradient descent optimization algorithms that are known to achieve good convergence. In all the experiments, the learning rates for SGD, Adam and Adamax were set to 1e-2, 1e-3 and 2e-3 respectively and the momentum was set to 0.5 for SGD. These values were decided upon after repetitive experiments.

A popular learning rate annealing method of step decay was also used for gradually decreasing the learning rate (lr) over the course of training so as to avoid a noisy convergence. The lr rate scheduler for the same was set to a decay factor of 0.1 every 7 epochs. Gradient clipping was also set to enforce a maximum gradient value of 0.25 to stop the gradient from going above this threshold thereby avoiding gradient explosion. The loss criterion chosen for all models was Cross Entropy Loss as this criterion is commonly used for n-class classification and also it combines LogSoftmax non linearity and NLL (Negative Log Likelihood) loss in one single class in Pytorch. The models have been trained using a CUDA enabled NVIDIA P40 GPU with 20 GB memory.

In transfer learning, there are different ways of transferring knowledge from pre-

trained models. In this research work, two transfer learning methods were analyzed:

i) Transfer Learning Method -1 (TLM-1): Fine-tuning without Freezing
ii) Transfer Learning Method -2 (TLM-2): Fine-tuning with Freezing

The first method involves taking the pre-trained model and fine-tuning it from head-to-toe for the bearing dataset. The second method also involves fine-tuning but it fine-tunes only selected block of layers, which are towards the end of the model, keeping the rest of the layers frozen.

8.3.1 Transfer Learning Method -1 : Fine-tuning Without Freezing

In this experiment, each model was first initialized with the corresponding pretrained weights for ImageNet and then the entire model was retrained on bispectrum images of bearing fault signals. During every epoch of training, the network was first trained on the train set and then validated against the validation set. Table 8.1 reports the classification results in terms of the metrics - validation accuracy, test accuracy, test loss and AUPRC (Area under Precision Recall Curve) values. Accuracy can be defined as the number of correctly predicted samples over the total number of samples in the dataset. Accuracy obtained for validation set is validation accuracy and likewise for test accuracy. The model state with the highest validation accuracy across all epochs was used to make predictions on the test set. This best validation accuracy value has been reported in Table 8.1. Average loss is the mean loss over losses from all the mini batches in one pass through network. All the loss values reported are average loss values. Accuracy values may not be a good metric in cases of class imbalanced datasets. Moreover, another commonly used metric, called area under the ROC (AUROC), can produce optimistic values on imbalanced datasets and is a preferred metric for balanced class data. Since the dataset is imbalanced, the area under the Precision-Recall curve (AUPRC) has been instead computed, which is preferred over AUROC when dealing with imbalanced datasets, for evaluating the performance of the models. Higher the value of AUPRC, better is the classification performance of the model.

GoogLeNet and ResNet-50 with Adam and Adamax optimizers perform the best with very high validation accuracies of 99.85%, 99.70%, 99.55%, 99.85% in this order and high test accuracies of 99.55%, 99.55%, 99.70%, 99.55% in this order and very

Table 8.1: Validation and Test results for the CNN models without freezing layers.

Model	**Optimizer**	**Validation accuracy (%)**	**Test loss**	**Test accuracy (%)**	**AUPRC (Micro-average)**
AlexNet	SGD	95.42	0.2468	94.96	0.9856
	Adam	99.85	0.0277	98.96	0.9994
	Adamax	98.67	0.0617	97.63	0.9982
VGG-19	SGD	94.10	0.2594	93.77	0.9810
	Adam	99.70	0.0831	99.26	0.9963
	Adamax	99.70	0.0132	99.11	0.9998
GoogLeNet	SGD	92.62	0.7691	90.65	0.9639
	Adam	99.85	0.0076	99.55	0.9999
	Adamax	99.70	0.0096	99.55	0.9999
ResNet-50	SGD	97.19	0.3601	94.51	0.9914
	Adam	99.55	0.0051	99.70	0.9999
	Adamax	99.85	0.0106	99.55	0.9999

low loss values of 0.0076, 0.0096, 0.0051, 0.0106. Following very closely behind this model is VGG-19 with Adam and Adamax optimizers. From the corresponding scores of AlexNet, it can be deduced that it is also performing very well.

The very high test accuracies of these models further confirms their excellent classification performance and absence of overfitting on train set. This excellent performance can be attributed to the fact that even though all the layers of the models were re-trained on the bearing dataset, the models were already initialized with the weights of the pre-trained models thereby already storing most of the important information for curves and edges needed by the models to distinguish basic shapes, hence making it easier and quicker for these models to customize their weights for the bearing dataset.

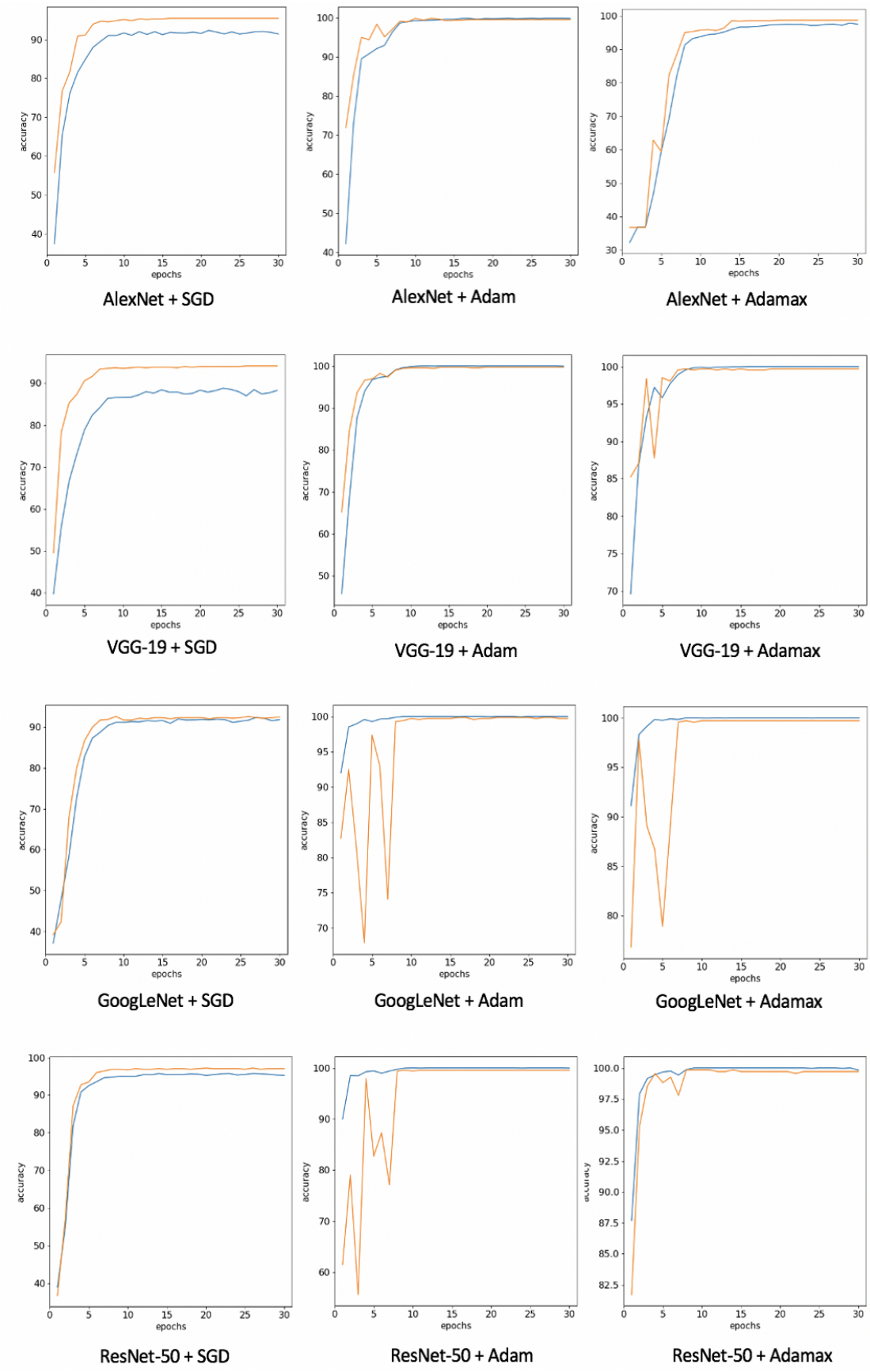
The area under the Precision-Recall curve (AUPRC) is a useful tool for evaluating soft classifiers like CNNs in terms of their quality of class separation. Since this is a multiclass classification setting, averaged AUPRC values have been computed by the micro-average method. Table 8.1 shows the micro-averaged values of AUPRC for the models. This metric has been computed by aggregating the contribution of all the classes by the one-vs-all technique and does not treat all classes equally, which again helps to avoid getting misleading results for an imbalanced dataset. Since high averaged values are obtained for AUPRC for all the models, it can be concluded that the trained models are able to separate the 7 classes well. These also show that the models are performing well on an imbalanced dataset.

Figure 8.6. contains the accuracy vs epochs and average loss vs epochs plots for training and validation. From the graphs, it is observed that the accuracy and loss curves converge very fast, each of them converging by the 15th epoch. This shows that models are able to transfer the knowledge they learnt from the ImageNet dataset to the bispectrum images of CWRU vibration signals and further improve their learning in just 15 passes of the entire dataset.

Another factor contributing to the good performance of the deep transfer networks, is the noiseless smooth convergence of both accuracy and loss observed from these graphs, highlighting their low dependence on big datasets and preventing researchers from worrying about sufficiency of data samples.

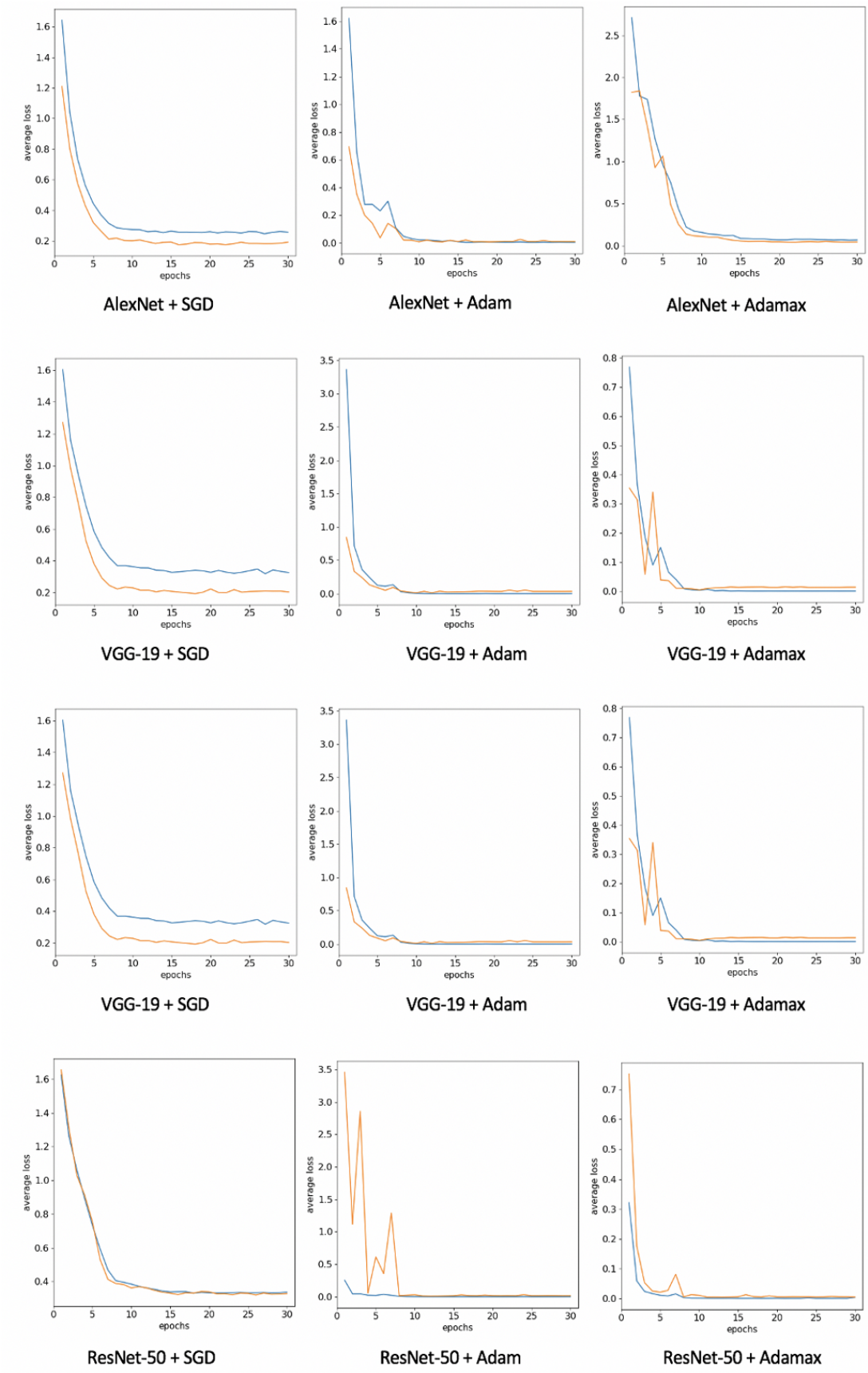
The AUPRC values can also be interpreted by plotting the precision recall curve. The nearer the area of the precision-recall curve to one, better is the predictive power of the model. The precision-recall curve for one of the models, i.e., AlexNet trained with SGD optimizer has been shown in Figure 8.7. The plot shows the AUPRC score micro-averaged over all the seven classes. It can be observed that the curve is almost aligning with the edges of the 1 by 1 plot area. This represents the high micro-averaged AUPRC score of 0.9856 and shows the model's strong ability to perform classification on imbalanced datasets.

