

University of Windsor

Scholarship at UWindor

Electronic Theses and Dissertations

Theses, Dissertations, and Major Papers

2017

DATA-DRIVEN TECHNIQUES FOR DIAGNOSING BEARING DEFECTS IN INDUCTION MOTORS

Maryam Farajzadeh Zanajni
University of Windsor

Follow this and additional works at: <https://scholar.uwindsor.ca/etd>

Recommended Citation

Farajzadeh Zanajni, Maryam, "DATA-DRIVEN TECHNIQUES FOR DIAGNOSING BEARING DEFECTS IN INDUCTION MOTORS" (2017). *Electronic Theses and Dissertations*. 5980.
<https://scholar.uwindsor.ca/etd/5980>

This online database contains the full-text of PhD dissertations and Masters' theses of University of Windsor students from 1954 forward. These documents are made available for personal study and research purposes only, in accordance with the Canadian Copyright Act and the Creative Commons license—CC BY-NC-ND (Attribution, Non-Commercial, No Derivative Works). Under this license, works must always be attributed to the copyright holder (original author), cannot be used for any commercial purposes, and may not be altered. Any other use would require the permission of the copyright holder. Students may inquire about withdrawing their dissertation and/or thesis from this database. For additional inquiries, please contact the repository administrator via email (scholarship@uwindsor.ca) or by telephone at 519-253-3000ext. 3208.

**DATA-DRIVEN TECHNIQUES FOR DIAGNOSING BEARING
DEFECTS IN INDUCTION MOTORS**

by

Maryam Farajzadeh-Zanjani

A Thesis

Submitted to the Faculty of Graduate Studies
through the Department of Electrical and Computer Engineering
in Partial Fulfillment of the Requirements for
the Degree of Master of Applied Science
at the University of Windsor

Windsor, Ontario, Canada

2017

© 2017 Maryam Farajzadeh-Zanjani

Data-driven Techniques for Diagnosing Bearing Defects in Induction Motors

by

Maryam Farajzadeh-Zanjani

APPROVED BY:

L. Rueda

Department of Computer Science
University of Windsor

J. Wu

Department of Electrical and Computer Engineering
University of Windsor

M. Saif, Adviser

Department of Electrical and Computer Engineering
University of Windsor

March 28, 2017

Declaration of Co-Authorship and Previous Publication

I. DECLARATION OF CO-AUTHORSHIP

I hereby declare that this thesis incorporates material that is the outcome of my research under the supervision of Dr. Mehrdad Saif. This thesis also incorporates the results of my research undertaken in collaboration with Dr. Rueda and Dr. Razavi-Far.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis, and have obtained written permission from each of the co-authors to include the co-authored material in my thesis.

I certify that, with the above qualification, this thesis and the research to which it refers is the product of my own work.

II. DECLARATION OF PREVIOUS PUBLICATION

This thesis partly includes the original papers that have been previously submitted, to be submitted, or published in peer reviewed journals and conferences as provided in the following table.

I certify that I have obtained a written permission from the copyright owner(s) to include the above published material(s) in my thesis. I certify that the above material describes work completed during my registration as graduate student at the University of Windsor.

I declare that, to the best of my knowledge, my thesis does not infringe upon anyones copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with

Thesis Chapter	Publication title/full citation	Publication status
Chapters 1-6	M. Farajzadeh-Zanjani, R. Razavi-Far and M. Saif, "Efficient sampling techniques for ensemble learning and diagnosing bearing defects under class imbalanced condition," 2016 IEEE Symposium Series on Computational Intelligence (SSCI), Athens, Greece, 2016, pp. 1-7.	published
Chapters 1-6	M. Farajzadeh-Zanjani, R. Razavi-Far, M. Saif and L. Rueda, "Efficient feature extraction of vibration signals for diagnosing bearing defects in induction motors," 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, 2016, pp. 4504-451.	published
Chapters 1-6	M. Farajzadeh-Zanjani, R. Razavi-Far, M. Saif, J. Zarei and V. Palade, "Diagnosis of Bearing Defects in Induction Motors by Fuzzy-Neighborhood Density-Based Clustering," 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), Miami, FL, 2015, pp. 935-940.	published
Chapters 1-6	Farajzadeh-Zanjani, M., Razavi-Far, R., Saif, M. Dimensionality reduction-based diagnosis of bearing defects in induction motors. (Conference paper)	prepared to submit
Chapters 1-6	Razavi-Far, R., Farajzadeh-Zanjani, M., Saif, M. A manuscript prepared for submission in a Journal	prepared to submit

the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or

Institution.

Abstract

Induction motors are frequently used in many automated systems as a major driving force, and thus, their reliable performances are of predominant concerns. Induction motors are subject to different types of faults and an early detection of faults can reduce maintenance costs and prevent unscheduled downtime. Motor faults are generally related to three components: the stator, the rotor and/or the bearings. This study focuses on the fault diagnosis of the bearings, which is the major reason for failures in induction motors.

Data-driven fault diagnosis systems usually include a classification model which is supported by an efficient pre-processing unit. Various classifiers, which aim to diagnose multiple bearing defects (i.e., ball, inner race and outer race defects of different diameters), require well-processed data.

The pre-processing tasks plays a vital role for extracting informative features from the vibration signal, reducing the dimensionality of the features and selecting the best features from the feature pool. Once the vibration signal is perfectly analyzed and a proper feature subset is created, then fault classifiers can be trained. However, classification task can be difficult if the training dataset is not balanced. Induction motors usually operate under healthy condition (than faulty situation), thus the monitored vibration samples relate to the normal state of the system expected to be more than the samples of the faulty state. Here, in this work, this challenge is also considered so that the classification model needs to deal with class imbalance problem.

Acknowledgements

This work would not be possible without the help and support of many people. Writing this thesis has had a big impact on me. It has been a period of intense learning for me, not only in the scientific arena, but also on a personal level. I would like to reflect on the people who have supported and helped me so much throughout this period.

Firstly, I would like to express my sincere gratitude to my supervisor Dr. Mehrdad Saif, Dean of the Faculty of Engineering, for his patience, motivation, immense knowledge and for his financial support throughout my master study. The door to Dr. Saif office was always open whenever I ran into a trouble spot or had a question about my research or writing this thesis. He supported me to conduct research, collaborate and improve in academic environment and also gave me the area to be creative and work toward different approaches to reach the goal. I want to thank him for all of the opportunities I was given to conduct my research and further my thesis.

In addition, I would like to thank my husband Dr. Roozbeh Razavi-Far for his encouragement, kind guidance, great supports and tireless efforts. He always helped me to surpass difficulties encountered during my study and research.

I would like to thank my thesis committee members, Dr. Jonathan Wu from the Electrical and Computer Engineering Department, and Dr. Luis Rueda from School of Computer Science for taking the time out of their busy schedule to participate in my seminar, reviewing my thesis and providing insightful comments to improve this work. Without their passionate participation and input, this thesis could not have been successfully conducted.

Finally, my deepest appreciation goes to my parents and my brother for their unconditional love, encouragement and support.

Table of Contents

	Page
Declaration of Co-Authorship and Previous Publication	iii
Abstract	vi
Acknowledgements	vii
List of Tables	x
List of Figures	xi
List of Abbreviations	xiv
1 Introduction	1
1.1 Background	1
1.2 Bearing Data	1
1.3 Goals and Contributions	2
1.4 Related Work	3
1.5 Outline	4
2 Induction Motors Principles and Previous Works	7
2.1 Induction Motors	7
2.2 Literature Review	10
2.3 Motivations and Case Study	13
2.4 Summary	14
3 Data-driven Diagnostic Scheme	15
3.1 Segmentation	15
3.2 Feature Extraction (FE)	16
3.2.1 Time-Domain	16
3.2.2 Frequency-Domain	17
3.2.3 Time-Frequency Domain	18
3.2.4 Singular Spectrum Analysis (SSA)	22

3.3	Feature Reduction	24
3.3.1	Feature Selection (FS)	24
3.3.2	Dimensionality Reduction (DR)	25
3.4	Classifiers	29
3.5	Summary	29
4	Data-driven Diagnostic Scheme Under Class Imbalance Condition	31
4.1	Class Imbalanced Learning	31
4.2	Data-Level Approaches	33
4.2.1	SMOTE	34
4.2.2	EMI-OS	34
4.3	Ensemble-Based Approaches	36
4.3.1	RUSBoost	40
4.3.2	SMOTEBagging	40
4.4	Algorithm-Level Approaches	43
4.4.1	Weighted Extreme Learning Machine	43
4.5	Summary	44
5	Experimental Results	45
5.1	Performance Measures	45
5.2	First Experiment - Combination of WPT and LDA	48
5.3	Second Experiment - Combination of EMD and DR Methods	52
5.4	Third Experiment - Diagnostic System under Class Imbalance Condition	57
5.5	Summary	71
6	Conclusions	73
6.1	Improvements and Future Works	75
	Appendix A Copyright Permissions	90
	Vita Auctoris	93

List of Tables

2.1	Fault frequencies	9
5.1	Confusion matrix for the multi-class problem	47
5.2	Comparison of the classification evaluation factors obtained by each classifier in first scenario.	50
5.3	Comparison of the classification evaluation factors obtained by each classifier in second scenario.	51
5.4	Comparison of the classification evaluation factors obtained by each classifier in third scenario.	51
5.5	Number of samples in each of the created datasets	60
5.6	Common features selected by all FS methods under LCI condition . .	70
5.7	Common features selected by all FS methods under MCI condition .	70
5.8	Common features selected by all FS methods under HCI condition . .	71

List of Figures

2.1	Components of a ball bearing [1]	8
2.2	Schematic diagram of the bearing including the dimensions and frequency characteristics of the defects [2].	9
2.3	The extrapolated distribution of failure in IMs[3]	11
3.1	Three-levels wavelet packet decomposition.	20
3.2	Empirical mode decomposition of the vibration signal containing characteristics of the inner race defects.	21
4.1	major techniques to tackle class imbalance problem [4].	39
5.1	Raw vibration signals of different classes: normal, ball defect, inner race defect, outer race defect. The faulty signals include different diameters of defects simulating incipient faults.	49
5.2	Performance evaluations (i.e., F-measure, MCC, MAvG and ROC area) obtained by each classifier on the first scenario.	50
5.3	Performance evaluations (i.e., F-measure, MCC, MAvG and ROC area) obtained by each classifier on the second scenario.	52
5.4	Performance evaluations (i.e., F-measure, MCC, MAvG and ROC area) obtained by each classifier on the third scenario.	53
5.5	Two dimensional features achieved by different DR techniques.	54
5.6	The performance measures obtained by means of each classifier trained by means of the features of different DR techniques. F-measure, MCC, ROC area and MAvG are presented in different panels from top to bottom.	56

5.7	Boxplot illustrates the distribution of the performance measures (i.e., F-measure, MCC, MA _v G and ROC area) attained by all classifiers for different DR techniques, where solid circles represent the distribution of the performance measures, the red dashes stand for maximum and minimum values, the solid squares stand for the average value for each DR method, and the red crosses stand for 1 and 99 percentiles of the performance values.	57
5.8	Block diagram of the proposed diagnostic system under class imbalance condition	59
5.9	Boxplot represents distributions of the performance measures attained by each feature extraction techniques	62
5.10	Boxplot represents distributions of the performance measures attained by each feature selection techniques	63
5.11	Boxplot represents distributions of the performance measures attained by each dimensionality reduction techniques	64
5.12	Boxplot represents distributions of the performance measures attained by each CIL technique	65
5.13	Boxplot illustrates distributions of the performance measures on the scenario of the Low imbalance ratio attained by each feature extraction technique w.r.t. various CIL techniques (i.e., the mean values attained by each CIL technique are connected for a better visual evaluation)	66
5.14	Boxplot illustrates distributions of the performance measures on the scenario of the medium imbalance ratio attained by each feature extraction technique w.r.t. various CIL techniques (i.e., the mean values attained by each CIL technique are connected for a better visual evaluation)	67

- 5.15 Boxplot illustrates distributions of the performance measures on the scenario of the Low imbalance ratio attained by each feature reduction technique w.r.t. various CIL techniques (i.e., the mean values attained by each CIL technique are connected for a better visual evaluation) . 68
- 5.16 Boxplot illustrates distributions of the performance measures on the scenario of the high imbalance ratio attained by each feature reduction technique w.r.t. various CIL techniques (i.e., the mean values attained by each CIL technique are connected for a better visual evaluation) . 69

List of Abbreviations

IM	Induction Motor
CWRU	Case Western Reserve University
FFT	Fast Fourier Transformation
LMD	Local Mean Decomposition
RUS	Random Under-Sampling
ROS	Random Over-Sampling
SMOTE	Synthetic Minority Over-Sampling Technique
FE	Feature Extraction
FS	Feature Selection
FR	Feature Reduction
DR	Dimensionality Reduction
CIL	Class Imbalance Learning
LCI	Low Class Imbalance
MCI	Medium Class Imbalance
HCI	High Class Imbalance
CSDT	Cost-Sensitive Decision Trees
CSNN	Cost-Sensitive Neural Networks
CBO	Cluster-Based Oversampling

EM	Expectation Maximization
EMI-OS	EM Imputation-based Over-Sampling
BEV	Bagging Ensemble Variation
WELM	Weighted Extreme Learning Machine
WPT	Wavelet Packet Transform
EMD	Empirical Mode Decomposition
IMF	Intrinsic Mode Functions
SSA	Singular Spectrum Analysis
SVD	Singular Value Decomposition
LFS	Linear Forward Selection
SFS	Sequential Forward Selection
SFFS	Sequential Floating Forward Search
mRMR	Minimal-Redundancy-Maximal-Relevance
PCA	Principal Component Analysis
LLC	Locally Linear Coordination
Gvmt	Linear Discriminant Analysis
NCA	Neighborhood Component Analysis
MCML	Maximally Collapsing Metric Learning
ROC	Receiver Operating Characteristic
MCC	Matthews Correlation Coefficient

Chapter 1

Introduction

1.1 Background

Induction motors (IMs) play a significant role in industries and impact on a wide portion of industrial applications. Hence, IMs' performance and safety should be monitored to prevent unexpected failures and decrease downtime and maintenance cost of the system [5]. In other words, an appropriate monitoring technique to assess the automated system condition is needed to guarantee its reliability, efficiency and controllability. The monitored abnormalities in IMs are mainly related to defects which are occurred in critical components such as the bearings, the stator and the rotor [5]. However, bearing defects are the major reason for induction motors failure [6, 3].

1.2 Bearing Data

The Case Western Reserve University (CWRU) Bearing Datasets [7] are one of the most frequently used reference data in the area of bearing diagnostics. CWRU dataset contain the vibration data of the bearing under normal condition and also with different defects including the ball, inner race and outer race. Defects ranging from 0.007 inch in diameter to 0.028 inch in diameter. The vibration data for healthy and faulty condition are recorded for motor loads of 0 to 3 horsepower.

The CWRU data has some latent characteristics in which some vibration signals are dominated by classical bearing fault features, and others are less clear or visualize

other fault symptoms. Localized defects in bearings create a sequence of broadband impulse responses in the acceleration signal as the bearing components recurrently strike the fault. The exact location of the defect determines the nature of the impulse response sequence [8].

CWRU data are available for public and contain fan and drive end vibration data as well as rpm. The datasets have become a standard reference in recent years and motivated many researchers in the field of bearing fault detection and diagnostic [8] so that Smith and Randall perform an extensive analysis on these datasets and ranked them as easily diagnosable to non-diagnosable. In this study, some data from non-diagnosable datasets (based on their research) is used.

1.3 Goals and Contributions

The large portion of IM's faults is related to bearing defects, hence, early detection and diagnosis of faults, which occurred in this component, is a substantial task. The goal of this work is to design a data-driven fault diagnosis system to detect bearing faults. To this aim, firstly the state-of-the-art data-driven techniques to diagnose bearing defects are studied. The preprocessing task to analyze the vibration signal is an inevitable part of a data-driven system. Thus, different preprocessing techniques are investigated to find the efficient method that could reveal the fault index better than others.

Secondly, this work aims to diagnose the bearing defects under the multi-class imbalanced condition. As in real world applications, data samples are often collected under skewed-class distribution, there is a need for some techniques for the ease of classification in the class imbalance condition and to facilitate training the fault classifiers. Diagnosing multiple bearing defects under the class imbalance condition is a challenging task since most of the classifiers are mainly devised for the class-balance distribution of data. Moreover, some of the classifiers can handle the class-imbalanced

data for the binary class situations.

Finally, this work proposes a sampling technique to diagnose bearing defects under class imbalance condition. The proposed technique is compared with the state-of-the-art techniques.

1.4 Related Work

Many works have been focused on the processing of the bearing vibration signal to identify the system state, i.e., normal condition or any defects. Vibration signal is mainly analyzed in three different domains; Time, Frequency and Time-Frequency domains. Various fault diagnosis system make use of data-driven techniques to extract informative features from the bearing vibration signal. For instance, Ravi and Mohanty [9] used Fast Fourier Transformation (FFT) to analyze frequency-domain features, and Liu and Han [10] extracted several time-frequency features by means of Local Mean Decomposition (LMD). In addition, time-domain features have been extensively studied due to their insensitivity to the change in motor load and the need for low computational efforts [11].

Apart from feature extraction that is an essential task in bearing vibration analysis, a qualified intelligent classification algorithm is needed to diagnose samples of bearing vibration accurately and efficiently. Data-driven diagnostic techniques are usually make use of a classification algorithm to diagnose faults [12, 13]. Various classification algorithms such as fuzzy systems [14] and neural networks [15, 2] have been considered for fault diagnosis. In this work also intelligent fault classifiers are used to diagnose bearing defects.

These fault classifiers aim to distinguish faulty (e.g., outer race defect, inner race defect or ball defect) and normal samples and determine the bearing health condition. However, these fault classification algorithms are typically based on the assumption that number of faulty samples are almost equal to number of normal samples. In

other words, most of these fault classifiers are not designed for skewed-class data distribution, while collected data in industrial processes are often imbalanced [16]. In fact, IMs operate in the normal condition, hence, samples of normal class expected to be greater than faulty ones. Since class imbalance problem endangers the classification performance, some techniques should be applied to deal with this problem. Different approaches to handle class imbalance in the level of data and algorithms are also considered in this study.

In data-level approach, sampling techniques which aim to provide a balance dataset is used. These methods tries to re-balance the dataset by under-sampling of the major class, e.g., Random Under-Sampling (RUS) or over-sampling of the minor class, e.g., Random Over-Sampling (ROS). In this approach, the goal is to create a dataset with the equal class distribution that can be used by the most of the fault classifiers. The benefits and the drawbacks of RUS and ROS as simple sampling methods along with state-of-the-art sampling techniques such as synthetic minority over-sampling technique (SMOTE) [17] and a novel sampling techniques based on missing data imputation, are studied in this work.

There exists other methods to tackle class imbalance problem for instance by performing some modification on classification algorithms, i.e., defining the weight or the cost of contribution of samples in the classification task [18].

Another approach is combinations of sampling and ensemble schemes which can lead to more versatile systems and obtain better performance [4]. In this approach, a re-balanced set of samples obtained by the sampling methods is used in the ensemble of the fault classifiers (i.e., Adaboost.M1 and Bagging) for diagnosing bearing defects under the class imbalanced condition.

1.5 Outline

The subsequent chapters of this study are structured as follows:

Chapter 2 discusses about induction motor faults and fault characteristic frequencies. Moreover, some published papers which have studied fault diagnosis of induction motors are reviewed. The goal of this work to propose a diagnostic scheme for bearing fault detection is briefly explained.

Chapter 3 defines different module of a data-driven diagnostic scheme which is commonly inclusive of preprocessing and classification modules. Then, the state-of-the-art techniques that have been applied in many literatures for each modules are explained. These techniques are basically related to feature extraction in three different domains; Time, Frequency and Time-Frequency and, feature reduction. The detailed modules of this diagnostic scheme, the role of each module and the important fault classifiers are described in this chapter.

Moreover, in real world applications, data samples are often collected under skewed-class distribution and there is a need to develop fault classification module to deal with this problem. The details of this problem is discussed in next chapter.

Chapter 4 describes the problem of class imbalanced in designing a data-driven fault diagnosis system. Moreover, different approaches to handle the problem of class imbalance are addressed and explained. Besides, a novel sampling technique to re-balance the data samples is proposed in this chapter.

Chapter 5 relates to experimental results obtained from the three different data-driven diagnosis system that operates under balanced and imbalanced conditions. Firstly, the results of a diagnostic system which makes use of wavelet packet transform and linear discriminant analysis are presented. Secondly, the experimental results for the combination of empirical mode decomposition and five different dimensionality reduction methods are provided. Finally, a data driven diagnostic system inclusive of all the state-of-the-art preprocessing techniques and class imbalance learning methods is described. The performance of each module w.r.t. class imbalance techniques are also provided.

Chapter 6 contains the conclusion of this work. This chapter firstly discusses

about the outcome of two experiments which focus on preprocessing techniques. Then, the performance of data-driven diagnostic techniques under the class imbalance condition is evaluated and the conclusion is made. The improvements for future works are also provided in this Chapter.

Chapter 2

Induction Motors Principles and Previous Works

2.1 Induction Motors

Induction motors are one of the most important components in many automated systems, and thus, their safe and reliable operations are of paramount importance to guarantee the driving force needed for the systems [19]. IMs can experience various type of faults, and thus, an accurate and early diagnosis of faults can prevent breakdown and an unscheduled maintenance and decrease service cost [20]. This work aims to diagnose bearing defects in IMs. These faults are basically categorized into distributed and local defects [21]. The former category of defects includes surface roughness, waviness, misaligned races and off-size rolling elements, on the contrary, the local defects include cracks, pits and spalls on the rolling surfaces [22]. Previous studies on IMs' source of failure hold the view that faults originated from the bearing cover a large proportion of failure distribution (41%) in comparison with stator winding (37%), rotor (10%), and other parts (12%) [6, 23]. Hence IM's failure diagnosis is concerned with bearing condition analysis has attracted the attention of many researchers [19, 22].

Generally, these defects may occur in different parts of the bearing as depicted in Figure 2.1. However, spalling of the races can be considered as the most frequent defect [22].

Bearings are one of the most important components in induction motors, and their

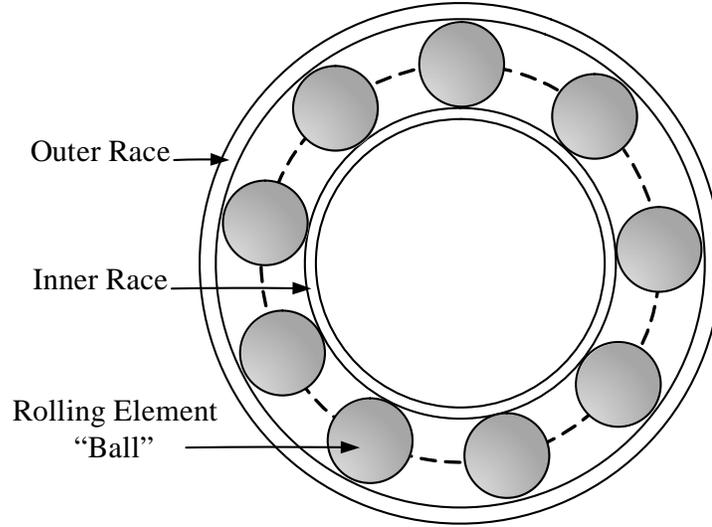


Figure 2.1 – Components of a ball bearing [1]

defects are the most frequent causes for machine breakdown. This has motivated a significant number of research efforts on vibration-based bearing defects diagnostics over the past twenty years [24]. Localized defects in bearings create a sequence of broadband impulse responses in the acceleration signal as the bearing components recurrently strike the fault. The exact location of the defect determines the nature of the impulse response sequence [8].