The total time taken for training the models for 30 epochs is reported in Table 8.4. It is observed that models take less time to train when their layers are frozen. The time values are decent considering the best accuracy and loss were achieved in half the time i.e. by the 15th epoch and the weights were not being learning from scratch.



(a) Accuracy vs epochs plots

Figure 8.6: Train and Validation plots for TLM-1. Blue curve represents Train results and orange curve represents Validation results.



(b) Average loss vs epochs plots

Figure 8.6: Train and Validation plots for TLM-1. Blue curve represents Train results and orange curve represents Validation results (Contd.)

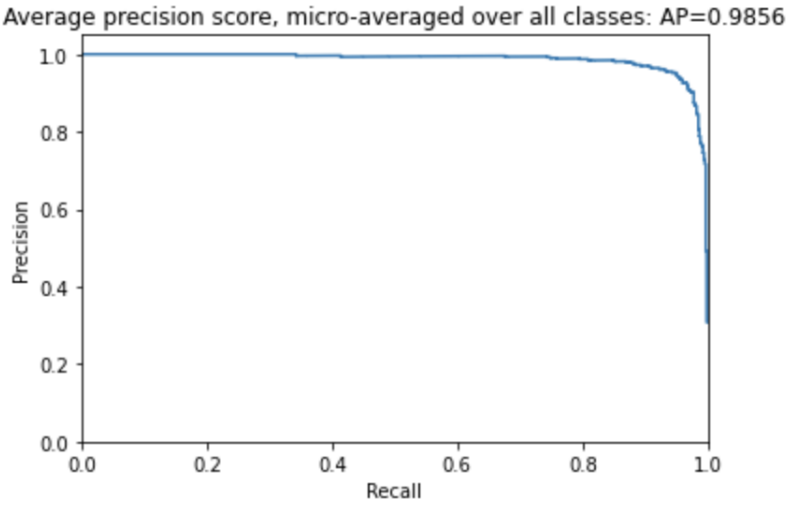


Figure 8.7: Precision-Recall curve for AlexNet model trained with SGD optimizer in TLM-1.

8.3.2 Transfer Learning Method -2 : Fine-tuning With Freezing

In this experiment, the models were again initialized with the corresponding pretrained weights for ImageNet but the fine-tuning process involved updation of weights of few selected layers. The models were fine-tuned by freezing the model layers closer to the input layer and training the layers closer to the classification output layer.

During the training of a convnet, the layers which come first in contact with the input perform the task of extracting basic edges and curves whereas the subsequent layers are responsible for detecting more complex shapes and figures and other visual information that the model finds to be popular in a class, thereby associating that information with that class.

Since any image can be disintegrated to a set of basic edges and curves, the initial layers are usually kept frozen so that their pretrained weights can be directly utilized and the computational resources can be used for training the remaining layers which are more important in terms of customising the model training to the dataset in hand. This approach accelerates training and also prevents overfitting of models on the dataset. For all the 4 networks, all the layers till the AdaptiveAvgPool2d layer were frozen and the rest of the layers were trained as depicted in Figure 8.3.

Table 8.2 tabulates the results when transfer learning was performed with the CNN models after freezing the initial set of layers. Unlike experiments in TLM-1; AlexNet

with Adamax clearly outperforms other models achieving a very high validation accuracy 99.11%, low test loss of 0.0659 and high test accuracy 98.96%. Its results with Adam optimizer are also good achieving the scores of 98.96%, 0.1584 and 98.07% in the same order. VGG-19 exhibits similar performance with these two optimizers. GoogLeNet and ResNet-50 are not far behind, both producing decent accuracies and losses with Adam and Adamax optimizers.

For GoogLeNet and ResNet-50, a slight dip in performance before freezing and after freezing layers is observed. From 99.85% validation accuracy and 0.0076 test loss value obtained before freezing to 95.57% and 0.1979 after freezing, GoogLeNet with Adam sees this dip in performance because now only some of its weights are trained to fit the bearing dataset.

Table 8.2: Validation and Test results for the CNN models after freezing layers.

Model	**Optimizer**	**Validation accuracy (%)**	**Test loss**	**Test accuracy (%)**	**AUPRC (Micro-average)**
AlexNet	SGD	88.64	0.5795	87.39	0.9426
	Adam	98.96	0.1584	98.07	0.9950
	Adamax	99.11	0.0659	98.96	0.9992
VGG-19	SGD	86.57	0.7166	85.46	0.9316
	Adam	98.82	0.1561	98.22	0.9967
	Adamax	98.96	0.1148	97.92	0.9962
GoogLeNet	SGD	85.10	0.9940	86.05	0.9234
	Adam	95.57	0.1979	95.25	0.9853
	Adamax	95.42	0.2163	94.66	0.9838
ResNet-50	SGD	92.47	0.7465	90.50	0.9635
	Adam	97.93	0.1289	96.44	0.9944
	Adamax	97.93	0.1313	95.99	0.9946

Table 8.3: Number of trainable parameters in the models

Model	**Total parameters trained without freezing layers**	**Total parameters trained with freezing layers**
AlexNet	57,032,519	54,562,823
VGG-19	139,598,919	119,574,535
GoogLeNet	5,607,079	7,175
ResNet-50	23,522,375	14,343

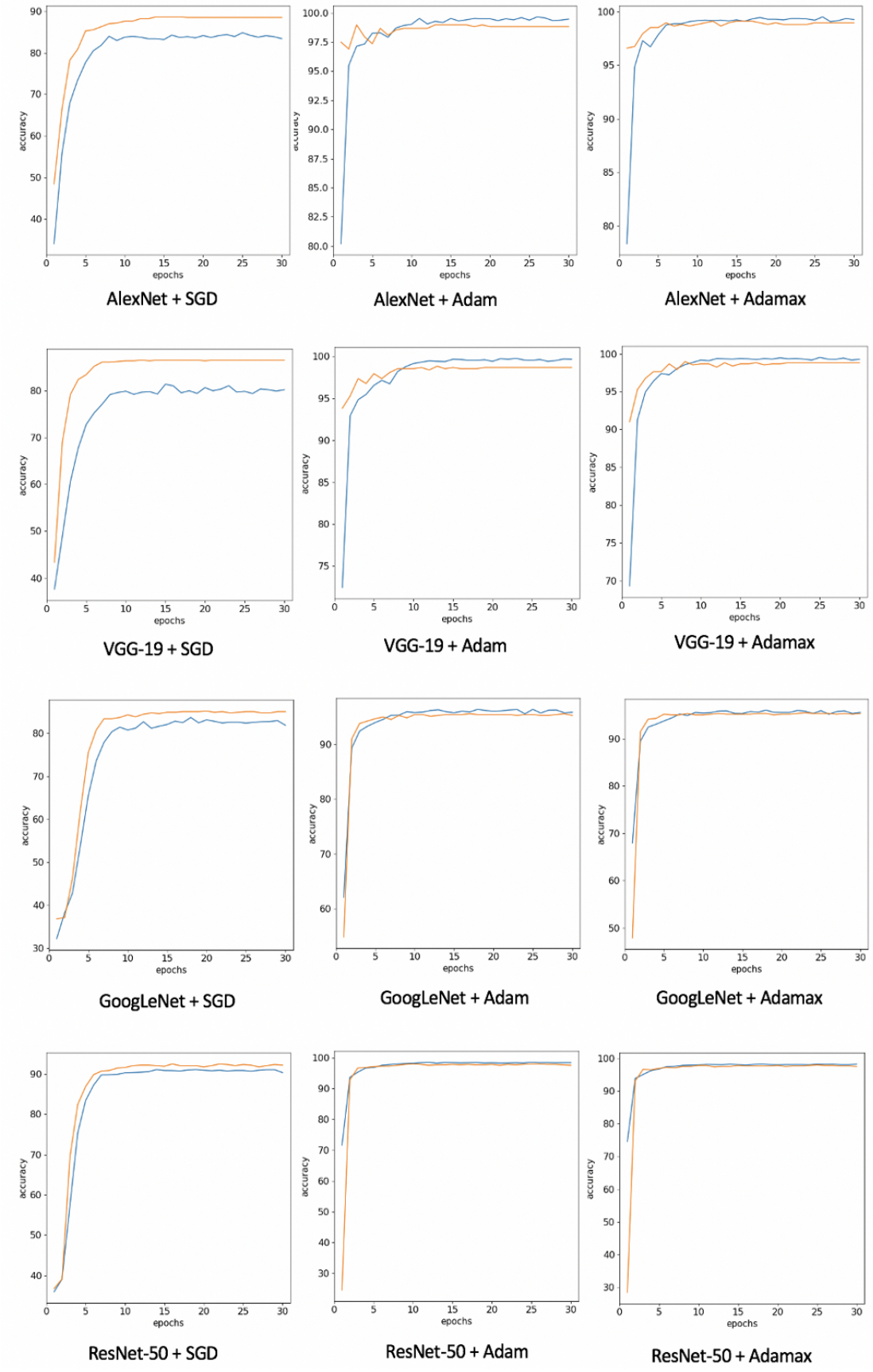
One can also see this from the number of parameters trained by these models as reported in Table 8.3. For GoogLeNet, the trainable parameters reduce from the order of 1e6 before freezing to 1e3 after freezing. For the same reason as discussed above, in the case of ResNet50 with Adamax, validation accuracy and test loss change from 99.85% to 97.93% and 0.0106 to 0.1313. However, in AlexNet and VGG-19 significant change in result scores is not seen. This is due to a larger proportion of weights being re-trained in the latter case as can be seen from Table 8.3.

In the former case consisting of ResNet-50 and GoogLeNet, the layers which were open to training were relatively less than the total number of layers in these networks, thus accounting to a visible effect on the training scores. Thus the performance of transfer learning depends on the number of layers or blocks kept unfrozen and this can be further experimented and explored in future.

Table 8.2 also shows the micro-averaged values of AUPRC for the models. It can be observed that the AUPRC scores for the models are again very high, hence showing their superior performance on the imbalanced dataset and high ability of class separation.