A number of works have studied characteristic defect frequency [22, 8, 25, 26, 27]. The frequency of the defect impulses can be explained by the following equations. These frequencies are also depicted in Figure 2.2.

Fundamental cage frequency:

$$F_c = \frac{1}{2}F_s \left(1 - \frac{D_b \cos(\theta)}{D_p} \right); \quad (2.1)$$

Ball defect frequency:

$$F_{bd} = \frac{D_p}{D_b} F_s \left(1 - \frac{D_b^2 \cos^2(\theta)}{D_p^2} \right); \quad (2.2)$$

Inner race defect frequency:

$$F_{id} = \frac{N}{2} F_s \left(1 + \frac{D_b \cos(\theta)}{D_p} \right); \quad (2.3)$$

Outer race defect frequency:

$$F_{od} = \frac{N}{2} F_s \left(1 - \frac{D_b \cos(\theta)}{D_p} \right). \quad (2.4)$$

where F_s is a shaft rotation frequency, D_b and D_p are the ball and the pitch diameters respectively, N is the number of rollers and θ is the bearing contact angle.

Drive end bearing defects frequencies are provided in Table 2.1[26].

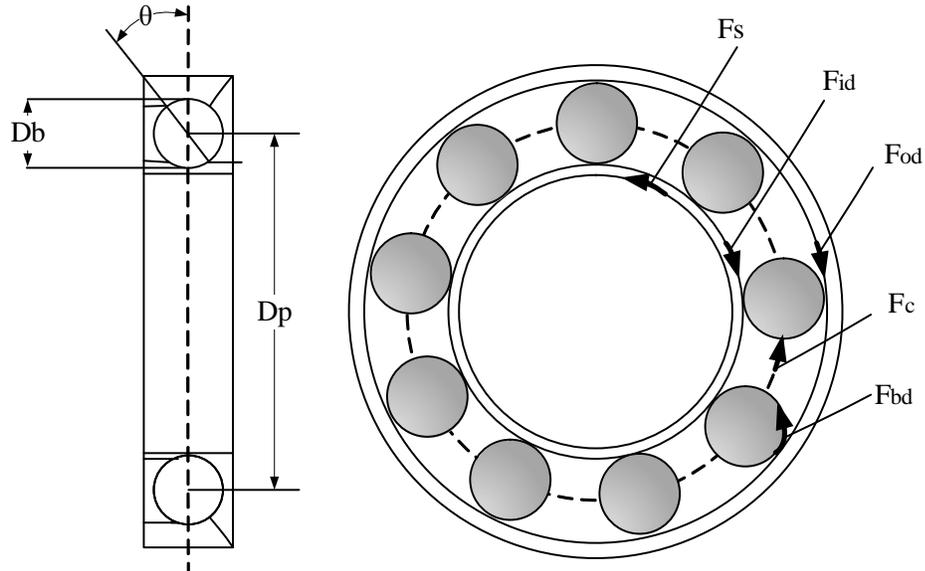


Figure 2.2 – Schematic diagram of the bearing including the dimensions and frequency characteristics of the defects [2].

Table 2.1 – Fault frequencies

Position	Inner Ring	Outer Ring	Cage Train	Rolling Element
Drive End	5.4152	3.5848	0.3983	4.7135

2.2 Literature Review

IMs have been extensively used as a primary source of power in various industries. As a result of their simplicity and rigidity, IMs gained a key role in many industries such as power plants, aerospace and petrochemical industries [28]. However, components aging and different type of defects can result in IMs failure and even system breakdown [29] and, thus, prompt diagnosis of IMs defects is strategically essential for industries, not only to improve reliability, efficiency and productivity of the systems, but also to minimize the maintenance cost [5, 30, 31, 32].

Various faults can occur in IMs that are typically categorized into two groups: mechanical damages (e.g., bearing defects, air gap eccentricity and broken rotor bars) and electrical damages (e.g., phase to ground, phase to phase and turn to turn connections in stator winding) [19, 23]. Recent studies on the source of faults in IMs show that bearing defects is the most frequent fault in IMs compared to stator winding, rotor bars and shaft/coupling [3, 23].

Moreover, a recent study [3] shows that bearing has the (69%) the distribution of induction motor faults. Figure 2.3 illustrates the extrapolated distribution of failure in IMs achieved by Austin H. Bonnet and Chunk Yung in their reliability paper[3]. Figure 2.3 shows that failures initiated in the bearing are covered almost two-thirds of fault distributions and the stator windings has almost one-fifth.

Hence, diagnosing bearing defects, which is the focus of this work, and performing prompt corrective actions can help preventing system failure and reduce the cost of unscheduled downtime.

Contemporary schemes to diagnose bearing defects in IMs are commonly based on the mathematical model of the system [33, 34]. However, performance of these diagnostic schemes highly depends on the accuracy of the model, which is not easy to obtain and subjected to inevitable assumptions and conditions. In this respect, data-driven model-free techniques have been extensively used for diagnosing faults in

IMs in recent years. This work also focuses on data-driven techniques for diagnosing bearing defects in IMs.

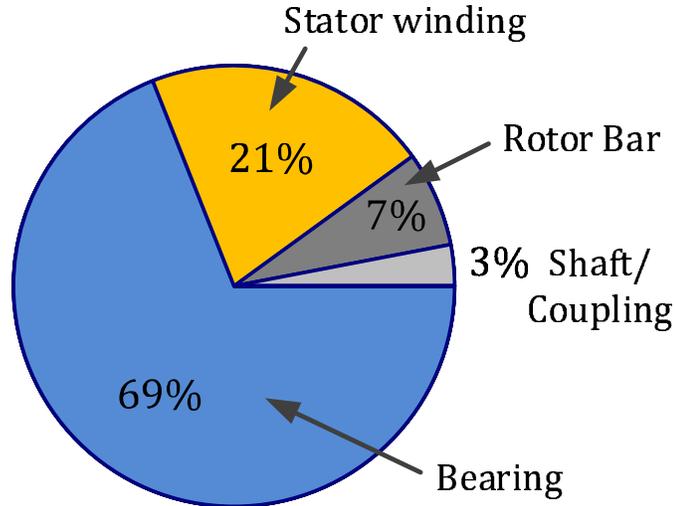


Figure 2.3 – The extrapolated distribution of failure in IMs[3]

Bearing defects can be divided into two categories of distributed and local [21]. The first group of defects contains surface waviness, roughness, off-size rolling elements and misaligned races. On the other hand, the local ones consist of spalls, pits and cracks over the races or the rolling elements [22]. These local defects are the most dominant mode of failure in IMs, which result in consecutive and periodic impulses in machine vibrations anytime a roller moves over the defective surface [35]. Then, a prevalent technique for diagnosing bearing defects in IMs is based on the processing of the raw vibrational signals to extract informative features for the use of fault classifiers [36, 37, 38, 39, 40, 41, 42, 43].

Various signal analysis techniques have been applied to the non-stationary and complex vibrational signals to extract discriminant features, including time-domain analysis, frequency-domain analysis (e.g., Fourier Transform) and time-frequency-domain analysis (e.g., Fast Fourier Transform, Wavelet Transform and Empirical Mode Decomposition) [22, 37, 38, 44, 45, 26, 46]. However, actual vibrational spectra

are usually represented by a considerable number of frequency or time-frequency components, i.e., high dimensional feature sets, which cannot be easily handled. These extracted features can be further processed by feature selection (FS) or dimensionality reduction (DR) techniques, in order to filter out redundant features, reduce the computational burden and provide a more informative and discriminant set of features for fault classifiers to recognize faulty and healthy states [47, 48].

The number of training samples for each class (i.e., normal and different classes of faults) plays a vital role in the performance of the diagnostic systems, however, in designing of the most diagnostic classifiers, it has been assumed that the observed samples from different classes have almost the same distribution, i.e., balanced class distribution. However, in real world applications, systems (e.g., induction motors) usually operate under the normal condition and, thus, samples of the normal state are expected to be more than samples of faulty classes. This can result in the collection of challenging imbalanced sets of samples, which further complicate the process of learning from the samples and diagnosing the faults. The main drawback with the imbalanced feature sets is that typical fault classifiers are usually biased w.r.t. the major class (i.e., normal state) and, therefore, there is a higher misclassification rate for the samples of the minor classes (i.e., faulty classes). Although a considerable number of data-driven techniques have been applied for diagnosing bearing defects in IMs [48, 49, 50, 8, 51], there is a need to design data-driven techniques for diagnosing bearing defects under the class imbalance conditions. Hence, in this work various state-of-the-art data-driven techniques have been designed for diagnosing bearing defects under the class imbalanced conditions.

One of the contributions of this study is such a general integrated scheme for diagnosing bearing defects from sets of features with class-imbalanced distributions of the samples. This diagnostic framework includes various state-of-the-art feature extraction, feature selection and dimensionality reduction techniques along with various advanced techniques for the class-imbalanced learning. These CIL techniques include

some state-of-the-art data-level, algorithm-level and ensemble-based approaches.

Besides, this work proposes a novel oversampling technique for class imbalance learning. The novelty of this approach is in generating a set of incomplete samples representative of the minor classes and imputing them by resorting to the expectation maximization algorithm to produce new synthetic samples of the minor classes. The proposed diagnostic scheme is verified w.r.t. the standard and widely-used Case Western Reserve University (CWRU) bearing datasets [7].

Over this diagnostic scheme, an empirical comparison of the performance of the data-driven techniques has been made with a triple objective. The former is to examine which technique outperforms the others in diagnosing bearing defects in IMs. The second one is to study the impacts of various feature extraction and reduction techniques in the diagnostic performance. The third one is the sensitivity analysis, where the performance of the diagnostic scheme is examined w.r.t. several datasets with different CI ratios. The diagnostic scheme is designed in a way that well-established conclusions can be extracted. The attained results show that the proposed novel technique outperforms other state-of-the-art class imbalance learning techniques in diagnosing bearing defects in terms of both performance measures and stability of the attained results.

2.3 Motivations and Case Study

The ultimate goal of this work is to design a diagnostic scheme which is able to detect multiple bearing defects (i.e., inner race defect, outer race defect and ball defect) in IMs, while the number of collected samples representative of each defect is less than the number of samples of the normal state. It is very common to collect vibration data with a skewed class distribution, since IMs often operate in the normal condition. A key requirement to design an efficient data-driven diagnostic system is then the ability of diagnosing multiple defects under the class imbalance problem.

Here, drive end and fan end bearing data from the Case Western Reserve University (CWRU) Bearing Data Center is used as a case study. More information about CWRU bearing data is also available in Chapter 1.

Various data-driven scenarios to diagnose bearing defects are investigated in this work. In order to analyze the efficiency of each scenario, CWRU bearing data is used. For each scenario, some datasets of CWRU are selected and the related methods are applied. Information related to the datasets used in each scenario is provided in Chapter 5.

2.4 Summary

This chapter provides basic information about induction motor faults and frequency characteristics. Different fault categories are mentioned; faults distribution and the importance of bearing faults are explained. After that the related works in the field of fault diagnosis of IMs are studied and some of these important literatures are referenced. Then, the goal of this study, to design a data-driven diagnostic system is stated. In the end, the motivation of the work which is multi-fault classification and the problem of class imbalance condition is briefly explained. Next chapter presents the data-driven diagnostic system including preprocessing module and intelligent fault classifiers.

Chapter 3

Data-driven Diagnostic Scheme

A data-driven diagnostic system generally consists of two main modules of preprocessing and classification. Preprocessing is an inevitable task for bearing fault detection. While the collected vibration signal is usually high frequency and non-stationary, there is a need for signal analysis to reveal the hidden characteristics and important signatures in normal and faulty signals. A well-processed signal provides the discriminant features for intelligent fault classifiers to enhance the accuracy of fault classification.

To this aim, preprocessing may consist of one or all of these tasks: segmentation, feature extraction, feature selection and dimensionality reduction. This chapter explains the state-of-the-art and widely used techniques for extracting important features from the data and analyzing the extracted features to reduce the dimension of the feature space and increase the accuracy of the diagnostic system.

3.1 Segmentation

Each collected dataset is indeed a vibration signal contains l representative samples. In this work, these samples are collected at 12kHz, which results into a time duration of approximately 10s. However, the collected samples during this interval represent various periods of the rotational procedure and, it is necessary therefore to segment the whole signal into successive intervals, resembling various non-overlapping samples collected at different time stamps. According to our experiments, fixing the length of segments to 1024 is a proper choice beyond which the performance measures begin to decrease. This value is fixed for all scenarios resulting into m non-overlapping

segments of length 1024.

3.2 Feature Extraction (FE)

Generally speaking, the efficiency of the fault detection techniques depends on the efficiency of the employed feature extraction (FE) methods that exploit informative features in fault diagnosis process. In other words, bearing fault diagnosis methods should process highly frequent raw signal in an efficient way to extract as many important features as possible which could deliver more discriminant information to classification model. These preprocessing methods mainly focus on three different domains; time, frequency, or timefrequency. In this study the state-of-the-art feature extraction methods commonly used in majority of data-driven diagnostic systems are explained and applied on CWRU bearing data.

3.2.1 Time-Domain

The simplest preprocessing method, also robust to load-changes issues, is the scheming of the statistical time-domain features. In order to obtain time-domain features of any vibration data in this study (e.g. Normal, Inner race defect, Ball defect and Outer race defect), firstly the vibration data is segmented with respect to the class labels. Once the representative samples of each class are divided into m none-overlapping folds (i.e., the intersection of all the folds is zero), different statistical measures are calculated for each fold. These time-domain measures are defined to create the feature vectors [*mean, root – mean – square, skewness, kurtosis, crest – factor, impulse – factor, margin – factor, entropy*].

Considering $k - th$ ($1 \leq k \leq m$) segment of the vibration signal \mathcal{S} , the eight features can be formulated in following, where l_k is the total number of samples in $k - th$ segment and x_{ik} stands for the $i - th$ sample in the $k - th$ segment. μ_k and σ_k^2 stand for the mean and the variance of the samples in the $k - th$ segment, respectively.

P stands for the probability mass function.

Mean

$$X_1 = \frac{\sum_{i=1}^{l_k} x_{ik}}{l_k} \quad (3.1)$$

Root Mean Square

$$X_2 = \left(\frac{\sum_{i=1}^{l_k} (x_{ik})^2}{l_k} \right)^{1/2} \quad (3.2)$$

Skewness

$$X_3 = \frac{\sum_{i=1}^{l_k} (x_{ik} - \mu_k)^3}{(l_k - 1)\sigma_k^3} \quad (3.3)$$

Kurtosis

$$X_4 = \frac{\sum_{i=1}^{l_k} (x_{ik} - \mu_k)^4}{(l_k - 1)\sigma_k^4} \quad (3.4)$$

Crest Factor

$$X_5 = \frac{\max(|x_{ik}|)}{\left(\frac{1}{l_k} \sum_{i=1}^{l_k} x_{ik}^2\right)^{1/2}} \quad (3.5)$$

Impulse Factor

$$X_6 = \frac{\max(|x_{ik}|)}{\frac{1}{l_k} \sum_{i=1}^{l_k} |x_{ik}|} \quad (3.6)$$

Margin Factor

$$X_7 = \frac{\max(|x_{ik}|)}{\left(\frac{1}{l_k} \sum_{i=1}^{l_k} \sqrt{|x_{ik}|}\right)^2} \quad (3.7)$$

Entropy

$$X_8 = \sum_{i=1}^{l_k} -P(x_{ik}) \log_2 P(x_{ik}) \quad (3.8)$$

3.2.2 Frequency-Domain

To extract frequency-domain features, the fundamental frequencies of the vibrational signals with the bearings defects and their amplitudes usually must be determined prior, which is not a realistic assumption in online monitoring applications, where the non-stationary conditions change the vibrational spectra. Here, the FE module

initially performs Fourier transform \mathcal{F} of each segment of the vibration signal, analyzes its frequency spectrum and extracts statistical features in the frequency domain. For the sake of a fair comparison, it calculates eight statistical features that can be formulated, similar to $\{X_1, \dots, X_8\}$ in Time-Domain section.

3.2.3 Time-Frequency Domain

Two state-of-the-art techniques are used to extract time-frequency domain features from the vibration signal. These dual-domain analysis techniques are Wavelet Packet Transform (WPT) and Empirical Mode Decomposition (EMD) that are widely used for non-stationary and nonlinear signal analysis. WPT and EMD map raw vibration data to more explanatory and useful features aiming to enhance classification accuracy [26, 52].

Wavelet Packet Transform (WPT)

WPT is an extension of wavelet transform, which aims to decompose signal into different frequency sub-bands with higher resolution and provide local structure analysis of the spectrum [53]. WPT makes use of low and high pass filters and iteratively decomposes both, approximation and details coefficients, into two parts, unlike Discrete Wavelet Transforms, which only consider the approximation coefficients for further decomposition; and, thus, it provides a richest signal analysis. WPT decomposition can be organized in a binary tree with leaves of equal-size that represent frequency sub-bands of the same width [52].

Figure 3.1 illustrates level by level decomposition of a signal up to level three. The time and the frequency presentation of a signal is on the top and on the leaves of a fully decomposed WPT tree, respectively. As each level of the tree is traversed there is an increase in the trade off between time and frequency resolution. In wavelet transforms, Figure 3.1, h_k and g_k are high pass and low pass wavelet filters, respectively. These are

known as quadratic mirror filters (QMF) that can be obtained by means of scaling function $\phi(t)$ and a selected wavelet function $\psi(t)$. WPT of a signal $x(t)$, can be explained by means of the following functions [52]:

$$W_{2n}(t) = \sqrt{2} \sum_k h(k) W_n(2t - k) \quad (3.9)$$

$$W_{2n+1}(t) = \sqrt{2} \sum_k g(k) W_n(2t - k) \quad (3.10)$$

where $W_0(t)$ can be described with $\phi(t)$ function, and $W_1(t)$ with the $\psi(t)$ function. Regarding to the recursive relation of components between the j^{th} and $(j+1)^{th}$ level, the signal can be decomposed as:

$$d_{j+1,2n} = \sum_m h(m - 2k) d_{j,n} \quad (3.11)$$

$$d_{j+1,2n+1} = \sum_m g(m - 2k) d_{j,n} \quad (3.12)$$

where $d_{j,n}$ stands for the wavelet coefficients of the level j in sub-band n , and m stands for the number of the wavelet coefficients.

In this study, the vibration signals are decomposed using discrete Meyer wavelet with Shannon entropy. Moreover, at each step of transformation, downsampling by 2 is performed leading to less number of samples in the final packs (i.e., leaves) for further decomposition. Given an arbitrary segment of time-domain vibration signal of l_k samples, WPT with a tree depth of j results in 2^j final packs. Each pack (i.e., $k - th$), at depth of j , contains around Q wavelet coefficients, where $Q = l_k/2^j$. The FE module then extract the eight different statistical features $\{X_1, \dots, X_8\}$ from the final 2^j packs of the tree by calculating the respective formula using the Q wavelet coefficients of each pack. This produces eight different statistical features for each wavelet packet and, thus, results in 8×2^j features. As an example, the wavelet packet decomposition which is done down to the fifth level ($j = 5$), results in 8×2^5 or 256

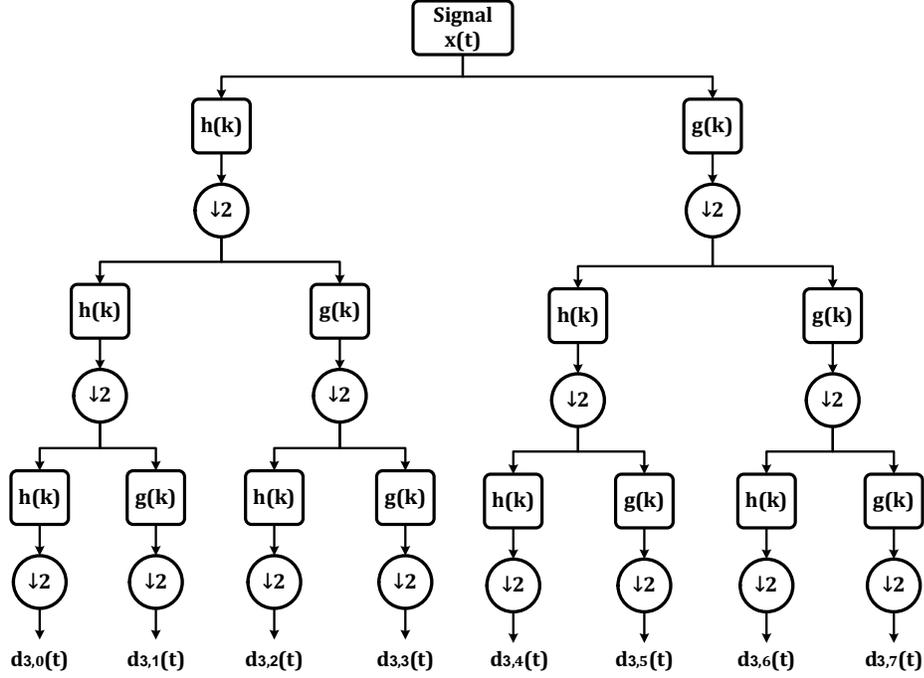


Figure 3.1 – Three-levels wavelet packet decomposition.

features.