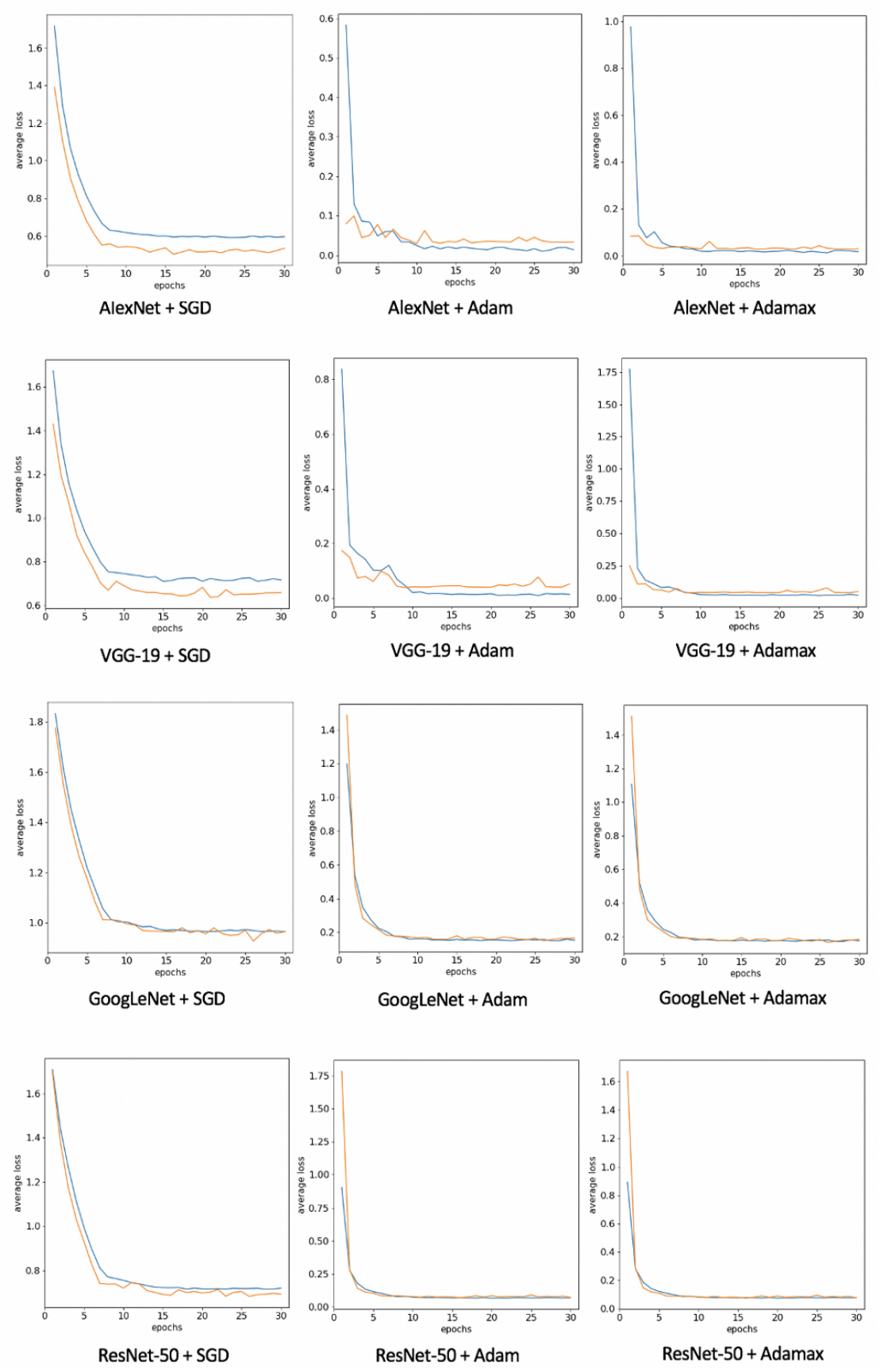
The accuracy vs epochs and loss vs epochs plots, in Figure 8.8, for training and validation with freezing layers show similar trends as in Figure 8.6, but it is observed that convergence is smoother and more stable this time with lesser fluctuations.

From the time scores in Table 8.4, it is seen that models take less time to train when their layers are frozen, for example, GoogLeNet with Adam takes 13m 50s to train for 30 epochs with freezing the layers, a total of 6m 27s less than the 20m 17s it took to train for the same number of epochs when layers were not frozen. In AlexNet, this delta is less for the same reason discussed above for loss and accuracy scores.



(a) Accuracy vs epochs plots

Figure 8.8: Train and Validation plots for TLM-2. Blue curve represents Train results and orange curve represents Validation results.



(b) Average loss vs epochs plots

Figure 8.8: Train and Validation plots for TLM-2. Blue curve represents Train results and orange curve represents Validation results. (Contd.)

Table 8.4: Time taken for training the models over 30 epochs.

Model	Optimizer	Training time without freezing layers	Training time with freezing layers
AlexNet	SGD	14 min 51 sec	13 min 51 sec
	Adam	14 min 57 sec	13 min 50 sec
	Adamax	14 min 57 sec	13 min 45 sec
VGG-19	SGD	17 min 43 sec	13 min 40 sec
	Adam	18 min 19 sec	15 min 2 sec
	Adamax	18 min 22 sec	13 min 44 sec
GoogLeNet	SGD	18 min 8 sec	11 min 11 sec
	Adam	20 min 17 sec	13 min 50 sec
	Adamax	20 min 28 sec	13 min 35 sec
ResNet-50	SGD	13 min 17 sec	11 min 19 sec
	Adam	14 min 45 sec	13 min 20 sec
	Adamax	15 min 29 sec	14 min 5 sec

8.3.3 Visualization of Feature Maps

Convolution neural networks learn to classify input images by mapping the input image's pixel values to a set of weights stored in the form of grids that are called kernels or filters. These filters, when applied on an image, convert the image into an array, called a feature map, that contains all the important discerning features.

Researchers have used visualization techniques to give insight into the function of intermediate feature layers of CNNs [194, 195]. These techniques offer a useful way of interpreting the contribution from different model layers towards the overall performance of the model and can also be useful for comprehending the learning process of a CNN model.

In this section, the feature maps generated by successive convolution layers when a model is trained on the bearing dataset, are visualized and attempted to be analyzed. Feature maps can be useful in understanding the discerning features in bispectrum images that the model thinks are contributing the most in bearing fault and severity classification, thus providing an insight into the performance of the trained deep CNN models. They also help in understanding the role of intermediate convolution layers in classification.

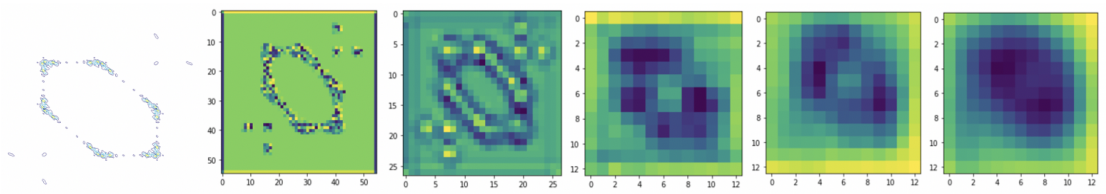
For this research work, the feature maps generated by the 5 convolution layers present in AlexNet architecture (in Figure 8.3) are plotted for TLM-1. This was performed by forwarding test images of each class through the AlexNet network that got trained with Adamax optimizer. When an input test image passes through a convolution network like AlexNet, every convolution layer outputs a three dimensional tensor which can be viewed as a stack of feature maps. From such a tensor of dimension, let's say (k,m,n), a (1,m,n) feature map is extracted and is plotted as a two dimensional image. So for the 224x224 input RGB image, the feature maps generated by the first 5 layers are of sizes 55x55, 27x27, 13x13, 13x13, 13x13 in the order. The 1st feature map for each of these convolution layers were generated as images. These feature map images were extracted for a bispectrum image per class.

Figure 8.9. shows seven bispectrum input images (one per class) and their corresponding feature map visualizations generated from the first 5 convolution layers while classifying these images. As discussed earlier and also evident from the figure, the 1st convolutional layer has all the basic edge and color detectors. As one goes deeper into the network, the abstraction level of the feature maps increases. The subsequent convolution layers appear to be detecting more complex shapes, textures and mesh patterns, all the time focusing more on the outline of the central ring and its surrounding objects present in these images.

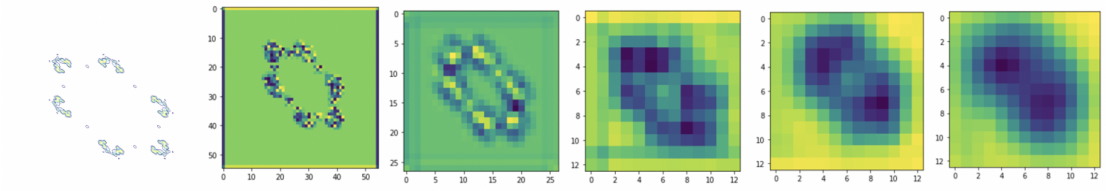
From colors and basic curves to complex shapes and mesh patterns, it can be observed that these visualisations have a hierarchical nature. But still it is difficult for the human eye to discern and understand the complex feature maps learned by layers as one goes deeper in the architecture. A possible future study to further understand these visualizations can involve visualizing filter maps of these convolution layers.

8.3.4 Results and Discussion

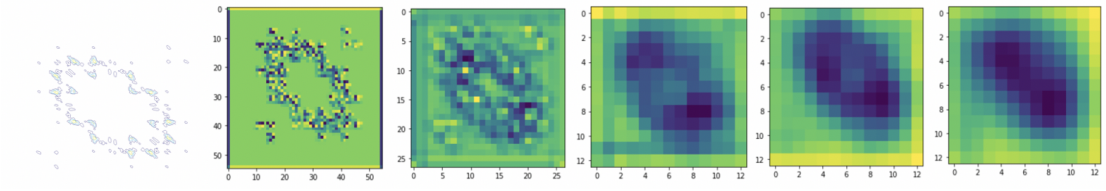
i) The results from experiments TLM-1 and TLM-2 are reasonably good. From the results in Table 8.1 for TLM-1, GoogLeNet and ResNet-50 with Adam and Adamax optimizers perform the best with very high test accuracies of 99.55%, 99.55%, 99.70%, 99.55% in this order and very low loss values of 0.0076, 0.0096, 0.0051, 0.0106. In TLM-2 on the other hand, AlexNet with Adamax clearly outperforms other models achieving a very high test accuracy of 98.96%