Empirical Mode Decomposition (EMD)

This study makes use of a powerful signal processing technique, so-called Empirical Mode Decomposition (EMD) [54], in the first level of feature extraction unit. EMD is a data analysis technique that can be applied to any complicated sequence of data. EMD can deal with nonlinear and non-stationary signals while traditional signal processing techniques, such as Fourier Spectral analysis, work based on the linear and stationary assumptions. Thus, EMD as an adaptive processing technique is applied to fault diagnosis of motor vibration signals. Figure 3.2 illustrates the IMFs and their respective residual extracted from vibration signals for inner race defects. EMD considers local characteristic time scales to decompose signal into a finite number of components referred to as “intrinsic mode functions” or IMFs. These IMFs are based on two essential conditions [44, 55]:

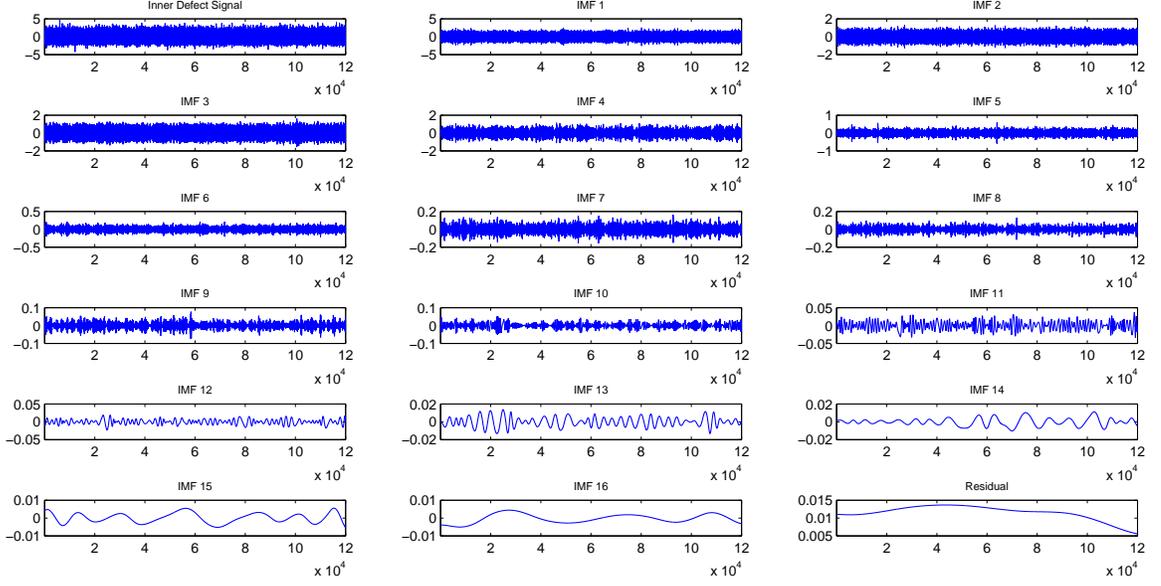


Figure 3.2 – Empirical mode decomposition of the vibration signal containing characteristics of the inner race defects.

1. The entire signal must contain either equal number of extrema and zero-crossings or the difference must be at most equal to one;
2. At each point, the mean value of the envelopes defined by local maxima and local minima becomes zero.

There exists a simple assumption that any signal can be explained by a set of different IMFs, where each IMF shows an oscillation mode among the data. Thus, given any signal \mathcal{S}_k , the decomposition method can be summarized as follows [55]:

1. Extract all the local extrema of \mathcal{S}_k .
2. Interpolate between the minima and the maxima to form lower and upper envelopes evp_{min} and evp_{max} , respectively, so that all the data points between them are completely covered.
3. Calculate the related mean $\mathcal{M} = \frac{evp_{min} + evp_{max}}{2}$.
4. Calculate the difference between the signal and the mean as $diff = \mathcal{S}_k - \mathcal{M}$.

5. If $diff$ satisfies two necessary conditions as an IMF, then it should be considered as the component $C_i = diff$. If not, then it can be treated as a sequence of data (i.e., \mathcal{S}_k), and returns to the first step to iterate the sifting procedure.
6. Compute C_i which is the finest scale or the shortest period of the i^{th} component, and then, subtract it from the data to determine the residual as follows $r_i = \mathcal{S}_k - C_i$
7. Iterate on the residual r_i if it still has at least 2 extrema.

At the end, the decomposition of the given signal \mathcal{S}_k results in a set of IMFs, $c_1, c_2, \dots, c_i, i = 1, \dots, \theta$ and a final residual term r_θ . By summing up the θ attained empirical modes and the corresponding residual r_θ , one can obtain \mathcal{S}_k as follows:

$$\mathcal{S}_k = \sum_{i=1}^{\theta} C_i + r_\theta \quad (3.13)$$

EMD decomposes the vibration signal into various IMFs with the same sample size as in the original signal. The size of IMFs is usually large, and thus, they can not be treated easily by means of online fault classifiers. To overcome this issue, eight statistical features, $\{X_1, \dots, X_8\}$, extracted from each intrinsic mode functions. As an example, selection of the first 7 IMFs and extraction of 8 fetures from them results in 56 (i.e., 7×8) statistical features in total.

3.2.4 Singular Spectrum Analysis (SSA)

SSA is another state-of-the-art technique to analyze the vibration signals. It makes use of the Hankel matrix and Singular Value Decomposition (SVD) for feature extraction [56, 57] and noise reduction from the vibration signals [58]. This can be considered as a non-parametric technique for time-series analysis, that aims to express any one-dimensional vibration signal $\mathcal{S}_k(t)$ by a collection of multiple independent

principal components, which may include white noise or different trends. This technique aims to extract the most informative components, i.e., new features, based on the respective singular values. This technique can be summarized as follows: (1) It initially embeds l_k data points of the signal $\mathcal{S}_k = (x_1, x_2, \dots, x_{l_k})$ into the Hankel (trajectory) matrix A as follows:

$$A = [X_1 : \dots : X_n] = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ x_2 & x_3 & \dots & x_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_m & x_{m+1} & \dots & x_{l_k} \end{pmatrix}_{m \times n} \quad (3.14)$$

Hence signal \mathcal{S}_k is projected into n lagged vector of size m , where $n = l_k - m + 1$ and $X_i = \{x_i, \dots, x_{i+m-1}\}^T \in \mathcal{R}^m$, ($i = 1, \dots, n$); (2) It then applies SVD on the trajectory matrix A to cancel out the noise, which results in two orthogonal matrices $U_{m \times m}$ and $V_{n \times n}$. Matrix A can be reconstructed through $A = U\Sigma V^T$, where Σ is a diagonal matrix, which composed of the square roots of the eigenvalues ζ_i of $A^T A$. The trajectory matrix A can be reformulated as follows:

$$A = \sum_{i=1}^q \sqrt{\zeta_i} U_i V_i^T \quad (3.15)$$

where $\{\zeta_1 \geq \zeta_2 \dots \geq \zeta_q \geq 0\}$ are eigenvalues, U_i denotes the respective orthonormal eigenvectors, $V_i = A^T U_i / \sqrt{\zeta_i}$ stand for the principal components and $q = \min(m, n)$; (3) It then splits the total set of m components into Δ disjoint sets and, consequently, groups the Eigentriples (i.e., $\sqrt{\zeta_i}, U_i, V_i$) of similar characteristics $\delta \in [1, \Delta]$, that are used then to reconstruct the trajectory matrix; (4) It then makes an average over the skew-diagonal elements of the newly grouped matrices to form a Hankel matrix. This is true since these values contribute into the same element in the newly derived vector $\psi_\delta = \{\psi_{\delta 1}, \psi_{\delta 2}, \dots, \psi_{\delta l_k}\} \in \mathcal{R}^{l_k}$, i.e., principal component. ψ_δ is indeed the projected vector of the δ -th disjoint matrix, which can be obtained through diagonal averaging

of the elements of the set as follows:

$$\psi_{\delta i} = \begin{cases} \frac{1}{i} \sum_{j=1}^i \phi_{j,i-j+1} & 1 \leq i \leq m \\ \frac{1}{m} \sum_{j=1}^m \phi_{j,i-j+1} & m \prec i \prec n \\ \frac{1}{l_k-i+1} \sum_{j=i-n+1}^m \phi_{j,i-j+1} & n \leq i \leq l_k \end{cases} \quad (3.16)$$

where $\phi_{j,i-j+1}$ stands for the elements of the δ -th disjoint matrix. It finally extracts reconstructed signals $\psi_{\delta} = \{\psi_{\delta 1}, \psi_{\delta 2}, \dots, \psi_{\delta l_k}\}$, $\delta = 1, \dots, \Delta$ as the most informative features. The FE module then calculates eight statistical features from each reconstructed signal. This results in extraction of 80 statistical features from each segment of the vibration signal (i.e., 8×10).

3.3 Feature Reduction

A crucial requirement for the practical implementation of an online diagnostic system is the capability of providing a small set of informative features, to guarantee efficient learning and immediate decision making and diagnosis. This can be performed by resorting to Feature Reduction (FR) techniques. The FR module (see the Figure 5.8) contains state-of-the-art techniques for Feature Selection (FS) and Dimensionality Reduction (DR).

3.3.1 Feature Selection (FS)

FS techniques aim to select and preserve a proper subset of n' informative features from all the extracted features n in order to attain more precise performance measures. FS techniques usually include a selection criterion and a search strategy. These techniques have been well studied and compared in [59, 60].

Here, two variants of Linear Forward Selection (LFS) [61] are used for feature selection. In general, LFS is a wrapper-based forward selection strategy, which can limit

the number of selected features in each forward selection step and, thus, significantly decreases the number of evaluations. These two LFS variants are Sequential Forward Selection (SFS) and Sequential Floating Forward Search (SFFS) [61]. Besides, in this work, another state-of-the-art technique, so-called Minimal-Redundancy-Maximal-Relevance (mRMR) [62], is used for feature selection.

SFS [61, 59] initially creates an empty subset, evaluates each candidate feature along with the current subset, i.e., previously selected features and, then, appends the best feature into the current subset. SFS algorithm, then, has a hill-climbing search mechanism and terminates when the preset number of features has been reached [61].

FSFS is a variant of floating techniques [63], which retreats features as long as the evaluation criterion is improving.

mRMR is a two-step FS algorithm, which merges the minimal-redundancy-maximal-relevance criterion and the wrapper [62]. This is a computationally inexpensive technique, which can find a small subset of informative features.

All of these FS techniques use the wrapper approach for the subset evaluation. The wrapper iteratively performs cross-validation and evaluates the estimated performance, i.e., F-measure, of a classifier to select a proper feature subset, i.e., those that maximize the F-measure during the search [64].

3.3.2 Dimensionality Reduction (DR)

DR techniques, on the other hand, aim to project the large scale set of features onto a lower dimensional data space so that the diagnosis process can be performed faster and even more efficient. Moreover, the obtained low dimensional features can decrease the storage space. These state-of-the-art dimensionality reduction techniques include two unsupervised approaches principal component analysis (PCA) [65] and locally linear coordination (LLC) [66]. The FR module also includes three supervised DR approaches: linear discriminant analysis (LDA) [67, 68], neighborhood component analysis (NCA) [69] and maximally collapsing metric learning (MCML) [70].

PCA is one of the most popular unsupervised technique for data analysis and feature reduction [71, 72]. PCA is a non-parametric technique which aims to extract latent information from data, while reducing the dimension of the feature space. PCA is indeed an orthogonal linear transformation of the patterns to a set of new coordinates in a way that the maximum variance by means of any projection of the patterns lies on the first coordinate, i.e., the first principal component, and the second largest variance lies on the second coordinate along with others.

LLC [66, 65] Locally Linear Coordination as an unsupervised dimensionality reduction method aims to map the different internal representation of the features into a single global coordination. The automatic alignment to new feature space is obtained by performing an eigensolver process on already trained model, unlike some methods that tries to adjust the objective function. The procedure summarized in two main steps: (1) create a mixture of local linear models using an EM-algorithm. (2) Perform local model alignment, (i.e. compute a weighted matrix by means of the obtained local models and their respective weights) to achieve the low-dimensional data representation [66].

Multi-Class LDA [67, 68] aims to discriminate the samples of different classes as much as possible by means of the linear combination so that the classifiers are able to determine the samples of each class easily and accurately in a new feature space. The goal of LDA, as an optimization problem, is finding the optimal projection matrix W while maximizing a separation function J which is defined as follows:

$$J = \frac{|W^T S_b W|}{|W^T S_w W|}, \quad (3.17)$$

in which S_b and S_w stand for the matrices of between class scatter and within class scatter, respectively. These matrices are computed by:

$$S_b = \frac{1}{n} \sum_{i=1}^c n_i (\mu_i - \mu)(\mu_i - \mu)^T \quad (3.18)$$

$$S_w = \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^{n_i} n_i (x_{ij} - \mu_i)(x_{ij} - \mu_i)^T \quad (3.19)$$

where n stands for the number of whole patterns, n_i indicates the number of patterns per i^{th} class, c stands for the number of class labels, μ_i is the sample mean for the i^{th} class, μ stands for the overall mean vector, and x_{ij} is the j^{th} sample in i^{th} class [68]. Given the generalized eigenvalue problem:

$$S_b W = \lambda S_w W \quad (3.20)$$

one can obtain the optimal projection matrix W , which maximizes the function J through eigenvalue decomposition of matrix $S_w^{-1} S_b$ and taking the eigenvector e_i which has the largest eigenvalue λ_i [68]. Finally, the projected data can be obtained by $y = W^T x$.

NCA Neighborhood Component Analysis [69] aims to learn a transformation matrix which maximizes the performance of the kNN classifier. NCA finds the expected leave-one-out performance by means of a random variant of kNN classification. Given a random arbitrary pattern x_i , it initially calculates the probability of drawing x_j as a neighborhood pattern for x_i as follows:

$$P_{ij} = \frac{\exp(-\|Wx_i - Wx_j\|^2)}{\sum_{k \neq i} \exp(-\|Wx_i - Wx_k\|^2)}, \quad P_{ii} = 0 \quad (3.21)$$

It then makes use of a random selection rule to compute the probability P_i where $i - th$ pattern can correctly classified as follows:

$$P_i = \sum_{j \in C_i} P_{ij} \quad (3.22)$$

where $C_i = \{j | c_j = c_i\}$ stands for a set of patterns, which belongs to the same class with x_i . The expected number of correctly classified patterns then can be computed

by:

$$f(W) = \sum_i \sum_{j \in C_i} P_{ij} = \sum_i P_i \quad (3.23)$$

NCA aims to maximize the above objective function by means of a gradient decent optimizer w.r.t. W which yields to a transformation matrix that best reveals similarity based on the class labels. NCA can project the data onto a low-dimensional feature space by restricting the matrix W in a non-square matrix format of size $q \times n$, ($q < n$). The projected feature space then can be computed by $y = Wx$.

MCML Maximally Collapsing Metric Learning (MCML) [70] is a supervised method that can perform dimensionality reduction on the data by means of a learned metric. MCML is based on the idea that an ideal metric is a metric in which the points belong to the same class could be projected into a location close to each other while the points of the other classes projected to a far location. The ideal metric can be achieved by solving the convex optimization problem which aims to collapse all patterns of the similar class to a particular point, and at the same time, project patterns of other classes into a very far distance point. MCML focus on learning Mahalanobis distances:

$$D[f(x_i), f(x_j)] = (x_i - x_j)^T A (x_i - x_j) \quad (3.24)$$

where x_i and x_j are two points in $X_{m \times n}$ space and A is positive semi-definite (PSD) matrix. Considering the most ideal scenario, when the distance of the patterns in the same class becomes zero and patterns of different classes are located on infinity far distance, it can be viewed as a linear projection, Wx , of mapping x while $A = W^T W$. MCML tries to learn matrix W such that the patterns of the same class mapped to a single location. This leads to create a convex optimization problem to find the optimal A that can be solved by means of a first order gradient method. Once the procedure for a full rank metric A is done, a low rank projection of A can also be calculated. After diagonalizing A into $A = \sum_{i=1}^n \lambda_i e_i e_i^T$, where $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_n$ are eigenvalues of

A and e_i are the corresponding eigenvectors, only q largest eigenvalues are considered. Hence, projecting the data onto a q -dimensional space can be specified by the rows of the low rank projection matrix $W_{q \times n}$. The result of projection is described as:

$$W = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_q}) [e_1^T; \dots; e_q^T] \quad (3.25)$$

3.4 Classifiers

The classifiers considered in this study are, Multi-Layer Perceptron (MLP), Naive Bayes (NB), k-nearest neighbor (kNN), random forest (RF) and decision tree (DT).

In order to estimate how accurately the fault classifiers perform in practice, 10-fold cross validation is applied. Then, the performances of different classifiers are evaluated. The performance measures considered in these experiments are weighted average of F-measure, weighted average of Matthews correlation coefficient (MCC), weighted average of receiver operating characteristic (ROC) area and Macro average geometric (MAvG) which are derived from the confusion matrix and explained in Chapter 5.

3.5 Summary

This chapter has discussed about a general data-driven diagnostic scheme inclusive of preprocessing and classification modules. The state-of-the-art techniques applied in majority of published papers to diagnose bearing defects are considered and explained in this chapter. Generally, the data-driven diagnostic system can be summarized as a four-step process, including segmentation, feature extraction, feature selection and fault classification. Each of these steps and the related techniques w.r.t three different domains, time-domain, frequency-domain and time-frequency domain, is discussed. In order to find the best scenario to diagnose bearing defects in multi-fault condition,

some of these techniques are implemented and applied on the bearing datasets and the experimental results are provided in Chapter 5. Moreover, considering that number of samples for class of normal is not always the same as the number of samples in minor classes or faults, so there is need to design a data-driven diagnostic system under class imbalance condition. This problem is explained in next Chapter.

Chapter 4

Data-driven Diagnostic Scheme Under Class Imbalance Condition

This chapter explains and proposes various approaches for diagnosing faults under class imbalance (CI) condition. Besides, CI is defined and it is mentioned that why class imbalance may happen. Moreover, the solutions to deal with this problem is addressed and related methods are explained.

4.1 Class Imbalanced Learning

The process of measuring and gathering raw data from different sources to analyze and detect anomalies, unusual trends and faults is crucial to enhance decision-making process. Many of the industrial processes, such as IMs, usually operate in the normal state. Thus, it is very common for the diagnostic systems to collect a large number of samples of the normal state in the batch of data, while only a very few faulty samples could be collected in practice. Various diagnostic schemes have been applied to industrial processes, however, detecting faults under the CI condition is a challenging task with growing attention from both academia and industry [73, 74].

CI can occur due to various reasons. Whether it is caused by error in data collection, limitations in time, cost, storage and privacy or because of the intrinsic properties of data. CI is an integral part of the real world problem and enormously common in practice [75, 16, 4]. An imbalanced dataset can be described as a set of samples, in which the proportion of the representative samples of one class is significantly larger than other class. The amount of this proportion brings up the definition of ‘imbalance

ratio', which is an important factor in selecting a proper classification technique. The imbalance ratio indicates the collected data are highly imbalance, moderate or low.

CI problem arises when the number of representative samples of one class far exceeds others. The major (minor) class in an imbalance dataset referred to a class with more (less) number of samples, while the rarer class is often the class of interest and should be detected with high accuracy. Standard classifiers usually require to accurately classify samples of the minor class as well as the major class. However, most of the classifiers usually require datasets with almost equal class distribution, while an imbalanced distribution results in misclassification of samples of the minor class and jeopardizes the classification performance. Besides, these classifiers usually consider the classification accuracy as a performance evaluation measure. However, the classification accuracy can be biased towards the overrepresented classes and cannot effectively reveal the prediction of the representative samples of the minor class [16].

Therefore, in CI conditions new techniques should be utilized to effectively classify faults or samples of the minor classes. These techniques are needed to construct classification models, that are able to accurately diagnose the most important and infrequent samples of faults [75]. Several techniques have been proposed to tackle CI problem [16, 75, 4, 76]. These techniques, depending on how they deal with the CI problem, can be divided into three main categories; algorithm-level approaches, data-level approaches and cost-sensitive methods [16]. The former category, addresses the problem by designing or modifying a classification algorithm in a way that it considers the importance of rare samples as well as frequent ones [4, 77]. The second category, often exploits different data processing techniques (i.e., sampling process) in order to re-balance the data distribution prior to feature extraction, feature selection and/or classification. These methods are generally more versatile because they can be integrated to any classification model [4, 17]. The third category includes cost-sensitive methods, which can belong to two previous categories. They either use different mis-

classification cost associated with each class sample (i.e., data-level approach) or alter the training procedure to take costs (i.e., algorithm-level approach) in order to bias the classifier toward the rare class [78, 79]. Various cost-sensitive learning techniques have been developed for CIL including cost-sensitive decision trees (CSDT) [80], cost-sensitive neural networks (CSNN) [81], cost-sensitive boosting methods, AdaCost, AdaC1, AdaC2 and AdaC3 [82, 83, 79]. Sometimes, cost-sensitive techniques are preferable to resampling techniques in some specific domains, however, definition of the misclassification cost is usually difficult and challenging [4, 16, 84, 18]. Cost-sensitive techniques are not then suitable for online monitoring applications and, thus, they are not considered in this work. Apart from these categories, ensemble-based techniques are successfully applied for CIL. Ensemble techniques (e.g., Bagging or Boosting) are usually merged with one of the above categories, in order to improve the accuracy. While cost-sensitive methods and algorithm-level approaches are usually problem-dependent, integration of data-level approaches and ensemble-based techniques are more versatile and have been effectively used for CIL in many works [75, 4, 76].

4.2 Data-Level Approaches

Data-level approaches, that are independent of the underlying classifier, are often a helpful solution for CIL. These methods try to reduce the effect of the skewed class distribution by adding samples of the minor class, e.g., random oversampling (ROS) or eliminating the available samples of the major class, e.g., random undersampling (RUS). ROS and RUS are simple methods to provide a balance class distribution; however, they are not ‘intelligent’, since they randomly choose their sampling subset, which may lead to over-fitting in the learning process or loss of important information. There exist some informed sampling techniques such as the synthetic minority oversampling technique (SMOTE) [17]. SMOTE has been widely used to

overcome the weakness of RUS and ROS. Another variation of informed sampling is the cluster-based oversampling (CBO) which aims to deal with the within-class imbalance problem along with the between-class imbalance problem [85].

Here, a novel oversampling technique has been developed which is based on imputation of missing values on samples of the minor class. This oversampling technique is based on expectation maximization (EM) [86], which is referred to as EM imputation-based over-sampling (EMI-OS). SMOTE and EMI-OS, as representative state-of-the-art sampling techniques are selected and applied for CIL with different imbalance ratio to re-balance the class distribution a priori.

4.2.1 SMOTE

SMOTE aims to generate some artificial samples to oversample the minor class instead of oversampling with replacement [17]. The number of created ‘synthetic’ samples depends on the required oversampling ratio. Given samples of the minor class, SMOTE generates some relevant synthetic samples as presented in Algorithm 1 [16, 17]. Each newly generated sample is along the line between a selected sample of the minor class x_i^{min} and its nearest neighbor x_j^{min} (see step 4), which provides a more general region for the minor class and can be effectively used in the CIL process. Thus, SMOTE avoids overfitting problem with ROS.

4.2.2 EMI-OS

EMI-OS is an EM imputation-based over-sampling technique, that aims to decrease imbalance ratio by generating new synthetic samples similar to samples of the minor class.

In order to gain a balanced dataset, it initially splits the dataset into subsets of samples of the minor class \mathcal{S}^{min} and major class \mathcal{S}^{maj} . It then iteratively induces random missing values on samples of the minor class, \mathcal{S}_t^{min} , where t stands for the

Algorithm 1: SMOTE [17]

INPUTS:

\mathcal{S}^{min} is a subset of samples of a minor class

α is the oversampling rate

k is the number of nearest neighbours

DEFINITIONS:

m^{min} is the # of samples of \mathcal{S}^{min}

for $i = 1, \dots, m^{min}$ **do**

for $j = 1, \dots, \alpha$ **do**

 1. Find a set k-nearest neighbors for x_i^{min}

$$\mathcal{N}_i = \{x_p \in \mathcal{S}^{min} \mid d_{ip} \leq d_{iq} \forall x_q \notin \mathcal{N}_i\}_{p=1}^k$$

 where d_{ip} (d_{iq}) stands for the Euclidean distance between x_i and x_p (x_i and x_q).