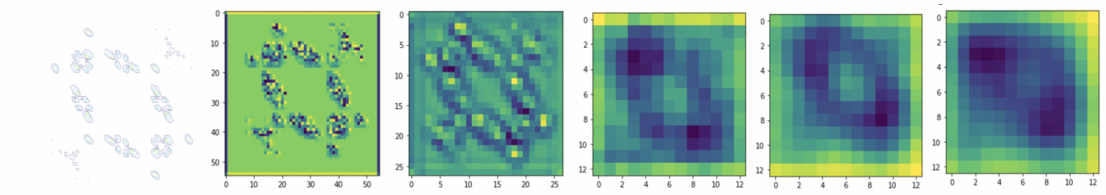
(a) Ball bearing with 0.007” fault



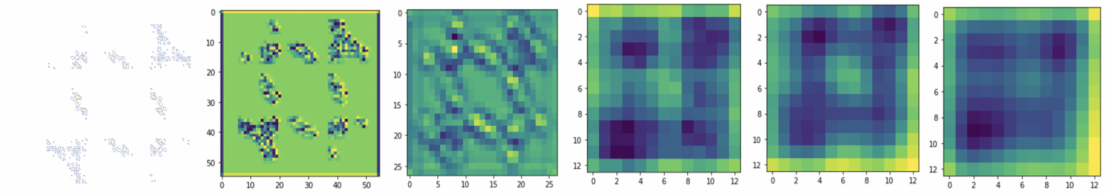
(b) Ball bearing with 0.014” fault



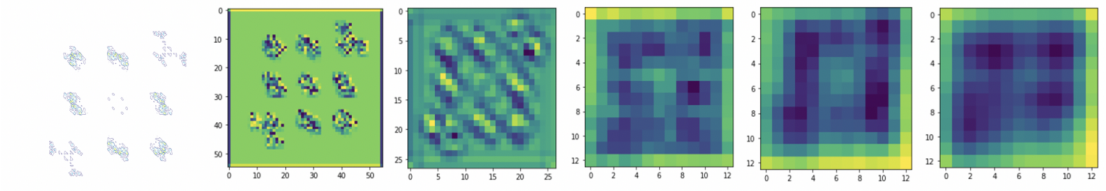
(c) Inner Race with 0.007” fault



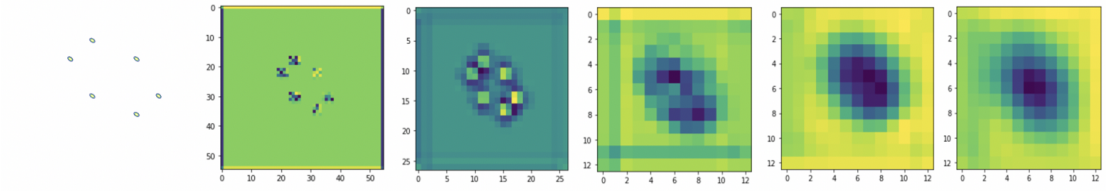
(d) Inner Race with 0.014” fault



(e) Outer Race with 0.007” fault



(f) Outer Race with 0.014” fault



(g) Normal class

Figure 8.9: Feature maps produced by the trained model - AlexNet + Adamax - in TLM-1 for Test images. Each row has images arranged from left to right as: Test input image, feature maps from 1st, 2nd, 3rd, 4th and 5th convolution layers.

and low test loss of 0.0659, as can be seen in Table 8.2. From these obtained test accuracy values, it is observed that the concern of overfitting for the experiments excluding freezing of layers, doesn't hold true and for a not so large dataset like ours, transfer learning can perform well and is robust even when retraining all the layers. Thus, transfer learning based CNN, trained with or without freezing, forms an excellent and robust classifier for small and class imbalanced bearing dataset.

ii) From observing the feature detection being performed by the network layer by layer, a small insight into how these transfer learning networks are discerning the important features and narrowing down the class associated with that image, is gained. It also highlights the efficiency and power of transfer learning in transferring knowledge gained from a dataset to classifying a different dataset.

8.4 FAULT DIAGNOSIS USING REAL POWER PLANT DATA

In this experiment, bearing data from 500 MW Kosti Thermal Power Plant commissioned by Bharat Heavy Electricals Limited, India [Figure 3.6.] has been used to test the proposed methodology. Because the bearing configuration of the pump of Kosti Thermal Power Plant matched with that used in CWRU dataset, the test samples acquired from the power-plant are appropriate to validate the trained model. So far, the diagnostic performance of higher order spectrum of vibration signals with deep transfer learning has not been explored on bearing data acquired from a Power Plant. This section attempts to do that.

8.4.1 Experimental Description

Data for normal baseline bearings was recorded using the deep groove ball bearing SKF6303-2Z/C3 with dimensions 17mm X 47mm X 14mm at drive-end with motor speed of 3000 RPM. Vibration data was collected at 12,000 samples/second. Vibration data was taken for 2 different timestamps and for both vertical and horizontal positions, leading to a total of 4 vibration signals.

These signals were bifurcated into sub signals of length 1024 and the same pre-processing steps were performed on these sub signals as done for CWRU dataset. Bispectrum arrays were constructed and the corresponding images were stored as

RGB images. These images were resized and normalized and were mixed with the test set images separated from the CWRU Dataset C. This new set of test images was then passed through the trained network model of AlexNet with Adamax, obtained by performing TLM -1 experiment on CWRU dataset. For one of the Power plant samples, feature maps were generated as done while experimenting on CWRU dataset.

Figure 8.10 shows the feature map visualizations generated by the first 5 convolution layers while classifying the chosen power plant sample.

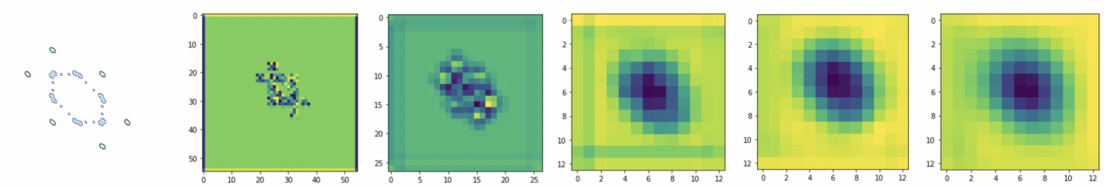


Figure 8.10: Feature maps produced by the trained model - AlexNet + Adamax - in TLM-1 for a sample Bispectrum Test image constructed from signal acquired from the Power Plant. Images are arranged from left to right as: Test input image, feature maps from 1st, 2nd, 3rd, 4th and 5th convolution layers.

8.4.2 Results and Discussion

The test accuracy, test loss and AUPRC scores achieved were 98.3%, 0.0561 and 0.9986 respectively. Furthermore, per instance classification results were obtained and it was observed that all the normal samples taken from the power plant were correctly classified. Hence, the performance of this methodology is further confirmed. Also, similar to the discussion on the visualization of feature maps generated for CWRU dataset samples, the feature maps generated for the power plant sample provide an insight into how the prominent features are being narrowed down as the image goes through the network.

8.5 CONCLUSION

Bearing fault diagnosis based on hand crafted features demands prior knowledge and might be unsuitable for detecting higher level abstract features that could be out of scope for the human mind. With the introduction of convolutional neural networks, the dependence on hand crafted features is removed but a requirement for a reasonable sized dataset is still there to ensure unbiased and robust training of the network. This section proposed a novel method based on deep neural network transfer learning and

higher spectral analysis such as Bispectrum with the goal of performing bearing fault and severity detection on small datasets, eliminating the need for hand tailored features or large datasets. Experiments involved popular network architectures - AlexNet, VGG-19, GoogLeNet and ResNet-50 - wherein these models were loaded with pre-trained weights and further trained on bispectrum images obtained from CWRU bearing fault dataset for the two cases of with and without freezing layers. Each model was trained using 3 optimizers - SGD, Adam and Adamax.

The use of CNN for classifying bispectrum images has already been investigated. [128] proposed a fault diagnosis method that is based on bispectrum analysis and CNN for detection of bearing faults in CWRU dataset. They used the same three optimizers and trained their CNN for 500 epochs on samples of one RPM and test them on samples of remaining 3 RPM, classifying the samples into Normal, Inner Race, Outer Race and Ball Bearing, obtaining average test accuracies of 91.97%, 96.27%, 96.74% and 90.10% for 1796, 1772, 1748, 1722 RPM values respectively. Though in this research work, experiments are performed wherein the train dataset has samples from all these RPM values and is of similar size, the proposed approach achieves an accuracy 99.7% in only 30 epochs.

Thus transfer learning on bispectrum images has a very good potential for diagnosing bearing faults and detecting fault severity . It can be used to speed up network training and save computational resources by reusing the weights of pre-trained models. Though the knowledge on the inner workings of these networks is limited and complex to understand, the feature maps of convolutional layers can be visualized to gain an insight into the same.

The efficacy of the proposed methodology is further confirmed by its good performance in classifying samples taken from an operating power plant.

8.6 CHAPTER SUMMARY

In this chapter, a novel methodology for fault diagnosis and classification using bispectrum images and deep transfer learning has been proposed and discussed. It is proved that deep convolutional neural networks, when trained on Bispectrum images of fault signals using transfer learning, provide highly accurate and reliable results for fault diagnosis that are at par with the state of the art results. These transfer learning based

models are able to quickly learn patterns from visual features of a vibration signal's bispectrum.

Four pretrained networks - Alexnet, VGG-19, GoogLeNet, ResNet-50 - have been fine tuned on bispectrum images prepared from vibration signals of machine ball bearing elements. Each network has been trained with 3 optimizers - Stochastic Gradient Descent (SGD), Adaptive Moment Estimation (Adam) and Adamax. These models are able to obtain high classification accuracy within a few epochs. Furthermore, the feature maps associated with intermediate convolution layers for one of these models have been generated and analysed.

CHAPTER IX
CONCLUSIONS, CONTRIBUTIONS AND SCOPE FOR FUTURE WORK

9.1 CONCLUSIONS

This research focused on fault diagnosis of power plant components using a range of vibration signature analysis techniques. It proposed novel algorithms comprising effective signal processing techniques in conjunction with advanced machine learning and deep learning techniques. The proposed algorithms were experimentally validated using bearing vibration data acquired from vibration lab and a real power plant. The effectiveness of the techniques was assessed by the analysis of experimental results concerning different types of faults.

This research was conducted in two phases to meet the research objectives. In the first phase, this research work extracted a comprehensive set of conventional statistical features, proposed novel statistical features and ranked them to find an optimal feature subset thereby decreasing the number of computations involved in fault detection using efficient machine learning techniques. In the second phase, a higher order spectrum, Bispectrum, was investigated for its sensitivity and effectiveness in fault diagnosis by applying deep transfer learning technique. Multiple conclusions were drawn from the outcomes of the two phases of research. These are discussed below:

9.1.1 Vibration Signature Analysis using Statistical Features and Machine Learning

In the first phase, extensive research was done to find advanced statistical parameters having more sensitivity to faults; and on the selection of an ideal set of dominant features from a comprehensive list of features in order to give them as input feature vectors to efficient machine learning classifiers for fault diagnosis. Statistical time domain features namely Hjorth Parameters and Normal negative log likelihood for Gaussian Mixture Model (GMM) were proposed and analyzed for the first time. Effectiveness of Hjorth parameters in the prediction of bearing faults was investigated

by scatter plots for all possible pairs of the three Hjorth parameters. It was observed that the pair of Hjorth Activity and Hjorth Mobility is capable of efficiently distributing the bearing vibration data into clusters and have great potential to be used as effective features for machine learning classifiers in performing accurate fault diagnosis. One more advantage of using Hjorth parameters is that their calculation involves variance, making cost of their computation very low. This accounts for their suitability in real-time tasks.

The diagnostic performance of Normal Negative Log Likelihood for GMM was compared with that of the established feature Normal negative log likelihood for Single Gaussian and it was proved that Normal negative log likelihood for GMM has better fault sensitivity as compared to its Single Gaussian counterpart.

These novel proposed features in addition to 27 other established statistical features for identification of bearing fault type and severity were investigated using a two-step approach to generate an optimal feature subset:

- Filter-based Ranking (using 3 ranking metrics)
- Filter-based Feature Subset Selection (using 11 search techniques)

The results obtained show that the newly proposed features (Hjorth parameters and Normal negative log likelihood for GMM) occupy a prominent place in the optimal subset and have a future potential of being used in a high dimensional dataset with multi-domain features.

The selected set of statistical features were further validated using Ensemble machine learning classifiers. It was observed that PART when used with Multiboost outperforms all the classifiers and gives the best performance with an accuracy of 96.75% on test dataset and 100% on the train dataset for fault type detection and 98.56% on test dataset and 100% on the train dataset for fault severity detection using optimal feature subset.

The performance of the features was further studied by inspecting the rules generated by them.

The diagnostic effectiveness of the features was further validated on a bearing dataset obtained from an operating thermal power plant. The overall accuracy and average F-measure achieved were 96.75% and 0.972 respectively.

9.1.2 Vibration Signature Analysis using Bispectrum Images and Deep Transfer Learning

In the second phase of the research, a higher order spectrum (Bispectrum) was investigated using Deep Transfer Learning for its sensitivity and effectiveness in fault diagnosis on insufficient labelled data of bearing fault images. Different deep transfer learning networks were used to analyze the performance of the proposed novel method.