 2. Select a nearest neighbor of x_i^{min} , $x_j^{min} \in \mathcal{N}_i$

 3. Select a random number γ , where $0 \leq \gamma \leq 1$.

 4. Create a new synthetic sample

$$x_i^{os(j)} = x_i^{min} + (x_j^{min} - x_i^{min}) \times \gamma$$

end

end

5. Return $m^{min} \times \alpha$ newly generated samples

$$\mathcal{S}_{new}^{min} = \{x_i^{os(j)} \mid \forall i, j : i = 1, \dots, m^{min}, j = 1, \dots, \alpha\}$$

iteration number, and forms \mathcal{S}_{mis}^{min} . EMI-OS then estimates missing values of \mathcal{S}_{mis}^{min} by resorting to the expectation maximization imputation (EMI) technique [86] and returns \mathcal{S}_{est}^{min} . It then selects only imputed samples of \mathcal{S}_{est}^{min} to form \mathcal{S}_{imp}^{min} and, then, merges these newly imputed samples that are representative of the minor class with the previous subset of samples of the minor class \mathcal{S}_t^{min} . This creates a new and larger collection \mathcal{S}_{t+1}^{min} of representative samples of the the minor class for the subsequent iteration. This procedure iteratively continues to create and combine samples of the minor class, until the number of samples of the minor class m_t^{min} reaches to the number of samples of the major class, m^{maj} . The pseudo-code of EMI-OS is presented in Algorithm 2.

4.3 Ensemble-Based Approaches

These techniques have been recently used for CIL [4]. Ensemble-based methods can be categorized into two major groups of (a) cost-sensitive ensembles and (b) ensemble learning along with data processing. The former, cost-sensitive ensembles, includes various cost-sensitive boosting techniques such as AdaCost, AdaC1, AdaC2 and AdaC3 [82, 83, 79]. The latter, ensemble learning along with data processing, can be categorized into three subgroups: (b.1) Boosting-based ensembles techniques which embed data processing techniques into boosting algorithms such as SMOTEBoost [87], RUSBoost [75], and DataBoost-IM [88], (b.2) Bagging-based ensembles can be divided into OverBagging [89], UnderBagging [90], and UnderOverBagging [89] strategies. Bagging-based ensembles include various algorithms, e.g., SMOTEBag [89], AsymmetricBag [91] and Bagging Ensemble Variation (BEV) [92] algorithms, and (b.3) Hybrid ensembles which merge both boosting and bagging algorithms to form an ensemble of ensembles such as EasyEnsemble and BalancedCascade [93]. Hybrid ensembles slightly under-perform SMOTEBag and RUSBoost, while they have a more complex structure as an ensemble of ensembles [94]. Hybrid ensembles are com-

Algorithm 2: EMI-based oversampling (EMI-OS)

INPUT: \mathcal{S} is a class imbalance dataset

DEFINITIONS:

\mathcal{S}^{min} is the subset of samples of the minor class

\mathcal{S}^{maj} is the subset of samples of the major class

m^{min} is the # of samples of the minor class

m^{maj} is the # of samples of the major class

C is the # of classes in \mathcal{S}

REQUIRE: $\mathcal{S}_{os}^{min} = \emptyset$

1. Copy samples of major class into \mathcal{S}^{maj}

for $i = 1, \dots, C - 1$ **do**

2. Set a counter $t = 1$

3. Copy samples of i^{th} minor class into $\mathcal{S}_t^{min(i)}$

while $m_t^{min(i)} \prec m^{maj}$ **do**

4. Induce random missing on $\mathcal{S}_t^{min(i)}$ to form $\mathcal{S}_{mis}^{min(i)}$

5. Call EMI subroutine to estimate missing values

$$\left[\mathcal{S}_{est}^{min(i)} \right] = \text{EMI}(\mathcal{S}_{mis}^{min(i)})$$

where EMI stands for the EM imputation [86]

6. Use imputed samples of $\mathcal{S}_{est}^{min(i)}$ to form $\mathcal{S}_{imp}^{min(i)}$

7. Merge the minor subsets

$$\mathcal{S}_{t+1}^{min(i)} = \mathcal{S}_{imp}^{min(i)} \cup \mathcal{S}_t^{min(i)}$$

8. $t = t + 1$

end

9. Create the oversampled subset of minor classes

$$\mathcal{S}_{os}^{min} = \mathcal{S}_{t+1}^{min(i)} \cup \mathcal{S}_{os}^{min}$$

end

10. Return the balanced dataset $\mathcal{S}_b = \mathcal{S}_{os}^{min} \cup \mathcal{S}^{maj}$

putationally expensive and not suitable for online monitoring applications, and thus, are not considered in this work. A critical review on this topic has been performed and efficient ensemble techniques for CIL are compared in [4]. Among various ensemble-based approaches for CIL (see Figure 4.1), RUSBoost [75] and SMOTEbag [89] are the most commonly used and robust techniques [4]. Moreover, RUSBoost is a simpler method with less computational complexity, which is able to perform fast on both binary and multi-class imbalance problems. Therefore, RUSBoost and SMOTEbag are considered in this work as representatives of the ensemble-based approaches.

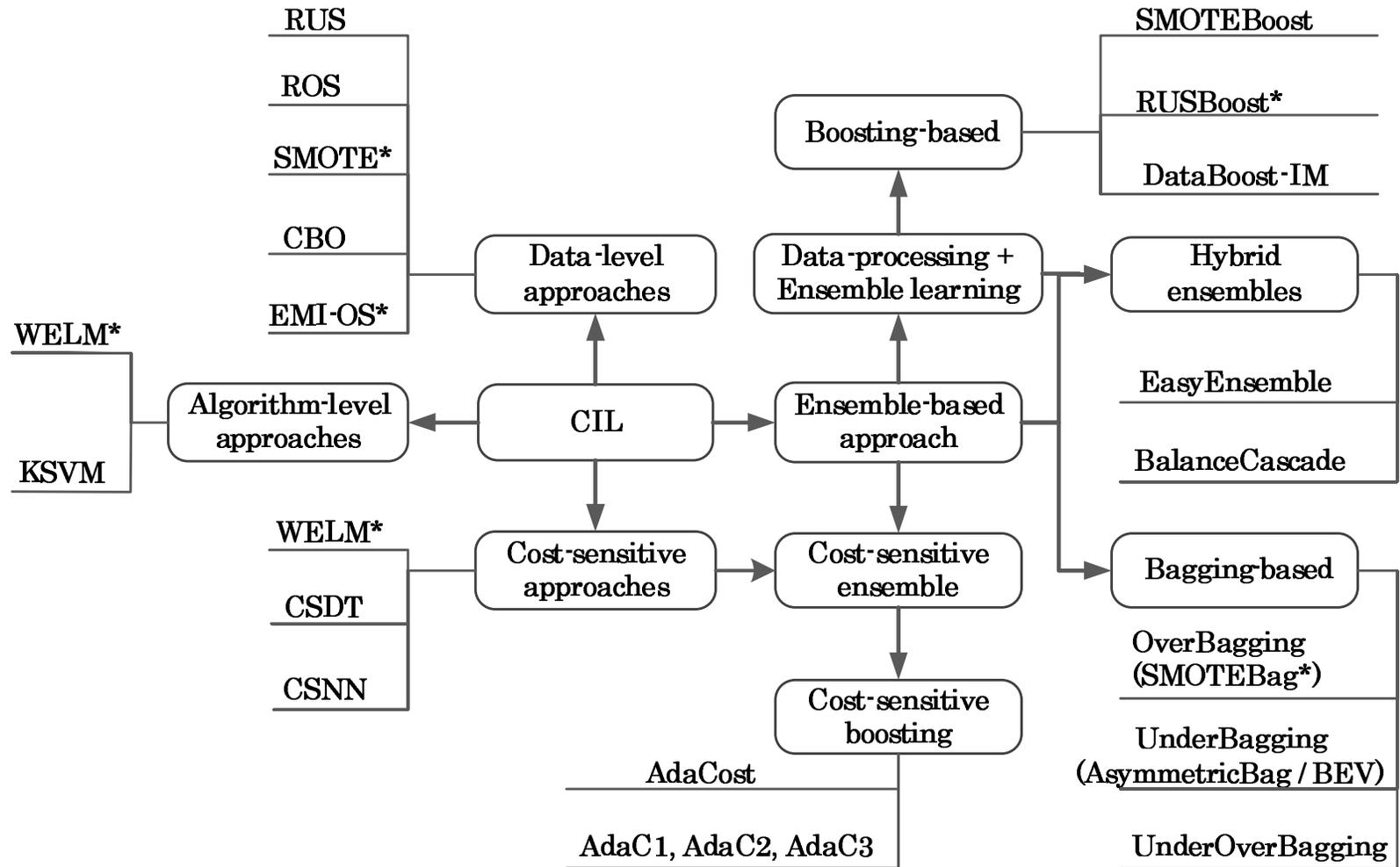


Figure 4.1 – major techniques to tackle class imbalance problem [4].

4.3.1 RUSBoost

RUSBoost algorithm is made of two important component; RUS and AdaBoost.M2 [75]. The former is a random undersampling method, which reduces number of samples of the major class and makes the RUSBoost learning process faster. The latter, boosting technique, aims to construct a composite classifier, which is often more accurate than each individual weak classifier. AdaBoost.M2 algorithm adjusts distribution weights of the samples, by assigning a higher weight to the misclassified samples including the representative samples of the minor class and also adjust voting weights of the individual base classifiers. This scheme also helps to avoid the loss of potentially useful samples, which can be occurred in RUS. The pseudo-code of RUSBoost is presented in Algorithm 3 [75].

4.3.2 SMOTEBagging

SMOTEBagging, so-called SMOTEBag (Algorithm 4) [89], is based on two components; SMOTE and Bagging algorithms. The former creates new synthetic samples by interpolating samples of the minor class to construct various balance subsets \mathcal{S}_t from the CI dataset \mathcal{S} and, then, iteratively makes use of the more balanced and diverse subsets \mathcal{S}_t to train T weak classifiers h_t and constructs a bagged ensemble.

SMOTEBag enhances the diversity among the base classifiers by changing the ratio of bootstrap replicates and generating various synthetic samples by SMOTE, over the bagging iterations. To construct these subsets, SMOTEBag resamples from the major class with replacement at rate 100%. It does not oversample each minor class separately. It uses a percentage value $\beta\%$ (range from 10 to 100) to control the number of newly generated synthetic samples of the minor classes. It initially resamples some original samples of the minor classes and, then, uses SMOTE to create new synthetic samples of the minor classes.

Algorithm 3: RUSBoost [75]

INPUTS:

$\mathcal{S} = \{(x_i, y_i) | x_i \in X, y_i \in Y\}_{i=1}^m$ is a CI dataset

h is the weak learning algorithm, i.e., Base-Classifier

T is the maximum # of iterations or Base-Classifiers

m' is the percentage of total samples of the minor class

DEFINITIONS:

m is the # of samples

1. Initialize the weight of each sample $w_i = \frac{1}{m}$

for $t = 1, \dots, T$ **do**

2. Do RUS on samples of the major class until m' condition satisfies and Form \mathcal{S}'_t with distribution w'_t

3. Construct a weak h_t based on \mathcal{S}'_t along with w'_t

4. Compute the error, ϵ_t w.r.t. S and w_t

$$\epsilon_t = \sum_{i: h_t(x_i) \neq y_i} w_t(i) (1 - h_t(x_i, y_i) + h_t(x_i, y))$$

5. Compute parameter $\gamma_t = \frac{\epsilon_t}{1 - \epsilon_t}$

6. Update the weight distribution w_t

$$w_{t+1}(i) = w_t(i) \gamma_t^{\frac{1}{2}(1 + h_t(x_i, y_i) - h_t(x_i, y \neq y_i))}$$

7. Normalize $w_{t+1}(i) = w_{t+1}(i) / \sum_i w_{t+1}(i)$

end

8. Return the final hypothesis:

$$H(x) = \underset{y \in Y}{\operatorname{argmax}} \sum_{t=1}^T h_t(x, y) \log \frac{1}{\gamma_t}$$

Algorithm 4: SMOTEBag [89]

INPUTS: \mathcal{S} is a CI dataset; T is the number of Base-Classifier; k is the number of nearest neighbors

DEFINITIONS: C is the number of classes in \mathcal{S} ; \mathcal{S}^{maj} is the subset of samples of the major class; $\mathcal{S}^{min}(i)$ is the subset of samples of the i^{th} minor class, $i = 1, \dots, C - 1$; \mathcal{S}_r^{maj} and $\mathcal{S}_r^{min}(i)$ are the resampled subsets of samples of the major and i^{th} minor class; $\mathcal{SWR}(A, B)$ is the function of sampling with replacement to draw B samples from the given set A .

for $t = 1, \dots, T$ **do**

1. $\beta = \frac{t}{T}$

2. Resample the major class with replacement at 100% as follows:

$$\mathcal{S}_r^{maj} = \mathcal{SWR}(\mathcal{S}^{maj}, m^{maj})$$

for $i = 1, \dots, C - 1$ **do**

3. Resample from original samples of the i^{th} minor class with replacement at the percentage of $\theta = \frac{m^{maj}}{m_i^{min}} \cdot \beta$

$$\mathcal{S}^{min}(i) = \mathcal{SWR}(\mathcal{S}^{min}(i), \theta \cdot m_i^{min})$$

4. Set $\alpha = \frac{m^{maj}}{m_i^{min}} \cdot (1 - \beta)$

5. Create new synthetic samples by means of SMOTE as follows:

$$\mathcal{S}_{new}^{min}(i) = \mathcal{SMOTE}(\mathcal{S}^{min}(i), \alpha, k)$$

end

6. Construct \mathcal{S}_t subset by means of samples of all classes as follows:

$$\mathcal{S}_t = \{\mathcal{S}_r^{maj} \cup \mathcal{S}_r^{min}(i) \cup \mathcal{S}_{new}^{min}(i) | \forall i\}_{i=1}^{C-1}$$

7. Train a base-classifier $h_t \rightarrow Y$ using \mathcal{S}_t

end

8. Find the composite hypothesis: $H(x) = \underset{y \in Y}{argmax} \sum_{t=1}^T h_t(x, y)$

4.4 Algorithm-Level Approaches

There exists another group of machine learning techniques, so-called algorithm-level approaches, to alleviate class imbalance problem [4, 77]. These techniques aim to adapt classification algorithms in order to bias the learning toward the minor class [78, 79]. Various efforts have been made to modify kernel mechanisms in the learning process such as Kernel-based support vector machines (KSVMs) [95]. Here, Weighted Extreme Learning Machine (WELM) [96] is used as a state-of-the-art algorithm-level approach, which also makes use of cost information.

4.4.1 Weighted Extreme Learning Machine

Extreme Learning Machine (ELM) is a generalized single hidden layer feedforward network (SLFN) with flexible processing nodes [97, 98]. ELM has been widely used for many machine learning applications due to its fast learning speed and good generalization performance. Weighted ELM (WELM) has been lately proposed to tackle the CI problem [96]. WELM can also belong to cost-sensitive approaches, since it uses a misclassification cost in a weighting scheme.

Given a set of multi-class samples $\{x_i, y_i\}_{i=1}^m$, where y_i is the vector of length C (i.e., the number of classes), WELM makes use of an $m \times m$ diagonal weight matrix W associated with each sample x_i and increases the corresponding weight of sample of the minor class(es), in order to enhance the impact of the minor class(es) and decreases the relative impact of the major class. WELM then reformulates the ELM optimization problem by means of the diagonal weight matrix W (the reader can refer to [96] for the detailed explanation), in order to maximize the marginal distance $2/||\lambda||$ and to minimize the weighted cumulative error ξ_i w.r.t. each sample:

$$\begin{cases} \text{Minimize:} & L_{ELM} = \frac{1}{2}||\lambda||^2 + \frac{\eta W}{2} \sum_{i=1}^m ||\xi_i||^2 \\ \text{Subject to:} & h(x_i)\lambda = y_i^T + \xi_i^T, \quad i = 1, \dots, m \end{cases} \quad (4.1)$$

where $\xi_i = \|H\lambda - Y\|^2$, H stands for the hidden layer output matrix, λ is the output weight, Y is the target vector and η is the regularization parameter to adjust the optimization terms. This optimization problem can be solved by means of the Karush-Kuhn-Tucker (KKT) theorem [99] which yields to:

$$\hat{\lambda} = H^\dagger Y = \begin{cases} H^T(\frac{1}{\eta} + W H H^T)^{-1} W Y & \text{if } m < l \\ (\frac{1}{\eta} + H^T W H)^{-1} W H^T Y & \text{if } m \succeq l \end{cases} \quad (4.2)$$

where l is the number of hidden nodes and H^\dagger stands for the Moore-Penrose generalized inverse of the matrix H [97, 98]. The output function of WELM then can be obtained as follows:

$$f(x) = \begin{cases} h(x) H^T(\frac{1}{\eta} + W H H^T)^{-1} W Y & \text{if } m < l \\ h(x) (\frac{1}{\eta} + H^T W H)^{-1} W H^T Y & \text{if } m \succeq l \end{cases} \quad (4.3)$$

where $f(x) = [f_1(x), \dots, f_C(x)]$ stands for the output function vector. The predicted class label of x then can be obtained by $\underset{i}{\operatorname{argmax}} f_i(x)$, $i \in [1, \dots, C]$.

4.5 Summary

In this chapter, class imbalance (CI) problem is explained. CI is a common problem in many real world case studies. CI happens since the number of samples from different classes (i.e., normal and faults) are not necessarily as equal as each other. In this work, the number of samples represent normal operation condition of IMs is larger than the number of samples that represents each class of fault. In this chapter, the solutions and the state-of-the-art techniques to deal with CI problem are addressed and a novel sampling technique is also introduced. Some parts of the next chapter describe the experimental results of the diagnostic system under class imbalance condition. The proposed scheme includes all the state-of-the-art preprocessing techniques explained in Chapter 3 and the CIL methods described in this Chapter.

Chapter 5

Experimental Results

This chapter includes three different experiments for diagnosing bearing faults in induction motors. Firstly the performance measures used for evaluating the accuracy of the diagnostic system in these experiments are explained. Then, the experimental results related to each of the experiments are explained in separate sections. First experiment relates to results obtained from the diagnostic system which is a combination of the wavelet packet transform as a FE method and liner discriminant analysis as a DR method. Six different intelligent fault classifiers are used in this experiment. In the second experiment, combination of empirical mode decomposition with five different dimensionality reduction techniques are evaluated and the experimental results are presented. Finally, in the third experiment combination of various state-of-the-art signal analysis methods with different feature reduction techniques under class imbalance condition is studied. All the FE and FS methods used in this experiment were explained in Chapter 3. Moreover, the CIL methods described in Chapter 4, are applied on CWRU bearing data and the achieved results are presented in this chapter.

5.1 Performance Measures

Performance evaluation is a challenging task under class imbalanced conditions. While accuracy is a performance measure in balanced conditions, it cannot be an appropriate measurement for imbalanced datasets as it has bias over the largest class [100]. Although, in two-class imbalanced problem, it is easy to provide some proper mea-

measurements to evaluate the accuracy of the classifiers, in multi-class imbalance conditions, it is difficult to carefully evaluate the classification performance. Thus, it is necessary to apply different appropriate metrics to provide an accurate and reliable assessment. These performance measures are weighted average of F-measure, MCC and ROC which are derived from the confusion matrix for imbalanced conditions. These performance metrics are defined below [100, 101]:

$$\text{Weighted } F - \text{measure} = \sum_{i=1}^c (w_i \cdot F\text{-measure}_i) \quad (5.1)$$

$$\text{Weighted } MCC = \sum_{i=1}^c (w_i \cdot MCC_i) \quad (5.2)$$

$$\text{Weighted } ROC = \sum_{i=1}^c (w_i \cdot ROC_i) \quad (5.3)$$

$$MAvG = \left(\prod_{i=1}^c Acc_i \right)^{\frac{1}{c}} \quad (5.4)$$

where c is the number of class, $F\text{-measure}_i$, MCC_i and ROC_i stand for the respective performance measures for a particular class i . Besides, w_i is the weight of the i^{th} class and Acc_i stands for the accuracy on class i , which can be calculated by [100, 101]:

$$w_i = \frac{n_i}{\sum_{i=1}^c n_i} \quad (5.5)$$

$$Acc_i = \frac{\text{correctly classified samples of class } i}{\text{total number of samples in class } i} \quad (5.6)$$

where n_i is the size of i^{th} class.

For a particular class i , the F-measure is formulated as follows:

$$F - \text{measure}_i = \frac{(2 \times \text{precision}_i \times \text{recall}_i)}{\text{precision}_i + \text{recall}_i} \quad (5.7)$$

in which:

$$precision_i = \frac{TP_i}{(TP_i + FP_i)} \quad (5.8)$$

$$recall_i = \frac{TP_i}{(TP_i + FN_i)} \quad (5.9)$$

where FP_i is the number of false positives, FN_i is the number of false negatives and TP_i and TN_i stand for true positives and true negatives for a specific class i , respectively.

Consider a confusion matrix Γ as depicted in Table 5.1 for a c -class problem where $\gamma_{p,q}$, ($p, q \in [1, c]$), referred to the value in row p and column q of the matrix Γ . Hence, TP_i , FP_i , FN_i and TN_i can be calculated as follows:

Table 5.1 – Confusion matrix for the multi-class problem

		Predicted			
		class 1	class 2	...	class c
Actual	class 1	$\gamma_{1,1}$	$\gamma_{1,2}$	$\gamma_{1,q}$	$\gamma_{1,c}$
	class 2	$\gamma_{2,1}$	$\gamma_{2,2}$	$\gamma_{2,q}$	$\gamma_{2,c}$
	...	$\gamma_{p,1}$	$\gamma_{p,2}$	$\gamma_{p,q}$	$\gamma_{p,c}$
	class c	$\gamma_{c,1}$	$\gamma_{c,2}$	$\gamma_{c,q}$	$\gamma_{c,c}$

$$TP_i = \gamma_{i,i} \quad (5.10)$$

$$FP_i = \sum_{p=1}^c \gamma_{p,i} - \gamma_{i,i} \quad (5.11)$$

$$FN_i = \sum_{q=1}^c \gamma_{i,q} - \gamma_{i,i} \quad (5.12)$$

$$TN_i = \sum_{p,q=1}^c \gamma_{p,q} - FN_i - FP_i - TP_i \quad (5.13)$$

The MCC per class, MCC_i , can be computed then as follows:

$$MCC_i = \frac{(TP_i \times TN_i - FP_i \times FN_i)}{\sqrt{(TP_i + FP_i)(TP_i + FN_i)(TN_i + FP_i)(TN_i + FN_i)}} \quad (5.14)$$

Moreover, area under ROC curve for the i^{th} class depends on true positives and false positives. ROC_i is a function of a varying threshold. The Mann-Whitney statistic is considered to calculate the ROC area for a particular class i (the reader can refer to [101, 102] for a more detailed explanation).

These performance measures that are obtained by means of different classifiers are compared together to find out whether the fault classifiers could properly classify the imbalanced data. For the sake of brevity, hereafter weighted average of F-measure, weighted average of MCC and weighted average of ROC are written as ‘F-measure’, ‘MCC’ and ‘ROC’, respectively.

5.2 First Experiment - Combination of WPT and LDA

This experiment aims to extract and select a proper set of features for diagnosing bearing defects in induction motors. Firstly, the vibration signals are analyzed by the wavelet packet transform, which is explained in Chapter 3, to extract informative time-frequency domain features.

Vibration signals for normal status, ball defect, inner race and outer race defects are depicted in Figure 5.1. The fault diameters of 0.021 and 0.028 *inches* are considered for ball and inner race fault state. Data for outer race defect includes 0.021 *inch* fault diameter that is gathered in three different positions (i.e., centered, orthogonal and opposite) relative to load zone. Data for 0.028 *inch* diameter is not available.

In this experiment, vibration signals are decomposed to eight levels using discrete Meyer wavelet with Shannon entropy. An eight-level decomposition generates a total of $2^8 = 256$ packs. Then, each pack is segmented and the entropy of each segment

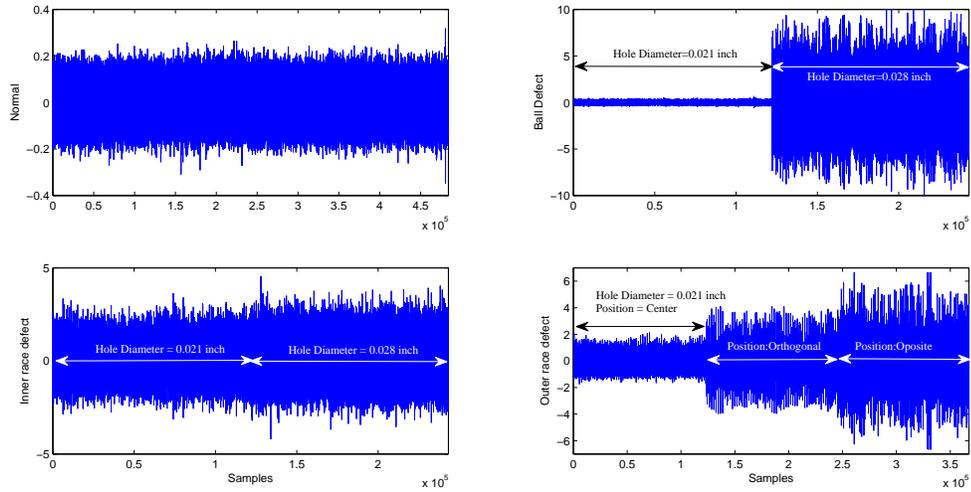


Figure 5.1 – Raw vibration signals of different classes: normal, ball defect, inner race defect, outer race defect. The faulty signals include different diameters of defects simulating incipient faults.

is calculated to create more informative feature space. For feature selection, three scenarios are considered as follows:

- 1) In the first scenario, all the features are fed to the classification module.
- 2) In the second scenario, the seven most informative packs based on the motor bearing vibration frequency, explained in Chapter 1, are selected. In this case, user knowledge about the parameters of the system is needed to calculate the fundamental frequencies of the faults.
- 3) In the third scenario, linear discriminant analysis (LDA) is applied on the obtained large scale data to reduce the number of features in an efficient way which is easier for the classifiers to discriminate different classes.

The classification results of the first scenario for each classifier are reported in Table 5.2 and illustrated in Figure 5.2. The bold entries in the Table 5.2 stand for the highest value of the respective performance measure in this scenario. The attained results show that RF outperforms other competitors according to all performance measures. In the first scenario, kNN, MLP, DT, NB, and SVM take the subsequent ranks, respectively, except MLP which has the higher rank according to the ROC

Table 5.2 – Comparison of the classification evaluation factors obtained by each classifier in first scenario.

Factors	First Scenario					
	MLP	NB	SVM	kNN	RF	DT
F-Measure	0.952	0.872	0.801	0.957	0.971	0.942
MCC	0.936	0.847	0.771	0.942	0.962	0.924
MAvG	0.951	0.820	0.746	0.952	0.964	0.937
ROC Area	0.998	0.962	0.874	0.968	0.999	0.968

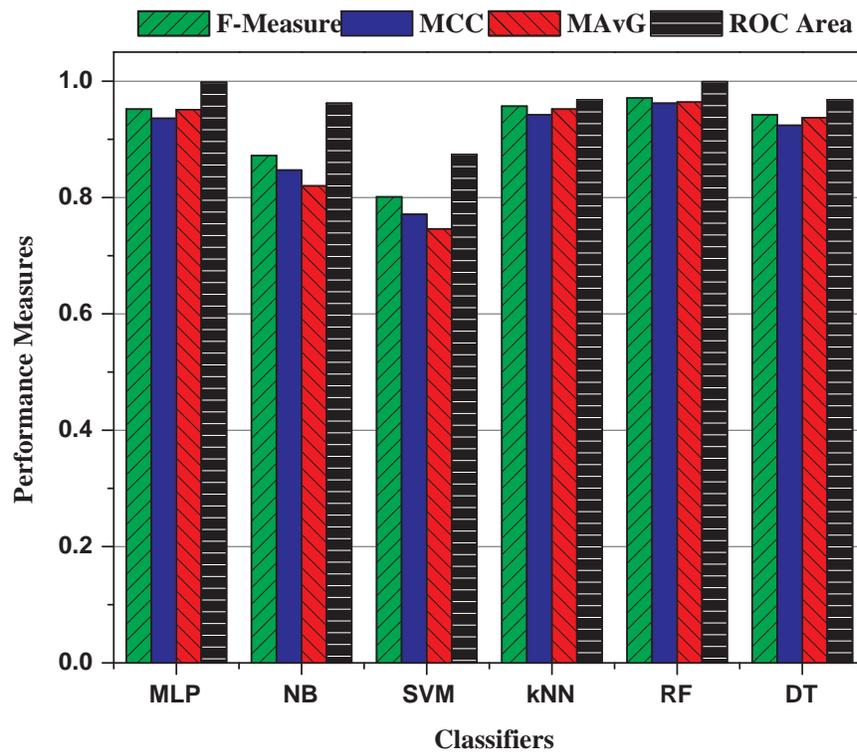


Figure 5.2 – Performance evaluations (i.e., F-measure, MCC, MAvG and ROC area) obtained by each classifier on the first scenario.

area.

In the second scenario, random forest still outperforms other competitors (see the bold entries in Table 5.3 for the second scenario). The performance measures of the second scenario are also illustrated in Figure 5.3. In the second scenario, DT, kNN,

Table 5.3 – Comparison of the classification evaluation factors obtained by each classifier in second scenario.

Factors	Second Scenario					
	MLP	NB	SVM	kNN	RF	DT
F-Measure	0.773	0.660	0.747	0.792	0.874	0.844
MCC	0.719	0.597	0.690	0.727	0.835	0.796
MAvG	0.700	0.510	0.659	0.760	0.855	0.815
ROC Area	0.890	0.900	0.835	0.856	0.978	0.919

MLP, SVM, and NB take the next ranks, respectively. MLP still has the second rank according to the ROC area measure. However, the figure shows that the performances of the trained classifiers in the second scenario are lower than those trained in the previous scenario. Performance estimation of the SVM classifier can be affected by cost and gamma parameters values. These parameters are adjusted by means of the grid search strategy [103].

Table 5.4 – Comparison of the classification evaluation factors obtained by each classifier in third scenario.

Factors	Third Scenario					
	MLP	NB	SVM	kNN	RF	DT
F-Measure	1	1	1	1	1	1
MCC	1	1	1	1	1	1
MAvG	1	1	1	1	1	1
ROC Area	1	1	1	1	1	1

In the last scenario, as displayed in Figure 5.4, there is a significant improvement on classification performances of all classifiers as they reached to a maximum performance equal to one. The detailed classification evaluation factors obtained by each classifier for third scenario is provided in Table 5.4.

It is shown that using LDA along with the wavelet packet transform leads to better classification performances. LDA carries out dimensionality reduction of the extracted features in a way that all the selected classifiers are able to perfectly separate

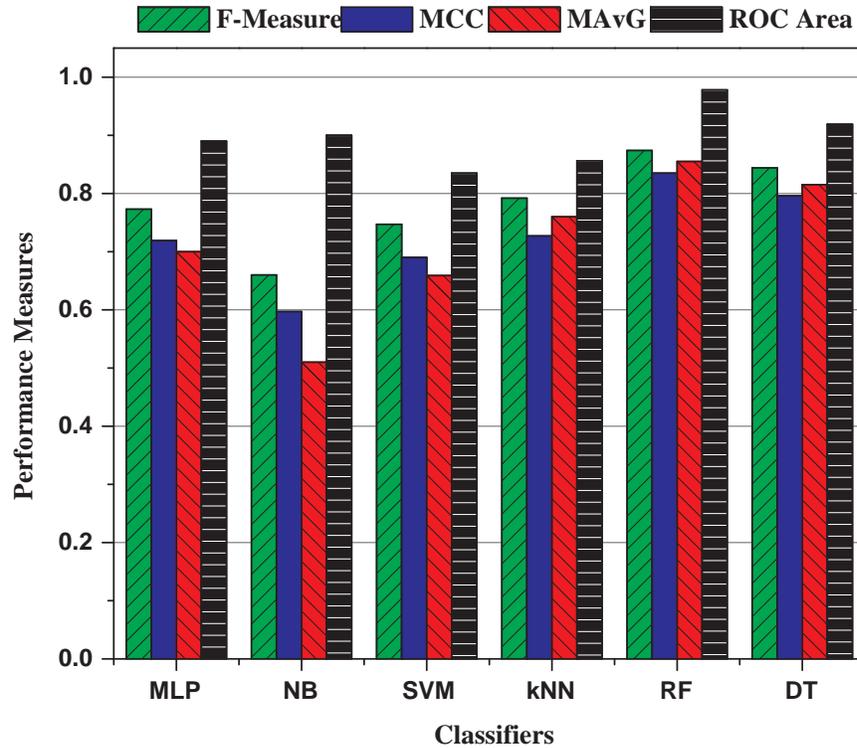


Figure 5.3 – Performance evaluations (i.e., F-measure, MCC, MAvG and ROC area) obtained by each classifier on the second scenario.

normal and fault classes. The performance of the MLP, SVM, and NB classifiers are significantly improved by means of the provided features in the third scenario. The computational time is also significantly reduced for the MLP classifier since it receives a much smaller feature set as input. In addition, RF can classify data very well compared to other classifiers in all three scenarios.

5.3 Second Experiment - Combination of EMD and DR Methods

This experiment focuses on the CWRU bearing datasets, which include vibration signals of a 2hp (horse power) motor (1750 rpm). These vibration signals are sampled at

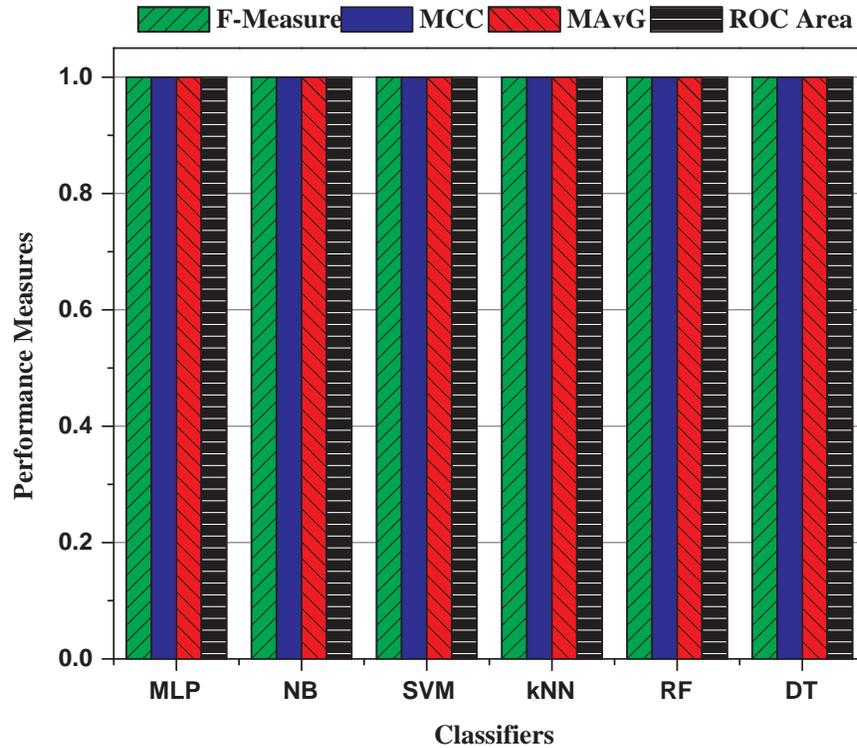


Figure 5.4 – Performance evaluations (i.e., F-measure, MCC, MAvg and ROC area) obtained by each classifier on the third scenario.

12kHz frequency representing various conditions including the normal, rolling element fault, inner race fault and outer race defects.

The fault diagnosis unit contains two major modules for feature extraction and dimensionality reduction. The former contains the empirical mode decomposition (EMD) which is one of the most attractive techniques to process the nonlinear and non-stationary vibration signals explained in chapter 3. The latter contains various state-of-the-art DR techniques named as PCA, LLC, LDA, NCA and MCML which are also described in Chapter 3. The prepared features by means of these DR techniques are then fed to the decision making module, which contains different classifiers, i.e., Multi-Layer Perceptron (MLP), Naive Bayes (NB), k Nearest Neighbor (kNN), Decision Tree (DT), and Random Forest (RF). In order to evaluate the effect of DR techniques on performance of classifiers, two different scenarios are considered

as follows:

1) In the first scenario, 16 IMFs (i.e., features) extracted by means of EMD from vibration signals are used in the fault classification module.

2) In the second scenario, five different DR methods are applied on the extracted features to reduce the number of features in an efficient way which is easier for the classifiers to discriminate different classes. Then these features of the new feature space fed as inputs to the fault classifiers.

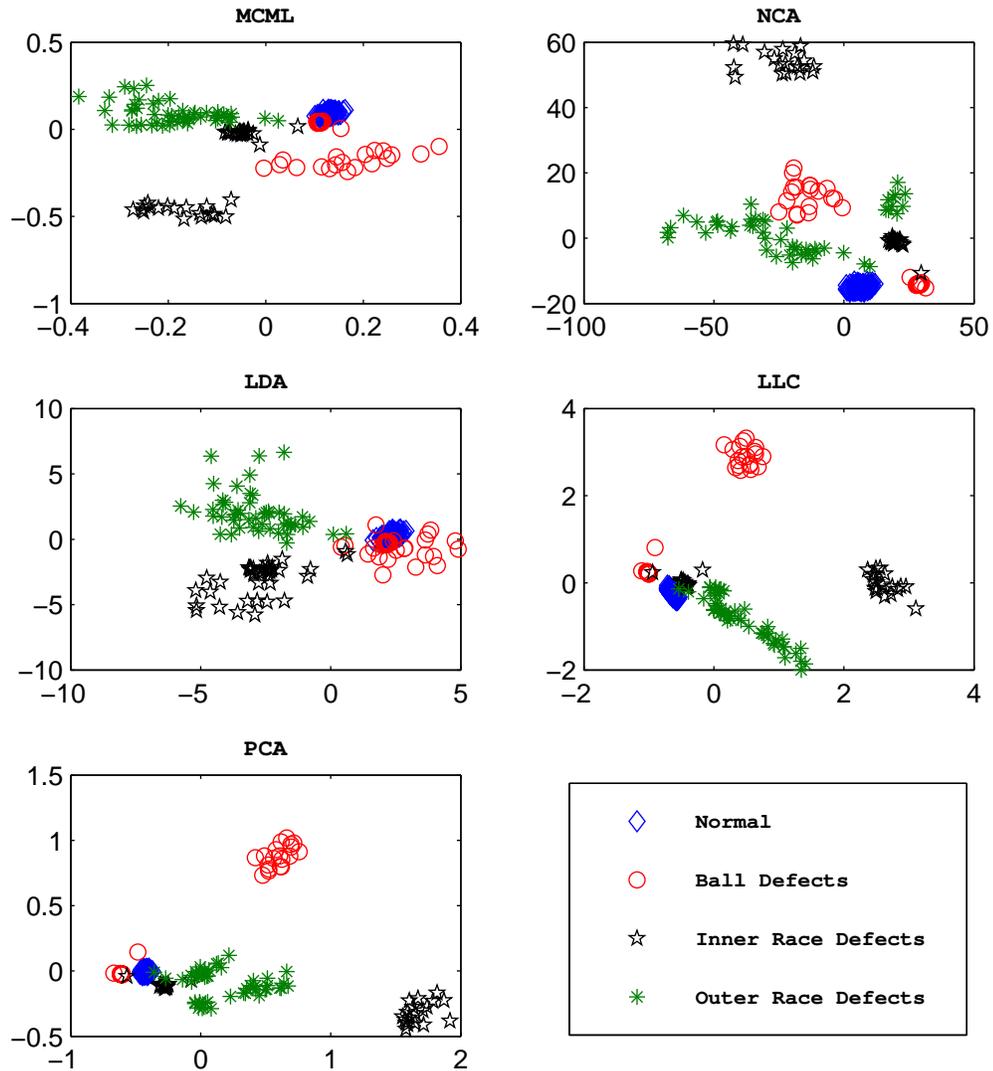


Figure 5.5 – Two dimensional features achieved by different DR techniques.

Figure 5.5 illustrates the achieved two-dimensional features, by means of each feature reduction method along with the class labels, i.e., normal, ball, inner race, and outer race. The attained features by means of each DR technique are then fed to the fault classifiers in the decision making module. The performances of fault classifiers are carefully evaluated through a 10-fold cross validation scheme by means of various suitable performance measures. These performance measures are weighted average of F-measure, weighted average of Matthews correlation coefficient (MCC), weighted average of receiver operating characteristic (ROC) area and Macro average geometric (MAvG) which are already defined in this Chapter.

Figure 5.6 illustrates the attained performance measures by means of each classifier trained by the features obtained through different DR techniques. For instance, the label ‘MCML-RF’ in the figure stands for the performance measure achieved by RF classifier which was trained by a feature set obtained through MCML DR technique. The first panel in the Figure 5.6 indicates that kNN and RF outperform other classifiers and MCML and LDA outperform other DR techniques in terms of F-measure. Other panels in the Figure 5.6 show more or less the same rank, except for the third panel which shows that the first scenario ‘No-DR’ (i.e., the right most five values) achieves a better rank in terms of ROC.

In order to compare these DR techniques, all the performance measures obtained by all classifiers are presented through a boxplot. Figure 5.7 presents the distribution of all performance measures obtained by all classifiers for each DR technique. This figure shows that MCML outperforms other DR techniques in providing useful and informative features for the fault classifiers. MCML is the most stable DR approach with the least variation (i.e., smallest box) and the highest mean value. LDA is positioned in the second rank; it is comparable with MCML except for the weak performances that are achieved along with DT and NB. NCA takes the third rank as a supervised DR technique, however, it achieves a few low performances along with NB which causes a larger variation compared to the NO-DR method (see also

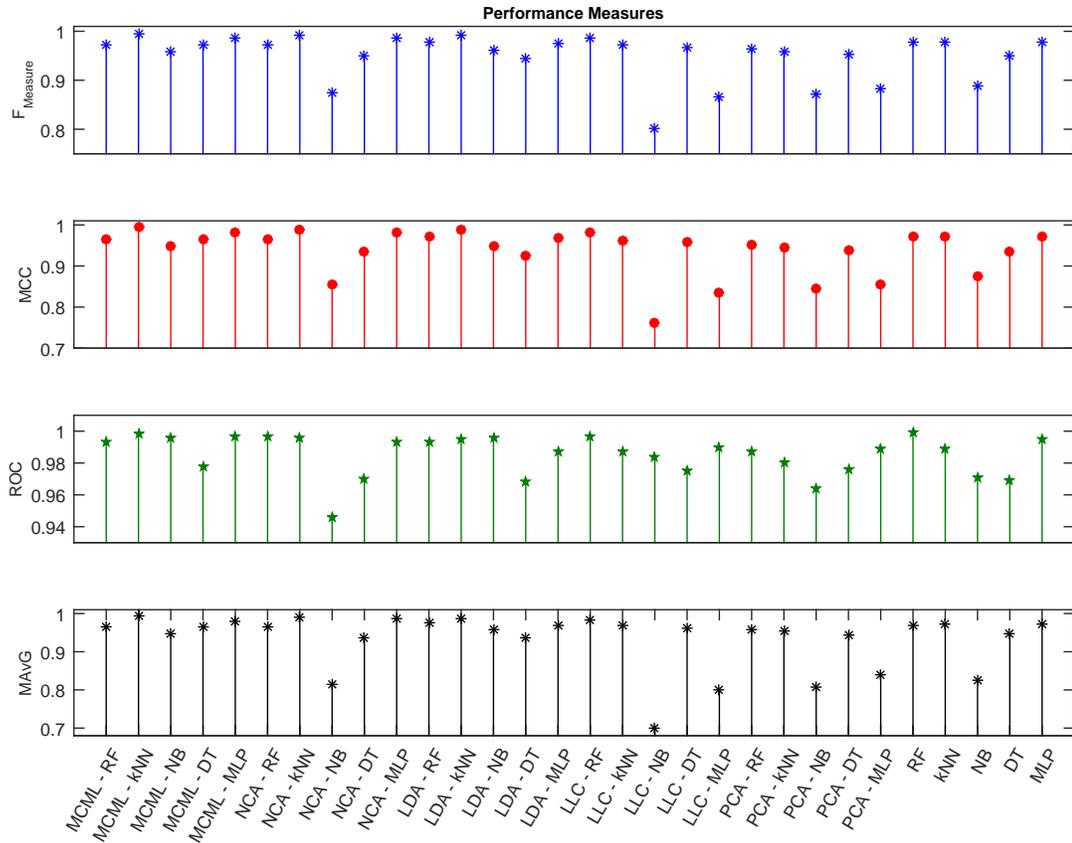


Figure 5.6 – The performance measures obtained by means of each classifier trained by means of the features of different DR techniques. F-measure, MCC, ROC area and MAvg are presented in different panels from top to bottom.

Figure 5.6 for more details). NO-DR, PCA and LLC take the subsequent ranks, respectively. This can be explained through the unsupervised nature of the PCA and LLC methods. NO-DR achieves the fourth rank compared to PCA and LLC with the cost of higher computational time. The lower performances of PCA and LLC are due to their unsupervised nature, i.e., they cannot explain some percent of variability among the features since they cannot see the class information. However, one can increase the variability explained by the PCA and LLC models by increasing the number of features with cost of increasing the computational time. LLC has also the most unstable results due to the weak performance achieved along with the NB classifier. This study also shows that using MCML along with the kNN and RF can

lead to the best performances.