The performance of the proposed algorithm was evaluated using metrics of validation accuracy, test accuracy, test loss and micro-averaged AUPRC scores. The proposed method includes two aspects, first is the construction of Bispectrum images from vibration signals followed by performing deep transfer learning on these images for highlighting the fault features and diagnosing them. Bispectrum images of bearing vibration signals were fed as input to four popular convolution neural networks (AlexNet, VGG-19, GoogLeNet and ResNet-50). The training of the convolution neural nets was performed using transfer learning to eliminate the requirement of large dataset that is usually needed for training deep learning models. In this research work, two Transfer Learning methods were implemented and analyzed:

- Transfer Learning Method -1 (TLM-1): Fine-tuning without Freezing
- Transfer Learning Method -2 (TLM-2): Fine-tuning with Freezing

The results proved that models take less time to train when their layers are frozen. The proposed approach achieved an accuracy of 99.7% in only 30 epochs. The diagnostic effectiveness of Deep Transfer Learning approach was further validated on a bearing dataset obtained from an operating thermal power plant. The test accuracy, test loss and AUPRC scores achieved were 98.3%, 0.0561 and 0.9986 respectively. The results obtained validate the feasibility and effectiveness of applying deep convolution neural networks trained on Bispectrum images of vibration signals using transfer learning for bearing fault diagnostics.

Hence this approach has a very good potential for diagnosing bearing faults and detecting fault severity. This approach also eliminated the need for manual feature extraction. The proposed Machine Learning and Deep Learning frameworks for Vibration Signature Analysis are compared in Table 9.1.

Table 9.1: Comparison of Proposed Machine Learning and Deep Learning frameworks

	Methodology I **Vibration Signature Analysis using Statistical Features and Machine Learning**	**Methodology II** **Vibration Signature Analysis using Bispectrum Images and Deep Transfer Learning**
Data dependencies	Machine learning algorithms proposed in this methodology perform well with handcrafted advanced features and can be applied on small and large vibration datasets for fault diagnosis.	Deep learning algorithms need large amounts of data to understand the problem domain. To apply deep learning on bispectrum image dataset, which is small in size, a deep transfer learning approach has been proposed in this methodology.
Hardware dependencies	This methodology can be implemented on low-end machines, like CPU, as the ML algorithms used are not very computationally expensive. The models used in this methodology have been executed on a 8 GB RAM machine with Intel Core i5-7200U CPU having 2.50 GHz speed.	This methodology involves training of deep convolutional neural networks. Since deep convolutional neural networks involve a large amount of matrix multiplication operations, high-end machines like GPU, are required for efficiently implementing this methodology. The CNN models used in this methodology have been trained using a CUDA enabled NVIDIA P40 GPU with 20 GB memory.
Execution time	Machine Learning algorithms used in this methodology are not computationally expensive and took very less time to train, ranging from few seconds to few minutes.	Deep learning models take a long time to train. If training from scratch, these networks can sometimes take upto several days to train. In this methodology the training time was much feasible, in the order of a few minutes, because transfer learning was used to train the models.

Feature engineering and dimensionality reduction	Feature engineering, or the process of extracting meaningful features from the dataset, requires time and expertise. This process is also crucial to reduce the complexity of the data and make patterns more visible to learning algorithms. For the applied Machine learning models used in this methodology, advanced fault sensitive hand-crafted features have been proposed and extracted, furthermore a filter based strategy for optimal feature subset selection is proposed.	CNNs, while getting trained, learn features from the input dataset on their own. Input images are convolved with kernels to form feature maps. These feature maps contain low-level features in the initial layers and high-level representations in the subsequent layers. Deep learning algorithms used in this methodology try to learn high-level features from input Bispectrum, therefore, eliminating the task of feature selection for dimensionality reduction.
In-class performance	A best test accuracy of 96.75% for fault type detection (4-class classification) and 98.56% for fault severity detection (12-class classification) was achieved by this methodology.	The proposed approach achieved a best test accuracy of 99.7% in 30 epochs for fault severity detection (7 class classification).
Interpretability of algorithms	Machine learning algorithms are easy to interpret. This methodology used Rule based Classifiers that generate" IF ..THEN rules" which are easily interpretable.	Deep learning algorithms can be difficult to interpret. In this methodology, an attempt to visualize the generated feature maps was made. It was noticed that as one goes deeper into the network, the abstraction level of the feature maps increases. The subsequent convolution layers appear to be detecting more complex shapes, textures and mesh patterns, which are difficult for the human eye to discern.

9.2 SCIENTIFIC CONTRIBUTIONS

In view of the research outcomes, the main contributions of this research work are as below:

i) Reviewing common faults in rotating machinery of power-plants to identify the most vulnerable component, i.e. bearing, and surveying signal processing techniques and artificial intelligence techniques for intelligent fault diagnosis using vibration signature analysis.

ii) Employing Empirical Mode Decomposition for pre-processing of vibration signals in order to make the proposed methodologies more effective and robust towards data collected from a noisy environment such as a real power plant.

iii) Generating a comprehensive list of time-domain statistical features and proposing new advanced statistical features, Hjorth Parameters and Normal negative log likelihood for Gaussian Mixture Model (GMM) for rolling element ball bearing fault detection and severity estimation.

iv) Devising a strategy to probe a comprehensive set of statistical time domain features with the objective of identifying optimal feature subset to feed as input to ensemble of Machine Learning classifiers for rolling element ball bearing fault detection and severity estimation and validating the optimal feature subset using data from real power plant. Proposed strategy can be used in future studies as a possible guideline when newly proposed features are used for vibration signature analysis along with machine learning algorithms.

v) Eliminating the need for manual feature extraction by proposing a Deep Transfer Learning based methodology that can efficiently analyze Bispectrum images of vibration signals for fault diagnosis. This methodology solves the most omnipresent problem that is faced with diagnosing most of the datasets - shortage of data samples - in this research, ball bearing vibration dataset.

vi) Visualizing the extracted feature maps, associated with intermediate convolution layers for one of the Deep Transfer models, to investigate the feature extraction ability of deep neural networks.

vii) Achieving Knowledge Transfer from laboratories to the real world by training the

proposed methodologies on vibration lab data and testing it on operational power plant data.

9.3 SCOPE FOR FUTURE WORK

Based on the findings of this research, following recommendations are made for future research in the field of Fault Diagnosis of rotating machinery components of Power Plant using Vibration Signature Analysis:

- In this research work, the fault detection capability of proposed novel algorithms was validated using experimental data containing vibration signals of bearings with a single type of fault. In future, vibration signals for multiple simultaneous faults in the components can be acquired and classified.
- The research work can be extended to accommodate fault diagnosis of other faulty components e.g. gear, rotor etc. by acquiring their vibration signals from vibration labs and power plants and testing the robustness and effectiveness of the proposed methodologies.
- Higher Order Spectral Analysis using Deep Transfer Learning can be further investigated for fault diagnosis by shifting the focus to orders higher than Bispectrum such as Trispectrum and Composite spectrum.
- An attempt is made in this research to visualize the feature maps to gain an insight into the classification severity detection performed by it but more in-depth investigations are necessary to have a better understanding of adaptive feature extraction capability of Deep Learning algorithms.

REFERENCES

[1] O. Pinkus, "Manual of bearing failures and repair in power plant rotating equipment," Electric Power Research Inst., Palo Alto, CA ,United States, Tech. Rep., 1991.

[2] C. De Michelis, C. Rinaldi, C. Sampietri, and R. Vario, "Condition monitoring and assessment of power plant components," in *Power Plant Life Management and Performance Improvement*. Elsevier, 2011, pp. 38–109.

[3] C. Kurien and A. K. Srivastava, "Condition monitoring of systems in thermal power plant for vibration, motor signature, noise and wear debris analysis," *World Scientific News*, vol. 91, pp. 31–43, 2018.

[4] C. Kurien and A. K. Srivastava, "Case study on the effectiveness of condition monitoring techniques for fault diagnosis of pumps in thermal power plant," *Mechanics and Mechanical Engineering*, vol. 23, no. 1, pp. 70–75, 2019.

[5] U. Suedmersen and W. Fortec-Forschungstechnik, "Transient vibration signature analysis at steam and gas turbines," 2006.

[6] G. Boon and K. De Bauw, "A journey through 30 years of vibration analysis on large turbines: a history of progress in technology and experience," 2009.

[7] H. D. M. de Azevedo, A. M. Araújo, and N. Bouchonneau, "A review of wind turbine bearing condition monitoring: State of the art and challenges," *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 368–379, 2016.

[8] J. K. Sinha, *Industrial Approaches in Vibration-Based Condition Monitoring*. CRC Press, 2020.

[9] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications," *Mechanical Systems and Signal Processing*, vol. 42, no. 1-2, pp. 314–334, 2014.

[10] A. K. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483–1510, 2006.

[11] M.-Y. Chow, "Guest editorial special section on motor fault detection and diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 47, no. 5, pp. 982–983, 2000.

[12] D. Brown, "Machine-condition monitoring using vibration analysis: a case study from a nuclear power-plant," *Brüel & Kjaer application notes*, vol. 209, p. 12, 1985.

[13] T. Barszcz, *Vibration-Based Condition Monitoring of Wind Turbines*. Springer, 2019.

[14] A. Rai and S. Upadhyay, "A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings," *Tribology International*, vol. 96, pp. 289–306, 2016.

[15] A. Kumar and R. Kumar, "Role of signal processing, modeling and decision making in the diagnosis of rolling element bearing defect: a review," *Journal of Nondestructive Evaluation*, vol. 38, no. 1, p. 5, 2019.

[16] N. Tandon and A. Choudhury, "A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings," *Tribology International*, vol. 32, no. 8, pp. 469–480, 1999.

[17] M. Lebold, K. McClintic, R. Campbell, C. Byington, and K. Maynard, "Review of vibration analysis methods for gearbox diagnostics and prognostics," in *Proceedings of the 54th meeting of the society for machinery failure prevention technology*, vol. 634, 2000, p. 16.

[18] Y. Wei, Y. Li, M. Xu, and W. Huang, "A review of early fault diagnosis approaches and their applications in rotating machinery," *Entropy*, vol. 21, no. 4, p. 409, 2019.

[19] A. Othman, N. S. Damanhuri, and N. Hamzah, "Fault diagnosis using feature extraction in power plant rotating machinery," *Institute of Research, Development and Commercialization*, 2009.

[20] H. Oh, J. H. Jung, B. C. Jeon, and B. D. Youn, "Scalable and unsupervised feature engineering using vibration-imaging and deep learning for rotor system diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 4, pp. 3539–3549, 2017.

[21] Q. Hu, Z. J. He, Y. Zi, Z. S. Zhang, and Y. Lei, "Intelligent fault diagnosis in power plant using empirical mode decomposition, fuzzy feature extraction and support vector machines," in *Key Engineering Materials*, vol. 293. Trans Tech Publ, 2005, pp. 373–382.

[22] Y. He, Z. Zhao, T. Luo, X. Lu, and J. Luo, "Failure analysis of journal bearing used in turboset of a power plant," *Materials & Design (1980-2015)*, vol. 52, pp. 923–931, 2013.

[23] M. Mehdizadeh and F. Khodabakhshi, "An investigation into failure analysis of interfering part of a steam turbine journal bearing," *Case Studies in Engineering Failure Analysis*, vol. 2, no. 2, pp. 61–68, 2014.

[24] N. Tazi, E. Chatelet, and Y. Bouzidi, "Wear analysis of wind turbine bearings," *International Journal of Renewable Energy Research (IJRER)*, vol. 7, no. 4, pp. 2120–2129, 2017.

[25] G. Pino, J. R. Ribas, and L. F. Guimarães, "Bearing diagnostics of hydro power plants using wavelet packet transform and a hidden markov model with orbit curves," *Shock and Vibration*, vol. 2018, 2018.

[26] H. Yang, J. Mathew, and L. Ma, "Vibration feature extraction techniques for fault diagnosis of rotating machinery: a literature survey," in *Asia-Pacific Vibration Conference*, no. 42460, 2003, pp. 801–807.

[27] Ş. N. Engin, K. Gülez, and M. Badi, "Advanced signal processing techniques for faulth diagnostics-a review," *Mathematical and Computational Applications*, vol. 4, no. 2, pp. 121–136, 1999.

[28] K. F. Tom, "A primer on vibrational ball bearing feature generation for prognostics and diagnostics algorithms," Army Research Lab Adelphi MD Sensors And Electron Devices Directorate, Tech. Rep., 2015.

[29] W. Caesarendra and T. Tjahjowidodo, "A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing," *Machines*, vol. 5, no. 4, p. 21, 2017.

[30] S. Fu, K. Liu, Y. Xu, and Y. Liu, "Rolling bearing diagnosing method based on time domain analysis and adaptive fuzzy-means clustering," *Shock and Vibration*, vol. 2016, 2016.

[31] L. Lu, J. Yan, and C. W. de Silva, "Dominant feature selection for the fault diagnosis of rotary machines using modified genetic algorithm and empirical mode decomposition," *Journal of Sound and Vibration*, vol. 344, pp. 464–483, 2015.

[32] R. K. Patel and V. Giri, "Feature selection and classification of mechanical fault of an induction motor using random forest classifier," *Perspectives in Science*, vol. 8, pp. 334–337, 2016.

[33] O. Janssens, V. Slavkovikj, B. Vervisch, K. Stockman, M. Loccufier, S. Verstockt, R. Van de Walle, and S. Van Hoecke, "Convolutional neural network based fault detection for rotating machinery," *Journal of Sound and Vibration*, vol. 377, pp. 331–345, 2016.

[34] P. K. Kankar, S. C. Sharma, and S. P. Harsha, "Fault diagnosis of ball bearings using machine learning methods," *Expert Systems with applications*, vol. 38, no. 3, pp. 1876–1886, 2011.

[35] M. Zeng, Y. Yang, J. Zheng, and J. Cheng, "Maximum margin classification based on flexible convex hulls for fault diagnosis of roller bearings," *Mechanical Systems and Signal Processing*, vol. 66, pp. 533–545, 2016.

[36] X. Wang, Y. Zheng, Z. Zhao, and J. Wang, "Bearing fault diagnosis based on statistical locally linear embedding," *Sensors*, vol. 15, no. 7, pp. 16 225–16 247, 2015.

[37] M. D. Prieto, G. Cirrincione, A. G. Espinosa, J. A. Ortega, and H. Henao, "Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 8, pp. 3398–3407, 2012.

[38] D. Dou and S. Zhou, "Comparison of four direct classification methods for intelligent fault diagnosis of rotating machinery," *Applied Soft Computing*, vol. 46, pp. 459–468, 2016.

[39] X. Yan and M. Jia, "A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," *Neurocomputing*, vol. 313, pp. 47–64, 2018.

[40] B. Sreejith, A. Verma, and A. Srividya, "Fault diagnosis of rolling element bearing using time-domain features and neural networks," in *2008 Third international Conference on Industrial and Information Systems*. IEEE, 2008, pp. 1–6.

[41] S. Abbasion, A. Rafsanjani, A. Farshidianfar, and N. Irani, "Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine," *Mechanical Systems and Signal Processing*, vol. 21, no. 7, pp. 2933–2945, 2007.

[42] P. E. William and M. W. Hoffman, "Identification of bearing faults using time domain zero-crossings," *Mechanical Systems and Signal Processing*, vol. 25, no. 8, pp. 3078–3088, 2011.

[43] K. A. Korba and F. Arbaoui, "SVM multi-classification of induction machine's bearings defects using vibratory analysis based on empirical mode decomposition," *International Journal of Applied Engineering Research*, vol. 13, no. 9, pp. 6579–6586, 2018.

[44] V. Sugumaran and K. Ramachandran, "Fault diagnosis of roller bearing using fuzzy classifier and histogram features with focus on automatic rule learning," *Expert Systems with Applications*, vol. 38, no. 5, pp. 4901–4907, 2011.

[45] C. Grover and N. Turk, "Rolling element bearing fault diagnosis using empirical mode decomposition and hjorth parameters," *Procedia Computer Science*, vol. 167, pp. 1484–1494, 2020.

[46] Y. Fan, C. Zhang, Y. Xue, J. Wang, and F. Gu, "A bearing fault diagnosis using a support vector machine optimised by the self-regulating particle swarm," *Shock and Vibration*, vol. 2020, 2020.

[47] Y. Chen, T. Zhang, Z. Luo, and K. Sun, "A novel rolling bearing fault diagnosis and severity analysis method," *Applied Sciences*, vol. 9, no. 11, p. 2356, 2019.

[48] Q.-q. Chen, S.-w. Dai, and H.-d. Dai, "A rolling bearing fault diagnosis method based on emd and quantile permutation entropy," *Mathematical Problems in Engineering*, vol. 2019, 2019.

[49] A. Kumar, C. Gandhi, Y. Zhou, H. Tang, and J. Xiang, "Fault diagnosis of rolling element bearing based on symmetric cross entropy of neutrosophic sets," *Measurement*, vol. 152, p. 107318, 2020.

[50] A. Kumar, C. Gandhi, Y. Zhou, R. Kumar, and J. Xiang, "Variational mode decomposition based symmetric single valued neutrosophic cross entropy measure for the identification of bearing defects in a centrifugal pump," *Applied Acoustics*, vol. 165, p. 107294, 2020.

[51] P. McFadden and M. Toozhy, "Application of synchronous averaging to vibration monitoring of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 14, no. 6, pp. 891–906, 2000.

[52] W. Wang, "Early detection of gear tooth cracking using the resonance demodulation technique," *Mechanical Systems and Signal Processing*, vol. 15, no. 5, pp. 887–903, 2001.

[53] G. Ibrahim, A. Albarbar, A. Abouhnik, and R. Shnibha, "Adaptive filtering based system for extracting gearbox condition feature from the measured vibrations," *Measurement*, vol. 46, no. 6, pp. 2029–2034, 2013.

[54] R. Alfredson and J. Mathew, "Frequency domain methods for monitoring the condition of rolling element bearings," *Transactions of the Institution of Engineers, Australia. Mechanical engineering*, vol. 10, no. 2, pp. 108–112, 1985.

[55] J. K. Sinha, "Bispectrum of a rotating shaft with a breathing crack," *Advances in Vibration Engineering*, vol. 7, no. 4, pp. 301–310, 2008.

[56] J. K. Sinha, "Higher order spectra for crack and misalignment identification in the shaft of a rotating machine," *Structural Health Monitoring*, vol. 6, no. 4, pp. 325–334, 2007.

[57] J. K. Sinha and K. Elbhbah, "A future possibility of vibration based condition monitoring of rotating machines," *Mechanical Systems and Signal Processing*, vol. 34, no. 1-2, pp. 231–240, 2013.

[58] L. Saidi, J. B. Ali, and F. Fnaiech, "Bi-spectrum based-emd applied to the non-stationary vibration signals for bearing faults diagnosis," *ISA transactions*, vol. 53, no. 5, pp. 1650–1660, 2014.

[59] L. Saidi, J. B. Ali, and F. Fnaiech, "Application of higher order spectral features and support vector machines for bearing faults classification," *ISA transactions*, vol. 54, pp. 193–206, 2015.

[60] W. Liu and J. Han, "Rolling element bearing fault recognition approach based on fuzzy clustering bispectrum estimation," *Shock and Vibration*, vol. 20, no. 2, pp. 213–225, 2013.

[61] F. Wang and L. Fang, "Bispectrum texture feature manifold for feature extraction in rolling bear fault diagnosis," *Mathematical Problems in Engineering*, vol. 2019, 2019.

[62] K. Elbhbah and J. K. Sinha, "Vibration-based condition monitoring of rotating machines using a machine composite spectrum," *Journal of Sound and Vibration*, vol. 332, no. 11, pp. 2831–2845, 2013.

[63] X. Xiong, S. Yang, and C. Gan, "A new procedure for extracting fault feature of multi-frequency signal from rotating machinery," *Mechanical Systems and Signal Processing*, vol. 32, pp. 306–319, 2012.

[64] D.-M. Yang, A. Stronach, P. MacConnell, and J. Penman, "Third-order spectral techniques for the diagnosis of motor bearing condition using artificial neural networks," *Mechanical Systems and Signal Processing*, vol. 16, no. 2-3, pp. 391–411, 2002.

[65] C. Capdessus, M. Sidahmed, and J. Lacoume, "Cyclostationary processes: application in gear faults early diagnosis," *Mechanical Systems and Signal Processing*, vol. 14, no. 3, pp. 371–385, 2000.

[66] W. A. Gardner, A. Napolitano, and L. Paura, "Cyclostationarity: Half a century of research," *Signal processing*, vol. 86, no. 4, pp. 639–697, 2006.

[67] L. Bouillaut and M. Sidahmed, "Cyclostationary approach and bilinear approach: comparison, applications to early diagnosis for helicopter gearbox and classification method based on HOCS," *Mechanical Systems and Signal Processing*, vol. 15, no. 5, pp. 923–943, 2001.

[68] C. J. Li, J. Ma, and B. Hwang, "Bearing condition monitoring by pattern recognition based on bicoherence analysis of vibrations," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 210, no. 3, pp. 277–285, 1996.

[69] Z. Feng, M. Liang, and F. Chu, "Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples," *Mechanical Systems and Signal Processing*, vol. 38, no. 1, pp. 165–205, 2013.

[70] F. Al-Badour, M. Sunar, and L. Cheded, "Vibration analysis of rotating machinery using time–frequency analysis and wavelet techniques," *Mechanical Systems and Signal Processing*, vol. 25, no. 6, pp. 2083–2101, 2011.

[71] V. Climente-Alarcon, J. Antonino-Daviu, M. Riera-Guasp, R. Puche-Panadero, and L. Escobar, "Application of the Wigner–Ville distribution for the detection of rotor asymmetries and eccentricity through high-order harmonics," *Electric Power Systems Research*, vol. 91, pp. 28–36, 2012.

[72] C. Hanxin, H. Yang, P. Chua, and L. G. Hian, "Fault diagnosis of water hydraulic motor by hilbert transform and adaptive spectrogram," in *2010 Prognostics and System Health Management Conference*. IEEE, 2010, pp. 1–6.

[73] Z. Peng and F. Chu, "Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography," *Mechanical Systems and Signal Processing*, vol. 18, no. 2, pp. 199–221, 2004.

[74] A. Djebala, N. Ouelaa, and N. Hamzaoui, "Detection of rolling bearing defects using discrete wavelet analysis," *Meccanica*, vol. 43, no. 3, pp. 339–348, 2008.

[75] N. Nikolaou and I. Antoniadis, "Rolling element bearing fault diagnosis using wavelet packets," *NDT & E International*, vol. 35, no. 3, pp. 197–205, 2002.

[76] H. Qiu, J. Lee, J. Lin, and G. Yu, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," *Journal of sound and vibration*, vol. 289, no. 4-5, pp. 1066–1090, 2006.

[77] J. Lin and L. Qu, "Feature extraction based on morlet wavelet and its application for mechanical fault diagnosis," *Journal of sound and vibration*, vol. 234, no. 1, pp. 135–148, 2000.

[78] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, 2018.

[79] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "Deep learning algorithms for bearing fault diagnostics-a review," in *2019 IEEE 12th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED)*. IEEE, 2019, pp. 257–263.

[80] R. Bal and S. Sharma, "Review on meta classification algorithms using WEKA," *International Journal of Computer Trends and Technology (IJCTT)*, vol. 35, pp. 38–47, 2016.

[81] Q. Wang, Y. B. Liu, X. He, S. Y. Liu, and J. H. Liu, "Fault diagnosis of bearing based on KPCA and KNN method," in *Advanced Materials Research*, vol. 986. Trans Tech Publ, 2014, pp. 1491–1496.

[82] A. Moosavian, H. Ahmadi, A. Tabatabaeefar, and M. Khazaee, "Comparison of two classifiers; K-nearest neighbor and artificial neural network, for fault diagnosis on a main engine journal-bearing," *Shock and Vibration*, vol. 20, no. 2, pp. 263–272, 2013.