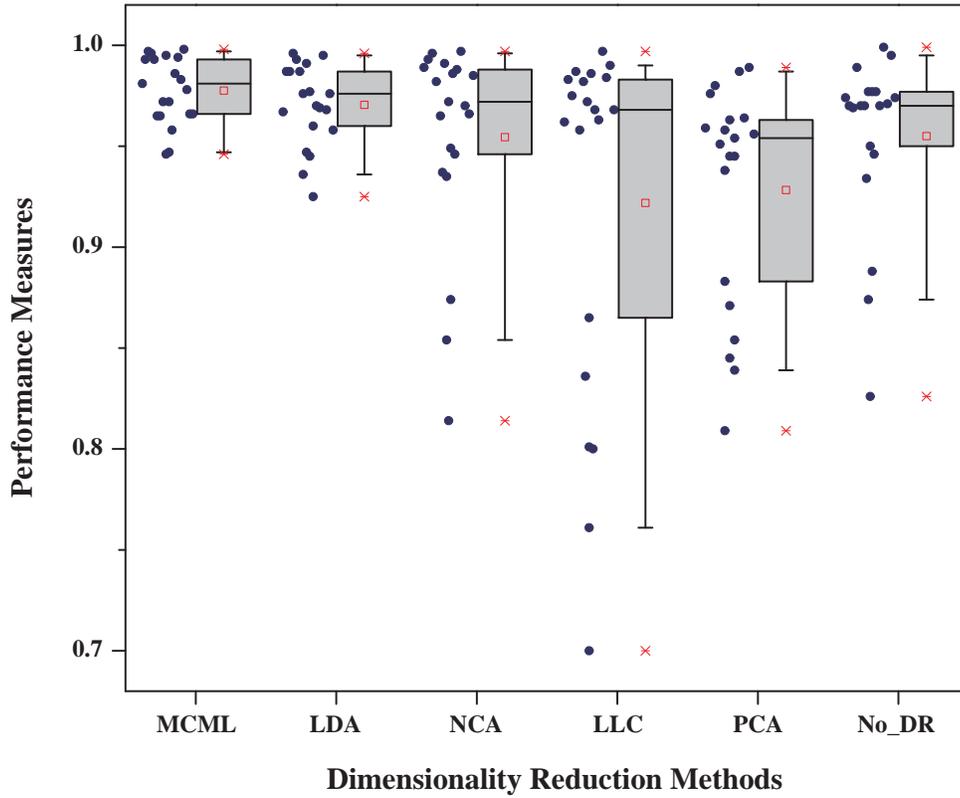


Figure 5.7 – Boxplot illustrates the distribution of the performance measures (i.e., F-measure, MCC, MAvG and ROC area) attained by all classifiers for different DR techniques, where solid circles represent the distribution of the performance measures, the red dashes stand for maximum and minimum values, the solid squares stand for the average value for each DR method, and the red crosses stand for 1 and 99 percentiles of the performance values.

5.4 Third Experiment - Diagnostic System under Class Imbalance Condition

The proposed diagnostic scheme of this experiment is presented in Figure 5.8. Figure 5.8 illustrates various state-of-the-art feature extraction and feature reduction

techniques (explained in Chapter3) along with CIL techniques. In addition, the fault classification module contains a novel and several state-of-the-art CIL techniques as explained in Chapter 4. These techniques are considered for diagnosing bearing defects under CI condition.

In order to better analyze the performance of diagnostic scheme under different CI condition, three different datasets D_1 , D_2 , D_3 with three different level of imbalance, low, moderate and high are created from the CWRU bearing vibration records. These datasets contain samples of normal state (i.e., as majority) and samples of different defects (i.e., as minority). Two different defect width, i.e. 0.007 inch and 0.021 inch, motor load of 2 hp, shaft speed of 1750 rpm with sampling rate of 12 KHz are considered in this study. The number of selected samples for class of normal and defects, imbalance ratios and defect width are specified in Table 5.5. In this table, the imbalance ratio of (20:1), (100:1) and (200:1) means that the number of samples in the class of major are 20, 100 and 200 times greater than the number of samples in the minor class and categorized as low, moderate and high class imbalance (i.e., LCI, MCI, HCI) datasets, respectively.

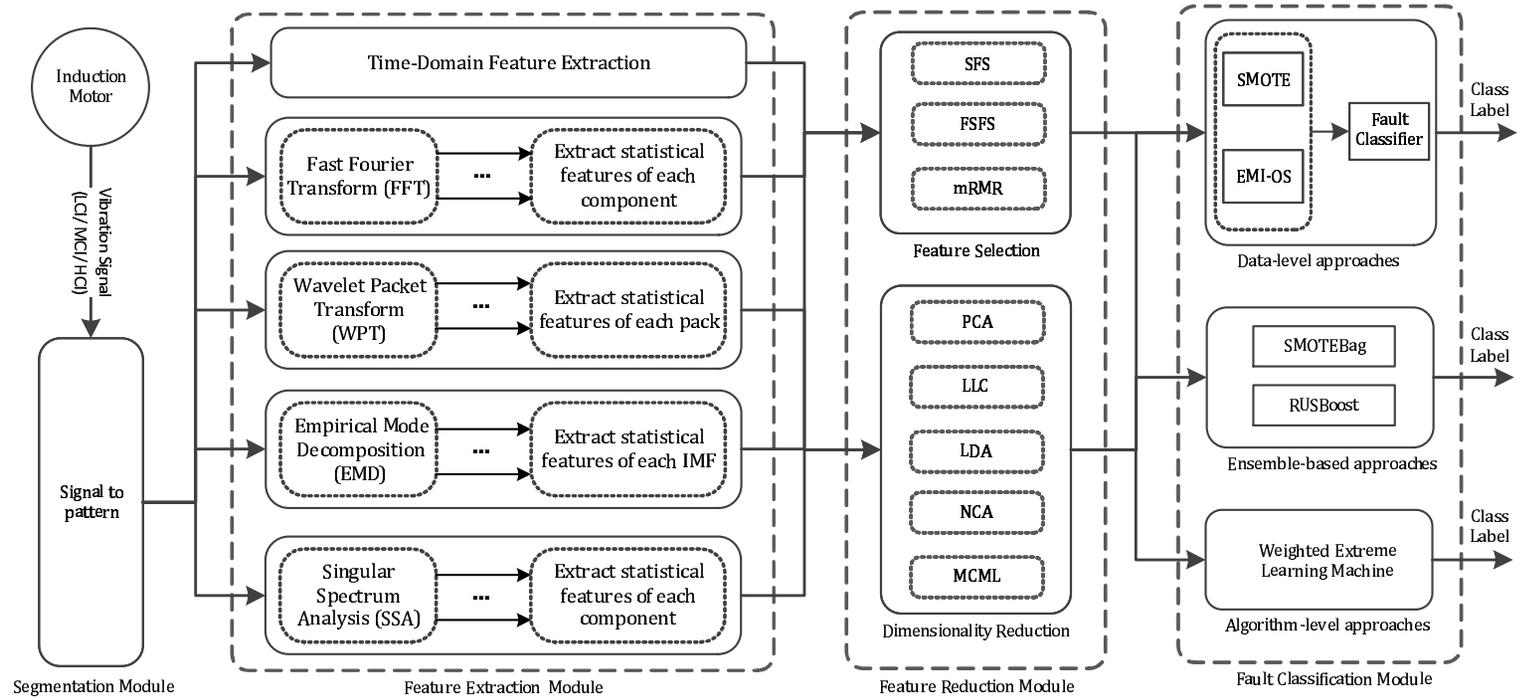


Figure 5.8 – Block diagram of the proposed diagnostic system under class imbalance condition

Table 5.5 – Number of samples in each of the created datasets

Class	Defect Width (<i>in.</i>)	Length of Signal			No. of Segmented Samples		
		D_1	D_2	D_3	D_1	D_2	D_3
Normal	-	819,200	819,200	819,200	800	800	800
Ball (DE)	0.007 and 0.021	40,960	8,192	4,096	40	8	4
Ball (FE)	0.007 and 0.021	40,960	8,192	4,096	40	8	4
Inner Race (DE)	0.007 and 0.021	40,960	8,192	4,096	40	8	4
Inner Race (FE)	0.007 and 0.021	40,960	8,192	4,096	40	8	4
Outer Race (DE)	0.007 and 0.021	40,960	8,192	4,096	40	8	4
Outer Race (FE)	0.007 and 0.021	40,960	8,192	4,096	40	8	4
Total		1,064,960	869,352	843,776	1040	848	824
Imbalance Ratio		20:1	100:1	200:1	20:1	100:1	200:1
		(LCI)	(MCI)	(HCI)	(LCI)	(MCI)	(HCI)

The fault classification module receives the CI sets (i.e., LCI, MCI and HCI) of samples as input and learns the relation between the features and defects to diagnose bearing defects under the CI conditions.

The CIL techniques, except for WELM, use the decision tree C4.5 algorithm as a base classifier, since it has been extensively used in CI domains [104, 105]; and perform K-fold cross validation on the input sets to avoid bias in the estimated performance caused by a single split of the available samples into training and test subsets.

Diagnostic results are compared in terms of weighted averages of ROC, MCC and F-measure. These performance measures are used to evaluate the performance of each module and to compare the impacts of each feature extraction and reduction technique in the diagnostic performance. Besides, this experiment also aims to evaluate and compare the diagnostic performance of CIL techniques w.r.t. feature extraction and reduction modules at different CI ratios. Figures 5.9, 5.10, 5.11, and 5.12 compares the state-of-the-art techniques in each module for diagnosing bearing defects over all three scenarios.

Figure 5.9 depicts the distribution of the performance measures attained through each feature extraction technique. The boxes illustrate the distribution range of the performance measures (dots) between the first and third quartiles, solid brown square represents the mean of the performance measures for each FE technique, solid red line indicates the median value of the performance measures attained through each FE technique, and green dash lines represents the outlier range. These feature extraction techniques are ranked w.r.t. the mean values as FFT, WPT, EMD, SSA and Time domain analysis. FFT has also the smallest box, which shows the minimum variation of the performance measures and the most stable FE technique. Time domain features produce the largest window and also has the maximum number of outliers, i.e., the most unstable diagnostic performances.

Figure 5.10 shows the distribution of the performance measures obtained through each feature selection technique. These feature selection methods are ranked w.r.t.

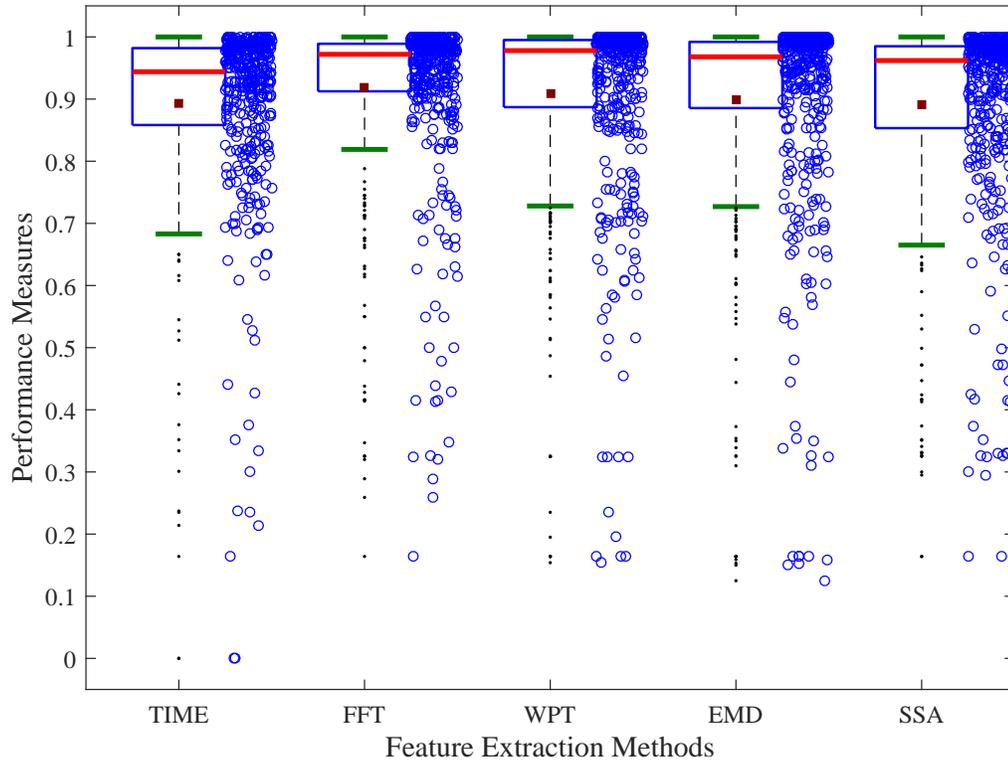


Figure 5.9 – Boxplot represents distributions of the performance measures attained by each feature extraction techniques

the mean values as FFS, FS and mrMR. FFS has also the smallest box, which indicates the least variation and maximum stability among other competitors. However, the performance measures attained by these feature selection methods are very close to each other.

Figure 5.11 represents the distribution of the performance measures achieved through each dimensionality reduction technique. These DR techniques are ranked w.r.t. the mean values as NCA, LDA, MCML, PCA and LLC. NCA has also the smallest box, which indicates the least variation and the most stable DR technique. LLC yields the largest box and the maximum number of outliers, which represent the maximum instability among other DR techniques.

Figure 5.12 illustrates the distribution of the performance measures attained by each CIL technique. These fault classification techniques are ranked w.r.t. the mean

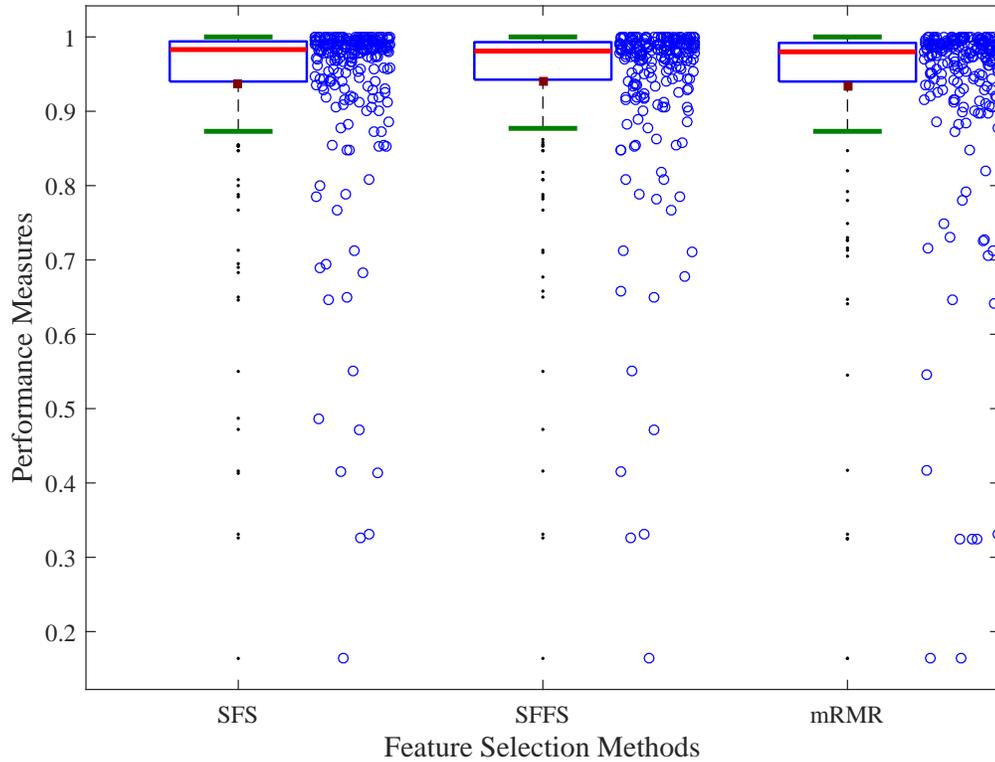


Figure 5.10 – Boxplot represents distributions of the performance measures attained by each feature selection techniques

values as EMI-OS, SMOTE, SMOTE BAG, WELM and RUSBoost. EMI-OS has also the smallest box, which represents the least variation and the most stable CIL technique. EMI-OS significantly outperforms other CIL techniques. RUSBoost and WELM yield the largest boxes and the maximum number of outliers, respectively, which result in the most unstable diagnostic performances.

Figure 5.13 shows the distribution of the performance measures obtained by each CIL technique through each FE technique for the scenario with the low imbalance ratio (LCI). The mean value of the performance measures obtained by each class imbalance learning technique is shown with a distinct marker. The average of performance measures over CIL methods w.r.t each FE techniques are calculated and connected by means of a distinct line for the sake of better comparison.

This explanation is valid for the rest of figures in this study.

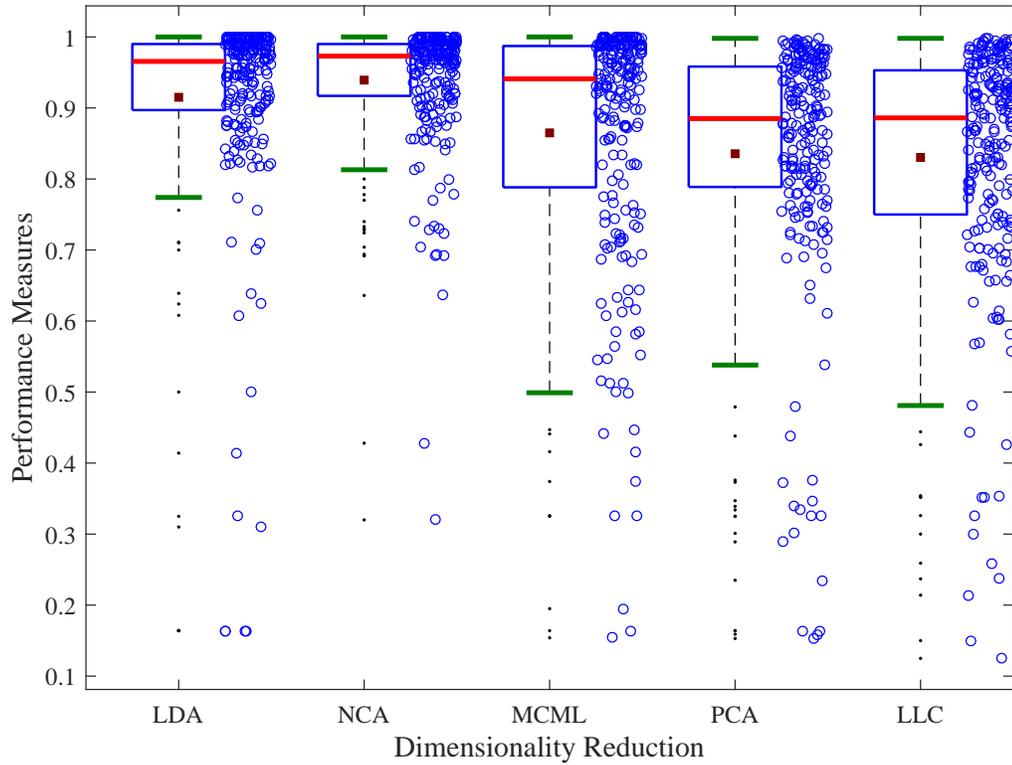


Figure 5.11 – Boxplot represents distributions of the performance measures attained by each dimensionality reduction techniques

Figure 5.13 shows that EMI-OS outperforms other CIL techniques regardless of the type of feature extraction, i.e., EMI-OS is insensitive to the type of extracted features. It is also the most stable fault classification technique. The rest of the CIL techniques are ranked as SMOTE, SMOTEBag, RUSBoost and WELM. The last two are the most unstable CIL methods.

In Figure 5.14, the level of imbalance is increased to the medium level (i.e., MCI). This figure represents the distribution of the performance measures obtained by each CIL technique through each feature extraction technique for the scenario with the medium imbalance ratio. Increasing the imbalance ratio does not change the rank of CIL techniques except for WELM and RUSBoost. This indicates that EMI-OS outperforms other CIL techniques regardless of the type of extracted features and the imbalance ratios.

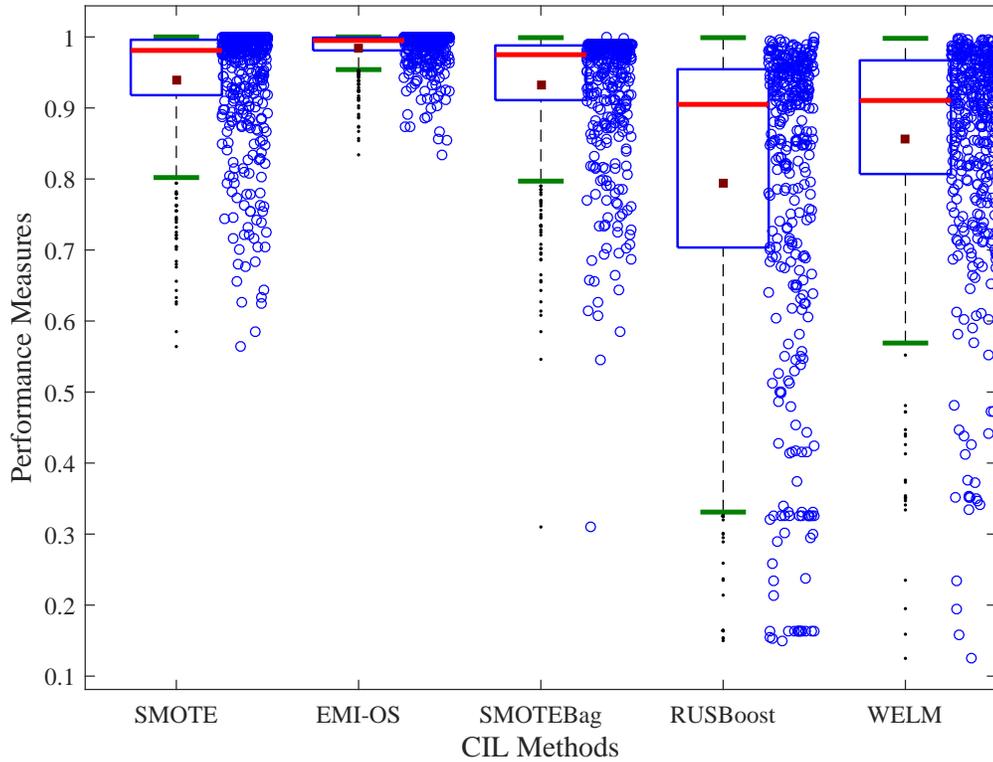


Figure 5.12 – Boxplot represents distributions of the performance measures attained by each CIL technique

SMOTE and SOMTEB_{ag} are the subsequent ranks with a slight difference. WELM outperforms RUSBoost with increasing the imbalance ratio and, thus, WELM and RUSBoost are ranked fourth and fifth for the scenarios with medium and high imbalance ratios.