[83] S. Mehta, X. Shen, J. Gou, and D. Niu, "A new nearest centroid neighbor classifier based on k local means using harmonic mean distance," *Information*, vol. 9, no. 9, p. 234, 2018.

[84] A. Sharma, R. Jigyasu, L. Mathew, and S. Chatterji, "Bearing fault diagnosis using weighted k-nearest neighbor," in *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, 2018, pp. 1132–1137.

[85] R. S. Gunerkar, A. K. Jalan, and S. U. Belgamwar, "Fault diagnosis of rolling element bearing based on artificial neural network," *Journal of Mechanical Science and Technology*, vol. 33, no. 2, pp. 505–511, 2019.

[86] J. B. Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech, "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals," *Applied Acoustics*, vol. 89, pp. 16–27, 2015.

[87] D. Dou, J. Jiang, Y. Wang, and Y. Zhang, "A rule-based classifier ensemble for fault diagnosis of rotating machinery," *Journal of Mechanical Science and Technology*, vol. 32, no. 6, pp. 2509–2515, 2018.

[88] Y. Li, J. Wang, L. Duan, T. Bai, X. Wang, Y. Zhang, and G. Qin, "Association rule-based feature mining for automated fault diagnosis of rolling bearing," *Shock and Vibration*, vol. 2019, 2019.

[89] J. G. Bazan, H. S. Nguyen, S. H. Nguyen, P. Synak, and J. Wróblewski, "Rough set algorithms in classification problem," in *Rough set methods and applications*. Springer, 2000, pp. 49–88.

[90] A. Wojna, "Analogy-based reasoning in classifier construction," in *Transactions on Rough Sets IV*. Springer, 2005, pp. 277–374.

[91] Y. Chen, G. Wang *et al.*, "Bearing fault diagnosis based on rough set," in *2010 2nd International Conference on Signal Processing Systems*, vol. 3. IEEE, 2010, pp. V3–706.

[92] X. Zhu, Y. Zhang, and Y. Zhu, "Intelligent fault diagnosis of rolling bearing based on kernel neighborhood rough sets and statistical features," *Journal of Mechanical Science and Technology*, vol. 26, no. 9, pp. 2649–2657, 2012.

[93] J. R. Yan, Y. Min, X. Cui, and Y. Huang, "Fault diagnosis of rolling bearing based on rough set and neural network," in *Applied Mechanics and Materials*, vol. 58. Trans Tech Publ, 2011, pp. 974–977.

[94] V. Muralidharan and V. Sugumaran, "Rough set based rule learning and fuzzy classification of wavelet features for fault diagnosis of monoblock centrifugal pump," *Measurement*, vol. 46, no. 9, pp. 3057–3063, 2013.

[95] W. Li, W. Pan, and S. Zhang, "Fault diagnosis using rough sets and BP networks," in *2010 International Conference on Mechanic Automation and Control Engineering*. IEEE, 2010, pp. 585–588.

[96] A. Wojna and R. Latkowski, "Rseslib 3: library of rough set and machine learning methods with extensible architecture," in *Transactions on Rough Sets XXI*. Springer, 2019, pp. 301–323.

[97] E. Zio, P. Baraldi, and G. Gola, "Feature-based classifier ensembles for diagnosing multiple faults in rotating machinery," *Applied Soft Computing*, vol. 8, no. 4, pp. 1365–1380, 2008.

[98] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: a review of classification and combining techniques," *Artificial Intelligence Review*, vol. 26, no. 3, pp. 159–190, 2006.

[99] T. G. Dietterich, "Ensemble methods in machine learning," in *International workshop on multiple classifier systems*. Springer, 2000, pp. 1–15.

[100] A. Sharma, M. Amarnath, and P. K. Kankar, "Novel ensemble techniques for classification of rolling element bearing faults," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 39, no. 3, pp. 709–724, 2017.

[101] N. Sikder, K. Bhakta, A. Al Nahid, and M. M. Islam, "Fault diagnosis of motor bearing using ensemble learning algorithm with FFT based preprocessing," in *International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*. IEEE, 2019, pp. 564–569.

[102] G. Xu, M. Liu, Z. Jiang, D. Söffker, and W. Shen, "Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning," *Sensors*, vol. 19, no. 5, p. 1088, 2019.

[103] P. Karimi and H. Jazayeri-Rad, "Comparing the fault diagnosis performances of single neural networks and two ensemble neural networks based on the boosting methods," *Journal of Automation and Control*, vol. 2, no. 1, pp. 21–32, 2014.

[104] L. I. Kuncheva, *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons, 2014.

[105] U.-P. Chong *et al.*, "Signal model-based fault detection and diagnosis for induction motors using features of vibration signal in two-dimension domain," *Strojniški vestnik*, vol. 57, no. 9, pp. 655–666, 2011.

[106] L. Hua, Y. Qiang, J. Gu, L. Chen, X. Zhang, and H. Zhu, "Mechanical fault diagnosis using color image recognition of vibration spectrogram based on quaternion invariable moment," *Mathematical problems in engineering*, vol. 2015, 2015.

[107] W. Li, M. Qiu, Z. Zhu, B. Wu, and G. Zhou, "Bearing fault diagnosis based on spectrum images of vibration signals," *Measurement Science and Technology*, vol. 27, no. 3, p. 035005, 2016.

[108] S. A. Khan and J.-M. Kim, "Automated bearing fault diagnosis using 2d analysis of vibration acceleration signals under variable speed conditions," *Shock and Vibration*, vol. 2016, 2016.

[109] B. Zhou and Y. Cheng, "Fault diagnosis for rolling bearing under variable conditions based on image recognition," *Shock and Vibration*, vol. 2016, 2016.

[110] C. Lu, Y. Wang, M. Ragulskis, and Y. Cheng, "Fault diagnosis for rotating machinery: A method based on image processing," *PloS one*, vol. 11, no. 10, p. e0164111, 2016.

[111] Y. Wang and Y. Cheng, "An approach to fault diagnosis for gearbox based on image processing," *Shock and Vibration*, vol. 2016, 2016.

[112] G. Xu, M. Liu, Z. Jiang, W. Shen, and C. Huang, "Online fault diagnosis method based on transfer convolutional neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 2, pp. 509–520, 2019.

[113] J. Zhang, C.-W. Liu, F.-R. Bi, X.-B. Bi, and X. Yang, "Fault feature extraction of diesel engine based on bispectrum image fractal dimension," *Chinese Journal of Mechanical Engineering*, vol. 31, no. 1, p. 40, 2018.

[114] S. Ma and F. Chu, "Ensemble deep learning-based fault diagnosis of rotor bearing systems," *Computers in Industry*, vol. 105, pp. 143–152, 2019.

[115] D. Verstraete, A. Ferrada, E. L. Droguett, V. Meruane, and M. Modarres, "Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings," *Shock and Vibration*, vol. 2017, 2017.

[116] D.-T. Hoang and H.-J. Kang, "Rolling element bearing fault diagnosis using convolutional neural network and vibration image," *Cognitive Systems Research*, vol. 53, pp. 42–50, 2019.

[117] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mechanical Systems and Signal Processing*, vol. 115, pp. 213–237, 2019.

[118] S. Tang, S. Yuan, and Y. Zhu, "Deep learning-based intelligent fault diagnosis methods toward rotating machinery," *IEEE Access*, vol. 8, pp. 9335–9346, 2019.

[119] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "Machine learning and deep learning algorithms for bearing fault diagnostics–a comprehensive review," *arXiv preprint arXiv:1901.08247*, 2019.

[120] S. R. Saufi, Z. A. B. Ahmad, M. S. Leong, and M. H. Lim, "Challenges and opportunities of deep learning models for machinery fault detection and diagnosis: A review," *IEEE Access*, vol. 7, pp. 122 644–122 662, 2019.

[121] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "Deep learning algorithms for bearing fault diagnostics–a comprehensive review," *IEEE Access*, vol. 8, pp. 29 857–29 881, 2020.

[122] Z. Zhao, Q. Zhang, X. Yu, C. Sun, S. Wang, R. Yan, and X. Chen, "Unsupervised deep transfer learning for intelligent fault diagnosis: An open source and comparative study," *arXiv preprint arXiv:1912.12528*, 2019.

[123] J. Tao, Y. Liu, D. Yang, F. Tang, and C. Liu, "Fault diagnosis of rolling bearing using deep belief networks," in *2015 International Symposium on Material, Energy and Environment Engineering*. Atlantis Press, 2015.

[124] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[125] Y.-M. Hsueh, V. R. Ittangihal, W.-B. Wu, H.-C. Chang, and C.-C. Kuo, "Fault diagnosis system for induction motors by CNN using empirical wavelet transform," *Symmetry*, vol. 11, no. 10, p. 1212, 2019.

[126] Y. Xie and T. Zhang, "Fault diagnosis for rotating machinery based on convolutional neural network and empirical mode decomposition," *Shock and Vibration*, vol. 2017, 2017.

[127] I. Mitiche, M. D. Jenkins, P. Boreham, A. Nesbitt, B. G. Stewart, and G. Morison, "Deep residual neural network for EMI event classification using bispectrum representations," in *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 2018, pp. 186–190.

[128] M. Sohaib and J.-M. Kim, "Fault diagnosis of rotary machine bearings under inconsistent working conditions," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 6, pp. 3334–3347, 2019.

[129] J. Lucas, B. Calef, and T. Kyono, "Recovering astronomical images with deep neural network supported bispectrum processing," in *Adv. Maui Opt. and Space Surveill. Technol. Conf*, 2018.

[130] L. Gagliano, E. B. Assi, D. K. Nguyen, and M. Sawan, "Bispectrum and recurrent neural networks: Improved classification of interictal and preictal states," *Scientific Reports*, vol. 9, no. 1, pp. 1–9, 2019.

[131] E. B. Assi, L. Gagliano, S. Rihana, D. K. Nguyen, and M. Sawan, "Bispectrum features and multilayer perceptron classifier to enhance seizure prediction," *Scientific Reports*, vol. 8, no. 1, pp. 1–8, 2018.

[132] D. Neupane and J. Seok, "Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review," *IEEE Access*, vol. 8, pp. 93 155–93 178, 2020.

[133] "Case Western Reserve University Bearing Data Center," available online: https://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website.

[134] "Fault data sets - Society for Machinery Failure Prevention Technology," available online: https://mfpt.org/fault-data-sets/.

[135] M. F. S. User's Manual, "Spectraquest inc," 2005.

[136] "Nice Ball Bearing," http://rbcbearings.com/ballbearings/index.htm.

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

[138] Q. Gao, C. Duan, H. Fan, and Q. Meng, "Rotating machine fault diagnosis using empirical mode decomposition," *Mechanical Systems and Signal Processing*, vol. 22, no. 5, pp. 1072–1081, 2008.

[139] R. Yan and R. X. Gao, "Rotary machine health diagnosis based on empirical mode decomposition," *Journal of Vibration and Acoustics*, vol. 130, no. 2, 2008.

[140] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 35, no. 1-2, pp. 108–126, 2013.

[141] W. Yang, R. Court, P. J. Tavner, and C. J. Crabtree, "Bivariate empirical mode decomposition and its contribution to wind turbine condition monitoring," *Journal of Sound and Vibration*, vol. 330, no. 15, pp. 3766–3782, 2011.

[142] F. Wu and L. Qu, "An improved method for restraining the end effect in empirical mode decomposition and its applications to the fault diagnosis of large rotating machinery," *Journal of sound and vibration*, vol. 314, no. 3-5, pp. 586–602, 2008.

[143] H. Jiang, C. Li, and H. Li, "An improved eemd with multiwavelet packet for rotating machinery multi-fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 36, no. 2, pp. 225–239, 2013.

[144] W.-L. Qin, W.-J. Zhang, and C. Lu, "Rolling bearing fault diagnosis: A data-based method using EEMD, information entropy and one-versus-one SVM," in *2016 12th World Congress on Intelligent Control and Automation (WCICA)*. IEEE, 2016, pp. 1016–1020.

[145] S. Cho, M. R. Shahriar, and U. Chong, "Identification of significant intrinsic mode functions for the diagnosis of induction motor fault," *The Journal of the Acoustical Society of America*, vol. 136, no. 2, pp. EL72–EL77, 2014.

[146] H. Han, S. Cho, S. Kwon, and S.-B. Cho, "Fault diagnosis using improved complete ensemble empirical mode decomposition with adaptive noise and power-based intrinsic mode function selection algorithm," *Electronics*, vol. 7, no. 2, p. 16, 2018.

[147] A. Ayenu-Prah and N. Attoh-Okine, "A criterion for selecting relevant intrinsic mode functions in empirical mode decomposition," *Advances in Adaptive Data Analysis*, vol. 2, no. 01, pp. 1–24, 2010.

[148] A.-O. Boudraa and J.-C. Cexus, "EMD-based signal filtering," *IEEE Transactions on Instrumentation and Measurement*, vol. 56, no. 6, pp. 2196–2202, 2007.

[149] B. Nayana and P. Geethanjali, "Identification of bearing faults using statistical time domain features and fused time-domain descriptor features," *Journal of Advanced Research in Dynamical and Control Systems.*, vol. 10, pp. 17–26, 2018.

[150] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalography and clinical neurophysiology*, vol. 29, no. 3, pp. 306–310, 1970.

[151] F. Mormann, T. Kreuz, C. Rieke, R. G. Andrzejak, A. Kraskov, P. David, C. E. Elger, and K. Lehnertz, "On the predictability of epileptic seizures," *Clinical neurophysiology*, vol. 116, no. 3, pp. 569–587, 2005.

[152] T. Cecchin, R. Ranta, L. Koessler, O. Caspary, H. Vespignani, and L. Maillard, "Seizure lateralization in scalp EEG using hjorth parameters," *Clinical Neurophysiology*, vol. 121, no. 3, pp. 290–300, 2010.

[153] D. Migotina, A. Rosa, and A. Fred, "Automatic k-complex detection using hjorth parameters and fuzzy decision," in *Proceedings of the 2010 ACM Symposium on Applied Computing*, 2010, pp. 979–980.

[154] F. J. Martinez-Albaladejo, A. Bueno-Crespo, and G. Rodriguez-Bermudez, "Testing extreme learning machine in motor imagery brain computer interface," *Journal of Intelligent & Fuzzy Systems*, vol. 33, no. 5, pp. 3103–3111, 2017.

[155] D. Shon, K. Im, J.-H. Park, D.-S. Lim, B. Jang, and J.-M. Kim, "Emotional stress state detection using genetic algorithm-based feature selection on EEG signals," *International Journal of environmental research and public health*, vol. 15, no. 11, p. 2461, 2018.

[156] H. Cai, J. Han, Y. Chen, X. Sha, Z. Wang, B. Hu, J. Yang, L. Feng, Z. Ding, Y. Chen *et al.*, "A pervasive approach to EEG-based depression detection," *Complexity*, vol. 2018, 2018.

[157] A. Rizal, R. Hidayat, and H. A. Nugroho, "Lung sound classification using empirical mode decomposition and the hjorth descriptor," *American Journal of Applied Sciences*, vol. 14, no. 1, pp. 166–173, 2017.

[158] M. Kaboli, R. Walker, G. Cheng *et al.*, "In-hand object recognition via texture properties with robotic hands, artificial skin, and novel tactile descriptors," in *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2015, pp. 1155–1160.

[159] C. L. Nikias and M. R. Raghuveer, "Bispectrum estimation: A digital signal processing framework," *Proceedings of the IEEE*, vol. 75, no. 7, pp. 869–891, 1987.

[160] C. L. Nikias and J. M. Mendel, "Signal processing with higher-order spectra," *IEEE Signal processing magazine*, vol. 10, no. 3, pp. 10–37, 1993.

[161] A. Bommert, X. Sun, B. Bischl, J. Rahnenführer, and M. Lang, "Benchmark for filter methods for feature selection in high-dimensional classification data," *Computational Statistics & Data Analysis*, vol. 143, 2020.

[162] A. Kumar, C. Gandhi, X. Liu, Y. Liu, Y. Zhou, R. Kumar, and J. Xiang, "A novel health indicator developed using filter-based feature selection algorithm for the identification of rotor defects," *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 2020.

[163] S. Fong, R. P. Biuk-Aghai, and R. C. Millham, "Swarm search methods in weka for data mining," in *Proceedings of the 2018 10th International Conference on Machine Learning and Computing*, 2018, pp. 122–127.

[164] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: optimization by a colony of cooperating agents," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 26, no. 1, pp. 29–41, 1996.

[165] M. Kang, J. Kim, and J.-M. Kim, "Reliable fault diagnosis for incipient low-speed bearings using fault feature analysis based on a binary bat algorithm," *Information Sciences*, vol. 294, pp. 423–438, 2015.

[166] D. T. Pham and M. Castellani, "The bees algorithm: modelling foraging behaviour to solve continuous optimization problems," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 223, no. 12, pp. 2919–2938, 2009.

[167] X.-S. Yang and S. Deb, "Cuckoo search via lévy flights," in *2009 World congress on nature & biologically inspired computing (NaBIC)*. IEEE, 2009, pp. 210–214.

[168] S. Deb, S. Fong, and Z. Tian, "Elephant search algorithm for optimization problems," in *2015 Tenth International Conference on Digital Information Management (ICDIM)*. IEEE, 2015, pp. 249–255.

[169] X.-S. Yang and X. He, "Firefly algorithm: recent advances and applications," *arXiv preprint arXiv:1308.3898*, 2013.

[170] X.-S. Yang, "Flower pollination algorithm for global optimization," in *International conference on unconventional computing and natural computation*. Springer, 2012, pp. 240–249.

[171] R. Tang, S. Fong, X.-S. Yang, and S. Deb, "Wolf search algorithm with ephemeral memory," in *Seventh International Conference on Digital Information Management (ICDIM 2012)*. IEEE, 2012, pp. 165–172.

[172] G.-G. Wang, X.-Z. Gao, K. Zenger, and L. d. S. Coelho, "A novel metaheuristic algorithm inspired by rhino herd behavior," in *Proceedings of the 9th EUROSIM Congress on Modelling and Simulation, EUROSIM 2016*, no. 142. Linköping University Electronic Press, 2018, pp. 1026–1033.

[173] W. W. Cohen, "Fast effective rule induction," in *Machine learning proceedings 1995*. Elsevier, 1995, pp. 115–123.

[174] E. Frank and I. H. Witten, "Generating accurate rule sets without global optimization," *Machine Learning: Proceedings of the Fifteenth International Conference*, 1998.

[175] A. Wojna, R. Latkowski, and Ł. Kowalski, "Rseslib: User guide," 2009.

[176] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.

[177] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[178] Y. Freund, R. E. Schapire *et al.*, "Experiments with a new boosting algorithm," in *Machine Learning: Proceedings of the Thirteenth International Conference*, vol. 96, 1996, pp. 148–156.

[179] G. I. Webb, "Multiboosting: A technique for combining boosting and wagging," *Machine learning*, vol. 40, no. 2, pp. 159–196, 2000.

[180] J. Kittler, "Combining classifiers: A theoretical framework," *Pattern analysis and Applications*, vol. 1, no. 1, pp. 18–27, 1998.

[181] L. I. Kuncheva, "Combining pattern classifiers: Methods and algorithms (kuncheva, li; 2004)," *IEEE Transactions on Neural Networks*, vol. 18, no. 3, pp. 964–964, 2007.

[182] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[183] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[184] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2015.

[185] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[186] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[187] R. Zhang, H. Tao, L. Wu, and Y. Guan, "Transfer learning with neural networks for bearing fault diagnosis in changing working conditions," *IEEE Access*, vol. 5, pp. 14 347–14 357, 2017.

[188] F. Yang, W. Zhang, L. Tao, and J. Ma, "Transfer learning strategies for deep learning-based PHM algorithms," *Applied Sciences*, vol. 10, no. 7, p. 2361, 2020.

[189] C. Che, H. Wang, Q. Fu, and X. Ni, "Deep transfer learning for rolling bearing fault diagnosis under variable operating conditions," *Advances in Mechanical Engineering*, vol. 11, no. 12, pp. 1–11, 2019.

[190] Z. Wu, H. Jiang, K. Zhao, and X. Li, "An adaptive deep transfer learning method for bearing fault diagnosis," *Measurement*, vol. 151, p. 107227, 2020.

[191] M. J. Hasan and J.-M. Kim, "Bearing fault diagnosis under variable rotational speeds using stockwell transform-based vibration imaging and transfer learning," *Applied Sciences*, vol. 8, no. 12, p. 2357, 2018.

[192] J. Zhu, N. Chen, and C. Shen, "A new deep transfer learning method for bearing fault diagnosis under different working conditions," *IEEE Sensors Journal*, vol. 20, no. 15, pp. 8394–8402, 2020.

[193] P. Cao, S. Zhang, and J. Tang, "Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning," *IEEE Access*, vol. 6, pp. 26 241–26 253, 2018.

[194] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European Conference on Computer Vision*. Springer, 2014, pp. 818–833.

[195] J. Grezmak, P. Wang, C. Sun, and R. X. Gao, "Explainable convolutional neural network for gearbox fault diagnosis," *Procedia CIRP*, vol. 80, pp. 476–481, 2019.

BRIEF PROFILE OF THE RESEARCH SCHOLAR

Chhaya Grover is a Ph.D. research scholar in the Department of Electronics Engineering in J. C. Bose University of Science and Technology YMCA, Faridabad, Haryana, India. She holds a Bachelor of Technology degree in Electronics and Communication from Institute of Engineering and Technology, Lucknow, India and Master of Technology degree in Electronics and Communication from National Institute of Technology, Kurukshetra, India. She has more than twenty years of teaching experience and currently working in the Department of Electronics and Communication Engineering in JSS Academy of Technical Education, Noida, India. Her research area is vibration signature analysis using advanced signal and image processing techniques for condition monitoring and fault diagnosis.

LIST OF PUBLICATIONS OUT OF THESIS

List of Published Papers

Sl. No.	Title of Paper	Name of Journal where published	No.	Volume & Issue	Year
1.	Optimal Statistical Feature Subset Selection for Bearing Fault Detection and Severity Estimation https://doi.org/10.1155/2020/5742053	Shock and Vibration Referred Journal (SCIE, SCOPUS)	ISSN: 1070-9622	Vol 2020 Special issue on "Advances in Fault Diagnosis and Defect Detection in Mechanical and Civil Engineering"	2020
2.	Rolling Element Bearing Fault Diagnosis using Empirical Mode Decomposition and Hjorth Parameters https://doi.org/10.1016/j.procs.2020.03.359	Science Direct, Elsevier's Procedia Computer Science Referred Journal (SCOPUS)	ISSN: 1877-0509	Vol 167 (2020)	2020
3.	Rolling Element Bearing Fault detection using Statistical Features and Ensemble classifiers. https://doi.org/10.35940/ijeat.C4836.029320	International Journal of Engineering and Advanced Technology (Referred Journal)	ISSN: 2249-8958	Vol-9 Issue-3	2020
4.	Vibration Signature Analysis using Rough Sets and Analogy-Based Reasoning Classification. https://doi.org/10.35940/ijeat.C4836.029320	International Journal of Recent Technology and Engineering (Referred Journal)	ISSN: 2277-3878	Vol-8 Issue-6	2020

List of Accepted Papers

Sl. No.	Title of the Paper	Name of Journal	Present Status	Year
5.	A Novel Fault Diagnostic System for Rolling Element Bearings using Deep Transfer learning on Bispectrum Contour Maps	Elsevier's "Engineering Science and Technology, an International Journal" ISSN: 2215-0986 Referred Journal (SCIE, SCOPUS)	Accepted	2021

List of Communicated Papers

Sl. No.	Title of the Paper	Name of Journal	Present Status	Year
6.	Meta classifier approach for improved bearing fault diagnosis	Walailak Journal of Science and Technology ISSN: 2228-835X Referred Journal (SCOPUS)	Under Review	2021