Figures 5.13 and 5.14 show that increasing the imbalance ratio results in significant reduction of the performance measures obtained by RUSBoost and WELM, i.e., these techniques are very sensitive to the imbalance ratio and their performances decrease by increasing the imbalance ratio. On the other hand, EMI-OS outperforms other CIL techniques and is robust to the change of the imbalance ratio. The performance measures obtained by EMI-OS are even improved by increasing the imbalance ratio, which is due to the use of EMI for producing a large number of discriminant samples.

Figure 5.15 shows the distribution of the performance measures obtained by each

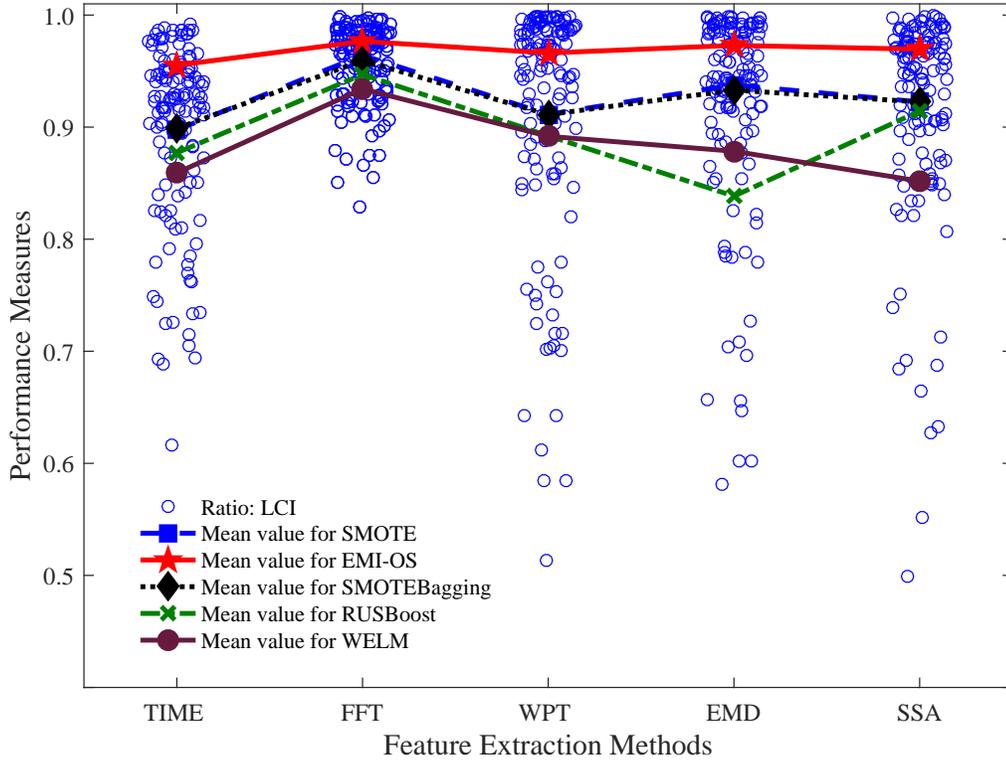


Figure 5.13 – Boxplot illustrates distributions of the performance measures on the scenario of the Low imbalance ratio attained by each feature extraction technique w.r.t. various CIL techniques (i.e., the mean values attained by each CIL technique are connected for a better visual evaluation)

CIL technique through each dimensionality reduction and feature selection technique for the scenario with the low imbalance ratio. The mean value of the performance measures obtained by each CIL technique is shown with a distinct marker. The figure shows that EMI-OS outperforms other CIL techniques regardless of the type of dimensionality reduction and feature selection techniques used for the feature reduction. EMI-OS is also the most stable fault classification technique. The rest of the CIL techniques are ranked as SMOTE, SMOTEBag, RUSBoost and WELM. The last two are the most unstable CIL techniques.

Considering feature reduction methods, increasing the level of imbalance to the high level could make an effect on some CIL techniques. Figure 5.16 shows the distribution of the performance measures obtained by each CIL technique through

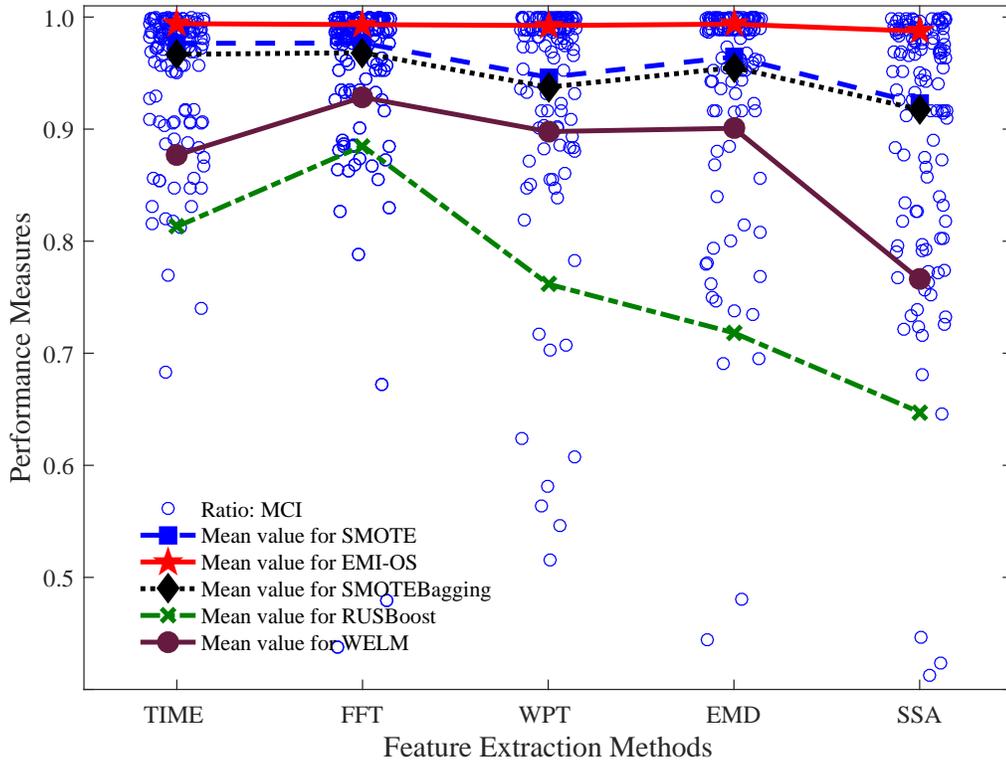


Figure 5.14 – Boxplot illustrates distributions of the performance measures on the scenario of the medium imbalance ratio attained by each feature extraction technique w.r.t. various CIL techniques (i.e., the mean values attained by each CIL technique are connected for a better visual evaluation)

each dimensionality reduction and feature selection technique for the scenario with the high imbalance ratio. The figure shows that increasing the imbalance ratio change the performance of WELM and RUSBoost considerably. Moreover, This indicates that EMI-OS outperforms other CIL techniques regardless of the type of feature reduction and the imbalance ratio. SMOTE and SOMTEBag are the subsequent ranks with a slight difference. WELM outperforms RUSBoost with increasing the imbalance ratio and, thus, WELM and RUSBoost are ranked fourth and fifth for the scenarios with medium and high imbalance ratios.

Figures 5.15 and 5.16 also confirm that increasing the imbalance ratio results in significant reduction of the performance measures obtained by RUSBoost and WELM, i.e., these techniques are very sensitive to the imbalance ratio. They also show that

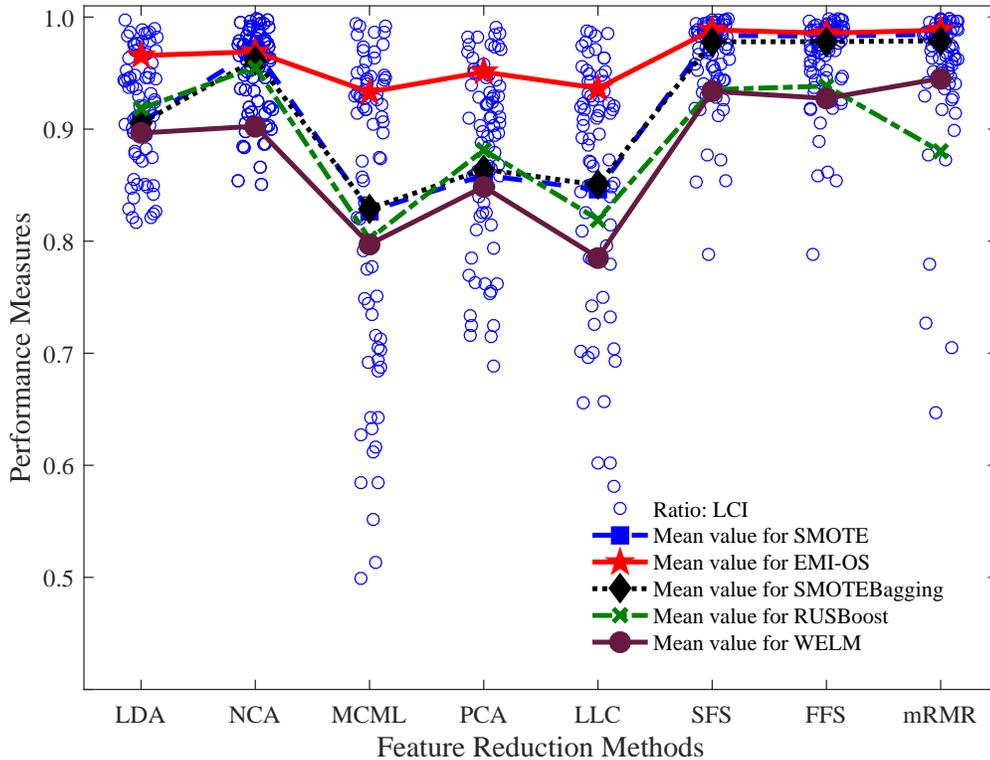


Figure 5.15 – Boxplot illustrates distributions of the performance measures on the scenario of the Low imbalance ratio attained by each feature reduction technique w.r.t. various CIL techniques (i.e., the mean values attained by each CIL technique are connected for a better visual evaluation)

EMI-OS outperforms other CIL techniques and is not very sensitive to the change of the imbalance ratio. They also indicate that the performance measures obtained by resorting to EMI-OS are even improved by increasing the imbalance ratio.

Figures 5.15 and 5.16 also illustrate the rank of dimensionality reduction and feature selection techniques. In general feature selection techniques outperform dimensionality reduction techniques as already represented in Figures 5.10 and 5.11. Feature selection techniques are also more stable compared to the dimensionality reduction techniques. NCA outperforms other dimensionality reduction techniques. The attained performance measures through NCA are comparable with the attained results through feature selection techniques. Figures 5.15 and 5.16 also shows that, with increasing the imbalance ratio, the difference between the attained results through

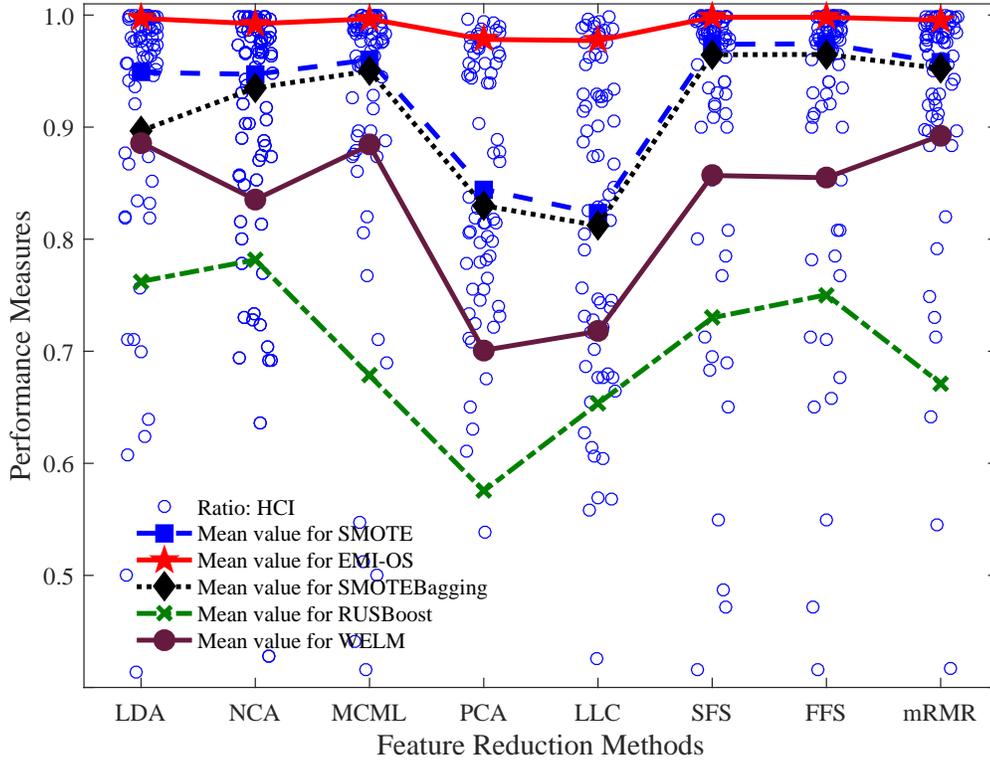


Figure 5.16 – Boxplot illustrates distributions of the performance measures on the scenario of the high imbalance ratio attained by each feature reduction technique w.r.t. various CIL techniques (i.e., the mean values attained by each CIL technique are connected for a better visual evaluation)

the feature selection techniques and the supervised dimensionality reduction techniques (NCA, LDA and MCML), is reduced. On the other hand, the unsupervised dimensionality reduction techniques (PCA and LLC) under-perform other competitors. This is particularly more significant for the scenario with the high imbalance ratio.

The attained results in Figures 5.15 and 5.16 show that feature selection techniques have first rank, followed by supervised and unsupervised dimensionality reduction techniques as second and third ranks, respectively, in diagnosing bearing defects under CI conditions.

In order to find the most relevant features that have the major impact on improvement of the classification performance, Table 5.6, Table 5.7 and Table 5.8 are

provided. The common features selected by all the three feature selection methods, i.e., SFS, SFFS and mRMR are presented in these tables. From Table 5.6, Table 5.7 and Table 5.8, it can be concluded that RMS, Mean value and Entropy are the most important features in LCI, MCI and HCI conditions, respectively; while margin factor does not have any effect on the performance of fault classifiers. In addition the most informative packs among 32 packs of the wavelet packet transform are packs number 10, 13, 18 and 23. For empirical mode decomposition, the first three IMFs play a vital role to provide proper features. Besides, components number 1, 2 and 9 extracted by means of SSA technique appear as important features in further processing by feature selection methods.

Table 5.6 – Common features selected by all FS methods under LCI condition

FE method	Feature No.	Statistical Feature	Component No.
Time	1, 3, 4, 5	RMS, Skewness, Kurtosis, Crest factor	-
FFT	1, 2, 8	RMS, Mean, Entropy	-
WPT	177	RMS	23 rd pack
EMD	1, 18, 24	RMS, Mean, Entropy	1 st and 3 rd IMFs
SSA	9, 24, 66	RMS, Entropy, Mean	2 nd , 3 rd and 9 th Comp.

Table 5.7 – Common features selected by all FS methods under MCI condition

FE method	Feature No.	Statistical Feature	Component No.
Time	2	Mean	-
FFT	1	RMS	-
WPT	73, 178	RMS, Mean	10 th and 23 rd packs
EMD	1, 10, 16	RMS, Mean, Entropy	1 st and 2 nd IMFs
SSA	18	Mean	2 nd Component

Table 5.8 – Common features selected by all FS methods under HCI condition

FE method	Feature No.	Statistical Feature	Component No.
Time	1, 6, 8	RMS, Impulse, Entropy	-
FFT	3, 8	Skewness, Entropy	-
WPT	98, 144	Mean, Entropy	13 rd and 18 th packs
EMD	1, 8	RMS, Entropy	1 st IMFs
SSA	65	RMS	9 th Component

5.5 Summary

In this chapter, the proper metrics to evaluate the performance of the diagnostic system are firstly introduced. Then, the chapter provides the experimental results related to three different experiments. The diagnostic system of the first experiment focuses on preprocessing methods. It considers the wavelet packet transform for analyzing the vibration signals and, then, liner discriminant analysis to provide a new feature set with a lower dimension and more discriminant ability.

The second experiment, focuses on designing an efficient multi-step pre-processing scheme to diagnose bearing defects in IMs. In this experiment, the combination of empirical mode decomposition with five different state-of-the-art dimensionality reduction techniques are considered. The diagnostic system makes use of the newly transformed features to train various fault classifiers about bearing faults. The experimental results provide a comparison to show the effect of various DR techniques (i.e., supervised and unsupervised) on the performance of the diagnostic system.

Lastly, the experimental results of the data-driven techniques in diagnosing bearing detects under class imbalance condition are presented. The proposed diagnostic system includes the state-of-the-art feature extraction and feature reduction methods and also contains the different class imbalance learning methods. The data-driven diagnostic scheme is shown in Figure 5.8. The proposed scheme is applied on the bearing datasets with three different imbalance ratios; Low, medium and high. The

achieved results are presented and the effect of CIL techniques along with the feature extraction and feature reduction methods are compared. From the obtained results, it can be seen that the proposed sampling technique, so-called EMI-OS, could perfectly enhance the accuracy of the data-driven diagnostic system under the class imbalance condition in all scenarios of this experiment. Moreover, the proposed method could maintain its performance in a high level, while different feature extraction and feature reduction techniques under low, medium and high imbalance ratios are applied.

Chapter 6

Conclusions

The ultimate goal of this study is to design an efficient integrated scheme for diagnosing bearing defects in IMs, under the class imbalance condition, which is quite relevant for actual diagnostic problems. To this aim, firstly an efficient pre-processing scheme, as the first experiment, is proposed for diagnosing bearing defects. The pre-processing scheme includes two main sub-modules: wavelet packet transform and linear discriminant analysis. The proposed method has been applied to the simulated bearing vibration signals, i.e., normal, ball defects, outer race defects and inner race defects with different diameters. The attained results have shown that the proposed method is effective in extracting discriminant features from the vibration signals, in a way that various fault classifiers in the diagnostic module can classify them with a very high accuracy.

Since the pre-processing module plays an important role to provide the most informative features for the diagnostic system, another state-of-the-art signal processing technique, called EMD, is investigated in the second experiment. The second scheme has also two main sub-modules: (1) extracting informative features, i.e., intrinsic mode functions, from the vibration signals by means empirical mode decomposition (EMD). The extracted IMFs are segmented to compute the entropy of each segment. The attained features are normalized, and, then, fed to the subsequent sub-modules; (2) feature reduction by means of different dimensionality reduction techniques to construct an informative set of small-sized features in both supervised and unsupervised manners, that best separate the different classes of bearing defects. To study the efficiency of these DR techniques, the state-of-the-art unsupervised and supervised

methods are implemented and compared for reducing the dimension of the features. The reduced features obtained by each DR technique are used to train various classifiers to diagnose bearing defects including normal state, ball defects, outer race defects and inner race defects with different diameters. The experimental results reveal the effectiveness of the developed feature extraction and reduction modules in providing discriminant features for diagnosing bearing defects.

The first and second experiments show the importance of feature extraction and dimensionality reduction techniques in processing the raw signals and providing useful features for the data-driven diagnostic system.

Moreover, in the real world applications, data samples are often collected under skewed-class distribution and there is a need for some techniques to ease the classification in class imbalance conditions and to facilitate training the fault classifiers. Hence, the classification module of a data-driven diagnostic system needs to be enhanced to be able to maintain a satisfied accuracy once encounters the class imbalance condition.

In order to diagnose multiple bearing defects under the class imbalance condition the third experiment is considered. The proposed scheme of the third experiment contains four modules including segmentation, feature extraction, feature reduction and fault classification. In the feature extraction module, several state-of-the-art techniques have been devised in the diagnostic scheme to extract informative sets of features from the vibrational signal. Feature reduction module includes various state-of-the-art feature selection and dimensionality reduction techniques, which aim to eliminate redundant features and generate discriminant and useful sets of features for the subsequent module. Fault classification module is made of some state-of-the-art data-level, algorithm-level and ensemble-based approaches for the class imbalance learning, that are adopted for diagnosing bearing defects.

In the fault classification module, a novel oversampling technique, called EMI-OS, for class-imbalance learning and diagnosing bearing defects has been developed, which

is based on the expectation maximization imputation technique. The novelty of the EMI-OS algorithm stands in producing a set of incomplete samples representative of each minor class and imputing them by resorting to the expectation maximization algorithm. This allows EMI-OS to generate new synthetic samples of the minor class. EMI-OS efficiently diagnoses bearing defects under the class-imbalance condition and improves performance measures. This improvement is more significant for the scenarios with the high class imbalance ratios. This scheme has been evaluated on the CWRU bearing datasets. Besides, three different scenarios are used to evaluate the sensitivity of the diagnostic system to the class imbalance ratio.

The achieved results show that the proposed novel oversampling technique, EMI-OS, outperforms other state-of-the-art CIL techniques in diagnosing bearing defects in terms of both performance measures and stability of the attained results. The experiments suggest that EMI-OS is a promising technique to deal with CI scenarios, extensible to classification of class-imbalance datasets and other areas of industrial applications, which looks to be a worthwhile direction for future research. This integrated diagnostic scheme also enables an empirical comparison in each module to study the impacts of the state-of-the-art feature extraction and reduction techniques in the diagnostic performance.

6.1 Improvements and Future Works

The proposed data-driven diagnosis system is applied on the CWRU bearing datasets. Further research can be conducted to evaluate the performance of this diagnostic system on the other bearing datasets with different operating loads. It is also noticeable that this work only considers the bearing defects since they are the major reason for induction motors failure. However, there exist other mechanical and electrical faults such as broken rotor bar or those that are initiated in stator windings. This diagnostic system can also be evaluated to diagnose these faults in the future works.

References

- [1] M. Farajzadeh-Zanjani, R. Razavi-Far, and M. Saif, “Efficient sampling techniques for ensemble learning and diagnosing bearing defects under class imbalanced condition,” in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, Dec 2016, pp. 1–7.
- [2] —, “Efficient feature extraction of vibration signals for diagnosing bearing defects in induction motors,” in *Proceedings of IEEE Joint Congress on Neural Networks (IJCNN)*, 2016, pp. 4505–4511.
- [3] A. H. Bonnett and C. Yung, “Increased efficiency versus increased reliability,” *IEEE Industry Applications Magazine*, vol. 14, no. 1, pp. 29–36, Jan 2008.
- [4] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, “A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 463–484, July 2012.
- [5] S. Nandi and H. Toliyat, “Condition monitoring and fault diagnosis of electrical machines—a review,” in *34th IAS Annual Meeting, IEEE Industry Applications Conf.*, vol. 1, 1999, pp. 197–204.
- [6] “Report of large motor reliability survey of industrial and commercial installations, part i,” *IEEE Trans. on Industry App.*, vol. IA-21, no. 4, pp. 853–864, 1985.
- [7] “Case western reserve university bearing data center website.” [Online]. Available: <http://csegroups.case.edu/bearingdatacenter/home>

- [8] W. A. Smith and R. B. Randall, “Rolling element bearing diagnostics using the case western reserve university data: A benchmark study,” *Mechanical Systems and Signal Processing*, vol. 64-65, pp. 100–131, 2015.
- [9] V. Rai and A. Mohanty, “Bearing fault diagnosis using FFT of intrinsic mode functions in hilberthuang transform,” *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2607 – 2615, 2007.
- [10] H. Liu and M. Han, “A fault diagnosis method based on local mean decomposition and multi-scale entropy for roller bearings,” *Mechanism and Machine Theory*, vol. 75, pp. 67 – 78, 2014.
- [11] M. Farajzadeh-Zanjani, R. Razavi-Far, M. Saif, J. Zarei, and V. Palade, “Diagnosis of bearing defects in induction motors by fuzzy-neighborhood density-based clustering,” in *IEEE International Conference on Machine Learning and Applications, ICMLA2015*, Miami, Florida, USA, December 9-11 2015.
- [12] R. Razavi-Far, H. Davilu, V. Palade, and C. Lucas, “Neuro-fuzzy based fault diagnosis of a steam generator,” in *7th IFAC Conf. on Fault Detection, Supervision and Safety of Technical Processes*, Barcelona, Spain, 2009, pp. 1180–1185.
- [13] R. Razavi-Far and M. Kinnaert, “A multiple observers and dynamic weighting ensembles scheme for diagnosing new class faults in wind turbines,” *Control Eng. Pract.*, vol. 21(9), pp. 1165–1177, 2013.
- [14] Z. Xu, J. Xuan, T. Shi, B. Wu, and Y. Hu, “A novel fault diagnosis method of bearing based on improved fuzzy ARTMAP and modified distance discriminant technique,” *Expert Syst. Appl.*, vol. 36, pp. 11 801–11 807, 2009.
- [15] R. Razavi-Far and M. Kinnaert, “Incremental design of a decision system for residual evaluation: A wind turbine application,” in *8th IFAC Conference Fault*

- Detection, Supervision and Safety of Technical Processes*, vol. 8(1), 2012, pp. 343–348.
- [16] H. He and E. A. Garcia, “Learning from imbalanced data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, Sept 2009.
- [17] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote: Synthetic minority over-sampling technique,” *J. Artif. Int. Res.*, vol. 16, no. 1, pp. 321–357, Jun. 2002.
- [18] X. Liu and Z. Zhou, “The influence of class imbalance on cost-sensitive learning: An empirical study,” in *Sixth International Conference on Data Mining (ICDM’06)*, Dec 2006, pp. 970–974.
- [19] M. Benbouzid, “Bibliography on induction motors faults detection and diagnosis,” *IEEE Trans. Ener. Con.*, vol. 14, pp. 1065–1074, 1999.
- [20] G. Singh and S. Kazzaz, “Induction machine drive condition monitoring and diagnostic researcha survey,” *Electric Power Systems Research*, vol. 64, pp. 145–158, 2003.
- [21] 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, pp. 469–480, 1999.
- [22] J. Zarei, “Induction motors bearing fault detection using pattern recognition techniques,” *Expert Syst. Appl.*, vol. 39, pp. 68–73, 2012.
- [23] A. Bellini, F. Filippetti, C. Tassoni, and G. A. Capolino, “Advances in diagnostic techniques for induction machines,” *IEEE Transactions on Industrial Electronics*, vol. 55, no. 12, pp. 4109–4126, Dec 2008.

- [24] R. Randall and J. Antoni, "Rolling element bearing diagnostics: A tutorial," *Mechanical Systems and Signal Processing*, vol. 25, no. 2, pp. 485 – 520, 2011.
- [25] J. Zarei and J. Poshtan, "Bearing fault detection using wavelet packet transform of induction motor stator current," *Tribology International*, vol. 40, pp. 763–769, 2007.
- [26] 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.
- [27] P. McFadden and J. Smith, "Model for the vibration produced by a single point defect in a rolling element bearing," *Journal of Sound and Vibration*, vol. 96, no. 1, pp. 69 – 82, 1984.
- [28] M. E. H. Benbouzid and G. B. Kliman, "What stator current processing-based technique to use for induction motor rotor faults diagnosis?" *IEEE Transactions on Energy Conversion*, vol. 18, no. 2, pp. 238–244, June 2003.
- [29] A. H. Bonnett and G. C. Soukup, "Cause and analysis of stator and rotor failures in three-phase squirrel-cage induction motors," *IEEE Transactions on Industry Applications*, vol. 28, no. 4, pp. 921–937, Jul 1992.
- [30] M. E. H. Benbouzid, "A review of induction motors signature analysis as a medium for faults detection," *IEEE Transactions on Industrial Electronics*, vol. 47, no. 5, pp. 984–993, Oct 2000.
- [31] R. M. Tallam, S. B. Lee, G. C. Stone, G. B. Kliman, J. Yoo, T. G. Habetler, and R. G. Harley, "A survey of methods for detection of stator-related faults in induction machines," *IEEE Transactions on Industry Applications*, vol. 43, no. 4, pp. 920–933, July 2007.

- [32] M. Kang, J. Kim, L. M. Wills, and J. M. Kim, "Time-varying and multiresolution envelope analysis and discriminative feature analysis for bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 12, pp. 7749–7761, Dec 2015.
- [33] D. U. Campos-Delgado and D. R. Espinoza-Trejo, "An observer-based diagnosis scheme for single and simultaneous open-switch faults in induction motor drives," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 2, pp. 671–679, Feb 2011.
- [34] H. Mekki, O. Benzineb, D. Boukhetala, M. Tadjine, and M. Benbouzid, "Sliding mode based fault detection, reconstruction and fault tolerant control scheme for motor systems," *{ISA} Transactions*, vol. 57, pp. 340 – 351, 2015.
- [35] J. Zarei, M. A. Tajeddini, and H. R. Karimi, "Vibration analysis for bearing fault detection and classification using an intelligent filter," *Mechatronics*, vol. 24, no. 2, pp. 151–157, 2014.
- [36] J. Qu, Z. Zhang, and T. Gong, "A novel intelligent method for mechanical fault diagnosis based on dual-tree complex wavelet packet transform and multiple classifier fusion," *Neurocomputing*, vol. 171, pp. 837 – 853, 2016.
- [37] Y. Tian, J. Ma, C. Lu, and Z. Wang, "Rolling bearing fault diagnosis under variable conditions using lmd-svd and extreme learning machine," *Mechanism and Machine Theory*, vol. 90, pp. 175 – 186, 2015.
- [38] M. Kang and J.-M. Kim, "Singular value decomposition based feature extraction approaches for classifying faults of induction motors," *Mechanical Systems and Signal Processing*, vol. 41, pp. 348 – 356, 2013.

- [39] P. Konar and P. Chattopadhyay, “Bearing fault detection of induction motor using wavelet and support vector machines (svms),” *Applied Soft Computing*, vol. 11, no. 6, pp. 4203 – 4211, 2011.
- [40] P. Baraldi, F. Cannarile, F. D. Maio, and E. Zio, “Hierarchical k-nearest neighbours classification and binary differential evolution for fault diagnostics of automotive bearings operating under variable conditions,” *Engineering Applications of Artificial Intelligence*, vol. 56, pp. 1–13, 2016.
- [41] V. C. M. N. Leite, J. G. B. da Silva, G. F. C. Veloso, L. E. B. da Silva, G. Lambert-Torres, E. L. Bonaldi, and L. E. d. L. de Oliveira, “Detection of localized bearing faults in induction machines by spectral kurtosis and envelope analysis of stator current,” *IEEE Transactions on Industrial Electronics*, vol. 62, no. 3, pp. 1855–1865, 2015.
- [42] Y. Trachi, E. Elbouchikhi, V. Choqueuse, and M. E. H. Benbouzid, “Induction machines fault detection based on subspace spectral estimation,” *IEEE Transactions on Industrial Electronics*, vol. 63, no. 9, pp. 5641–5651, 2016.
- [43] D. Z. Li, W. Wang, and F. Ismail, “An enhanced bispectrum technique with auxiliary frequency injection for induction motor health condition monitoring,” *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 10, pp. 2679–2687, 2015.
- [44] 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. 12, pp. 108 – 126, 2013.
- [45] B. Li, M. Chow, Y. Tipsuwan, and J. Hung, “Neural-network-based motor rolling bearing fault diagnosis,” *IEEE Trans. Ind. Elec.*, vol. 47, pp. 1060–1069, 2000.

- [46] F. Dalvand, A. Kalantar, and M. S. Safizadeh, "A novel bearing condition monitoring method in induction motors based on instantaneous frequency of motor voltage," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 1, pp. 364–376, 2016.
- [47] X. Jin, M. Zhao, T. W. S. Chow, and M. Pecht, "Motor bearing fault diagnosis using trace ratio linear discriminant analysis," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 5, pp. 2441–2451, 2014.
- [48] T. W. Rauber, F. de Assis Boldt, and F. M. Varejo, "Heterogeneous feature models and feature selection applied to bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 1, pp. 637–646, Jan 2015.
- [49] C. M. Vong, P. K. Wong, and W. F. Ip, "A new framework of simultaneous-fault diagnosis using pairwise probabilistic multi-label classification for time-dependent patterns," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 8, pp. 3372–3385, Aug 2013.
- [50] D. He, R. Li, and J. Zhu, "Plastic bearing fault diagnosis based on a two-step data mining approach," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 8, pp. 3429–3440, Aug 2013.
- [51] A. Boudiaf, A. Moussaoui, A. Dahane, and I. Atoui, "A comparative study of various methods of bearing faults diagnosis using the case western reserve university data," *Journal of Failure Analysis and Prevention*, vol. 16, no. 2, pp. 271–284, 2016.
- [52] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: A review with applications," *Signal Processing*, vol. 96, Part A, pp. 1–15, 2014.

- [53] L. Eren and M. J. Devaney, “Bearing damage detection via wavelet packet decomposition of the stator current,” *IEEE Transactions on Instrumentation and Measurement*, vol. 53, no. 2, pp. 431–436, 2004.
- [54] N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N. Yen, C. Tung, and 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 A: Mathematical, Physical and Engineering Sciences*, vol. 454(1971), pp. 903–995, 1998.
- [55] G. Rilling, P. Flandrin, P. Goncalves *et al.*, “On empirical mode decomposition and its algorithms,” *IEEE-EURASIP workshop on nonlinear signal and image processing*, vol. 3, pp. 8–11, 2003.
- [56] B. Muruganatham, M. Sanjith, B. Krishnakumar, and S. S. Murty, “Roller element bearing fault diagnosis using singular spectrum analysis,” *Mechanical Systems and Signal Processing*, vol. 35, no. 12, pp. 150 – 166, 2013.
- [57] H. Jiang, J. Chen, G. Dong, T. Liu, and G. Chen, “Study on hankel matrix-based SVD and its application in rolling element bearing fault diagnosis,” *Mechanical Systems and Signal Processing*, vol. 52-53, pp. 338–359, 2015.
- [58] R. Golafshan and K. Y. Sanliturk, “SVD and hankel matrix based de-noising approach for ball bearing fault detection and its assessment using artificial faults,” *Mechanical Systems and Signal Processing*, vol. 70-71, pp. 36 – 50, 2016.
- [59] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, 2003.
- [60] M. Kudo and J. Sklansky, “Comparison of algorithms that select features for pattern classifiers,” *Pattern Recognition*, vol. 33, no. 1, pp. 25–41, 2000.

- [61] M. Gutlein, E. Frank, M. Hall, and A. Karwath, “Large-scale attribute selection using wrappers,” in *2009 IEEE Symposium on Computational Intelligence and Data Mining*, March 2009, pp. 332–339.
- [62] H. Peng, F. Long, and C. Ding, “Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, Aug 2005.
- [63] P. Pudil, J. Novovičová, and J. Kittler, “Floating search methods in feature selection,” *Pattern Recogn. Lett.*, vol. 15, no. 11, pp. 1119–1125, Nov. 1994.
- [64] R. Kohavi and G. H. John, “Wrappers for feature subset selection,” *Artificial Intelligence*, vol. 97, no. 1, pp. 273 – 324, 1997.
- [65] L. Van der Maaten, E. Postma, and H. Van den Herik, “Dimensionality reduction: A comparative review,” *Technical Report TiCC TR 2009-005*, 2009.
- [66] Y. W. Teh and S. Roweis, “Automatic alignment of local representations,” in *In Advances in Neural Information Processing Systems 15*. MIT Press, 2003, pp. 841–848.
- [67] A. Martinez and A. Kak, “PCA versus LDA,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, 2001.
- [68] L. Rueda, B. J. Oommen, and C. Henrquez, “Multi-class pairwise linear dimensionality reduction using heteroscedastic schemes,” *Pattern Recognition*, vol. 43, no. 7, pp. 2456 – 2465, 2010.
- [69] J. Goldberger, G. Hinton, S. Roweis, and R. Salakhutdinov, “Neighbourhood components analysis,” in *Advances in neural information processing systems*, 2004, pp. 513–520.

- [70] A. Globerson and S. Roweis, “Metric learning by collapsing classes.”
- [71] I. Jolliffe, *Principal Component Analysis*, 2nd ed. Springer, 2002.
- [72] R. Razavi-Far, E. Zio, and V. Palade, “Efficient residuals pre-processing for diagnosing multi-class faults in a doubly fed induction generator, under missing data scenarios,” *Expert Syst. Appl.*, vol. 41, no. 14, pp. 6386–6399, 2014.
- [73] Z.-B. Zhu and Z.-H. Song, “Fault diagnosis based on imbalance modified kernel fisher discriminant analysis,” *Chemical Engineering Research and Design*, vol. 88, no. 8, pp. 936 – 951, 2010.
- [74] G. Ditzler and R. Polikar, “Incremental learning of concept drift from streaming imbalanced data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 10, pp. 2283–2301, 2013.
- [75] C. Seiffert, T. M. Khoshgoftaar, J. V. Hulse, and A. Napolitano, “Rusboost: A hybrid approach to alleviating class imbalance,” *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 40, no. 1, pp. 185–197, 2010.
- [76] T. M. Khoshgoftaar, J. V. Hulse, and A. Napolitano, “Comparing boosting and bagging techniques with noisy and imbalanced data,” *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 41, no. 3, pp. 552–568, May 2011.
- [77] G. Wu and E. Y. Chang, “Kba: kernel boundary alignment considering imbalanced data distribution,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 786–795, June 2005.
- [78] R. Barandela, J.S.Sanchez, V. Garcia, and E. Rangel, “Strategies for learning in class imbalance problems,” *Pattern Recognition*, vol. 36, no. 3, pp. 849–851, 2003.

- [79] C. Ling, V. Sheng, and Q. Yang, “Test strategies for cost-sensitive decision trees,” *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 8, pp. 1055–1067, 2006.
- [80] M. Maloof, “Learning when data sets are imbalanced and when costs are unequal and unknown,” *Proc. Intl Conf. Machine Learning, Workshop Learning from Imbalanced Data Sets II*, 2003.
- [81] M. Kukar and I. Kononenko, “Cost-sensitive learning with neural networks,” in *Proc. European Conf. Artificial Intelligence*, 1998, pp. 445–449.
- [82] W. Fan, S. Stolfo, J. Zhang, and P. Chan, “Adacost: Misclassification cost-sensitive boosting,” in *6th Int. Conf. Mach. Learning*, 1999, pp. 97–105.
- [83] Y. Sun, M. Kamel, A. Wong, and Y. Wang, “Cost-sensitive boosting for classification of imbalanced data,” *Pattern Recognition*, vol. 40, no. 12, pp. 3358–3378, 2007.
- [84] K. McCarthy, B. Zabar, and G. Weiss, “Does cost-sensitive learning beat sampling for classifying rare classes?” in *Proceedings of the 1st International Workshop on Utility-based Data Mining*, ser. UBDM '05. New York, NY, USA: ACM, 2005, pp. 69–77.
- [85] T. Jo and N. Japkowicz, “Class imbalances versus small disjuncts,” *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 1, pp. 40–49, 2004.
- [86] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 1, pp. 1–38, 1977.
- [87] N. Chawla, A. Lazarevic, L. Hall, and K.W. Bowyer, “Smoteboost: Improving prediction of the minority class in boosting,” in *Proc. Knowl. Discov. Databases*, 2003, pp. 107–119.

- [88] H. Guo and H. Viktor, “Learning from imbalanced data sets with boosting and data generation: The databoost-im approach,” *SIGKDD Expl. Newsl.*, vol. 6, pp. 30–39, 2004.
- [89] S. Wang and X. Yao, “Diversity analysis on imbalanced data sets by using ensemble models,” in *Computational Intelligence and Data Mining, 2009. CIDM '09. IEEE Symposium on*, March 2009, pp. 324–331.
- [90] R. Barandela, R. Valdovinos, and J. Sanchez, “New applications of ensembles of classifiers,” *Pattern Anal. App.*, vol. 6, pp. 245–256, 2003.
- [91] D. Tao, X. Tang, X. Li, and X. Wu, “Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 7, pp. 1088–1099, 2006.
- [92] C. Li, “Classifying imbalanced data using a bagging ensemble variation ({BEV}),” in *Proc. 45th Annual Southeast Regional Conference*, ACM, Ed., 2007, pp. 203–208.
- [93] X. Y. Liu, J. Wu, and Z. H. Zhou, “Exploratory undersampling for class-imbalance learning,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 2, pp. 539–550, April 2009.
- [94] M. Galar, A. Fernandez, E. Barrenechea, and F. Herrera, “Eusboost: Enhancing ensembles for highly imbalanced data-sets by evolutionary undersampling,” *Pattern Recognition*, vol. 46, no. 12, pp. 3460–3471, 2013.
- [95] X. Hong, S. Chen, and C. Harris, “A kernel-based two-class classifier for imbalanced data sets,” *IEEE Trans. Neural Networks*, vol. 18, no. 1, pp. 28–41, 2007.
- [96] W. Zong, G.-B. Huang, and Y. Chen, “Weighted extreme learning machine for imbalance learning,” *Neurocomputing*, vol. 101, pp. 229 – 242, 2013.

- [97] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: Theory and applications,” *Neurocomputing*, vol. 70, pp. 489 – 501, 2006.
- [98] G. B. Huang, H. Zhou, X. Ding, and R. Zhang, “Extreme learning machine for regression and multiclass classification,” *IEEE Transactions on systems, man, and cyberneticsPART B: Cybernetics*, Vol. 42, No. 2,, April 2012.
- [99] R. Fletcher, *Practical Methods of Optimization, Constrained Optimization*. John Wiley & Sons, Inc, 1981, vol. 2.
- [100] J. Snchez-Crisostomo, R. Alejo, E. Lpez-Gonzlez, R. Valdovinos, and J. Pacheco-Snchez, “Empirical analysis of assessments metrics for multi-class imbalance learning on the back-propagation context,” in *Advances in Swarm Intelligence*, ser. Lecture Notes in Computer Science. Springer International Publishing, 2014, vol. 8795, pp. 17–23.
- [101] K. Goseva-Popstojanova, G. Anastasovski, and R. Pantev, “Using multiclass machine learning methods to classify malicious behaviors aimed at web systems,” in *Software Reliability Engineering (ISSRE), 2012 IEEE 23rd International Symposium on*, 2012, pp. 81–90.
- [102] T. Fawcett, “An introduction to roc analysis,” *Pattern Recogn. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.
- [103] Y. JinLiang, Z. Yongli, and Y. Guoqin, “Power transformer fault diagnosis based on support vector machine with cross validation and genetic algorithm,” in *Advanced Power System Automation and Protection (APAP), 2011 International Conference on*, vol. 1, 2011, pp. 309–313.
- [104] D. J. Drown, T. M. Khoshgoftaar, and N. Seliya, “Evolutionary sampling and software quality modeling of high-assurance systems,” *IEEE Transactions on*

Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 39, no. 5, pp. 1097–1107, 2009.

- [105] C. T. Su and Y. H. Hsiao, “An evaluation of the robustness of mts for imbalanced data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 10, pp. 1321–1332, 2007.

Appendix A

Copyright Permissions



Home Create Account Help



Requesting permission to reuse content from an IEEE publication

Title: Efficient sampling techniques for ensemble learning and diagnosing bearing defects under class imbalanced condition

Conference Proceedings: 2016 IEEE Symposium Series on Computational Intelligence (SSCI)

Author: Maryam Farajzadeh-Zanjani; Roozbeh Razavi-Far; Mehrdad Saif

Publisher: IEEE

Date: 6-9 Dec. 2016

Copyright © 2016, IEEE

LOGIN

If you're a copyright.com user, you can login to RightsLink using your copyright.com credentials. Already a RightsLink user or want to learn more?

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

- 1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
- 2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
- 3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

- 1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
- 2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
- 3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK

CLOSE WINDOW

Copyright © 2017 Copyright Clearance Center, Inc. All Rights Reserved. [Privacy statement](#). [Terms and Conditions](#).
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



Title: Efficient feature extraction of vibration signals for diagnosing bearing defects in induction motors

Conference Proceedings: 2016 International Joint Conference on Neural Networks (IJCNN)

Author: Maryam Farajzadeh-Zanjani; Roozbeh Razavi-Far; Mehrdad Saif; Luis Rueda

Publisher: IEEE

Date: 24-29 July 2016

Copyright © 2016, IEEE

LOGIN
If you're a [copyright.com](#) user, you can login to RightsLink using your [copyright.com](#) credentials. Already a RightsLink user or want to [learn more?](#)

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

- 1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
- 2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
- 3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

- 1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
- 2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
- 3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK

CLOSE WINDOW



Title: Diagnosis of Bearing Defects in Induction Motors by Fuzzy-Neighborhood Density-Based Clustering

Conference Proceedings: 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)

Author: M. Farajzadeh-Zanjani; R. Razavi-Far; M. Saif; J. Zarei; V. Palade

Publisher: IEEE

Date: 9-11 Dec. 2015

Copyright © 2015, IEEE

LOGIN

If you're a [copyright.com](#) user, you can login to RightsLink using your [copyright.com](#) credentials. Already a [RightsLink](#) user or want to [learn more?](#)

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

- 1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
- 2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
- 3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

- 1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
- 2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
- 3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK

CLOSE WINDOW

Vita Auctoris

Maryam Farajzadeh-zanjani was born in 1986 in Mashhad, Iran. She completed her bachelor degree in computer hardware engineering in 2008. She received her MSc degree in computer engineering from Eastern Mediterranean University in Cyprus. She is currently a candidate for the Master's degree in Electrical and Computer Engineering at the University of Windsor and hopes to graduate in spring 2017. She has multi-discipline background in electrical, computer engineering and computer science. She has done research projects in cross-cutting areas with priorities in computational intelligence, machine learning, data mining, optimization, modeling, cybernetics, cyber security and fault diagnosis systems.