



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

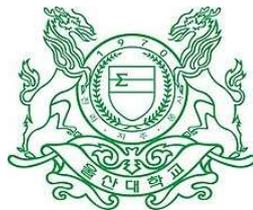
저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

DOCTOR OF PHILOSOPHY

**FAULT DIAGNOSIS OF ROTATING MACHINES BASED
ON ADVANCED SIGNAL PROCESSING METHODS AND
DATA-DRIVEN ARTIFICIAL INTELLIGENCE
TECHNIQUES**



**The Graduate School
of the University of Ulsan
Department of Computer Engineering**

ALEXANDER E. PROSVIRIN

**Fault Diagnosis of Rotating Machines Based on Advanced
Signal Processing Methods and Data-Driven Artificial
Intelligence Techniques**

Dissertation

for the Degree of

Doctor of Philosophy

(Computer Engineering)

University of Ulsan

Alexander E. Prosvirin

February 2021

**Fault Diagnosis of Rotating Machines Based on Advanced
Signal Processing Methods and Data-Driven Artificial
Intelligence Techniques**

Supervisor: Professor Jong-Myon Kim

A Dissertation

Submitted to
the Graduate School of the University of Ulsan
In Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy
(Computer Engineering)

by

Alexander E. Prosvirin

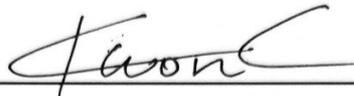
Department of Computer Engineering
University of Ulsan, Korea
February 2021

©2021 – Alexander E. Prosvirin

All rights reserved

Fault Diagnosis of Rotating Machines Based on Advanced Signal Processing Methods and Data-Driven Artificial Intelligence Techniques

This certifies that the dissertation
of Alexander E. Prosvirin is approved.



Prof. Kwon, Yung-Keun, Committee Chair



Prof. Kim, Jong-Myon, Advisor



Prof. Yoon, Seok-Hoon, Committee Member



Dr. Chung, Hyun-Sang, Committee Member



Dr. Kim, Ki-Hong, Committee Member

Department of Computer Engineering
University of Ulsan, Republic of Korea

February 2021

VITA

Alexander E. Prosvirin was born in Moscow, USSR, in 1991. He received his degree of Engineer at specialty “Control and Informatics in Technical Systems” from the Moscow State University of Mechanical Engineering “MAMI” (MSUME-MAMI), now called Moscow Polytechnic University, Moscow, Russian Federation, in 2013. Since September 2016, he has been working towards his Ph. D. Degree in the Department of Electric/Electronic and Computer Engineering, University of Ulsan, South Korea, under the supervision of Professor Jong-Myon Kim. His Ph. D. dissertation is based on exploring Advanced Signal Processing and Data-Driven Artificial Intelligence techniques for fault diagnosis of rotating machinery. His current research interests include artificial intelligence in general and deep learning particularly, signal processing, fault feature extraction, detection, and identification.

DEDICATION

This work is dedicated to my parents for their unconditional love and endless support throughout my studies and thereafter.

Thanks, mom and dad for always believing in me and for encouraging me to pursue my dreams.

ACKNOWLEDGMENTS

This has been a long journey of assorted movements. The movements of distresses and soothes, desolation and hope, joy and apprehension, a sense of accomplishment, and at the same time endeavoring for progress. Eventually, all the hard work pays off. First and foremost, I would like to express my profound gratitude to my advisor, Professor Jong Myon Kim, for his kind supervision and continuous support for my Ph.D. study and related research work, for his patience, motivation, and sharing of immense knowledge. Throughout my Ph.D., Professor Kim has provided me the kind support both in academic and non-academic matters. Without his kind help the vision of both this research work and the arc of the research direction would be hazy or to be apt short-sighted. I am indebted to him for his guidance, trust, and generous support in this long journey of my studies. Besides my advisor, I extend my gratitude to my Ph.D. supervisory committee for taking out the time to invest in the improvement of my dissertation.

I am grateful to all my fellow laboratory members in the Ulsan Industrial Artificial Intelligence Laboratory (UIAI Lab), University of Ulsan, for the thought-provoking discussions, and for all the fun we have had in the last four and half years. I am thankful to, no particular order, Dr. Piltan Farzin, Dr. M. M. Manjurul Islam, Jaeyoung Kim, Inkyu Jeong, and other laboratory mates for their valuable feedback and recommendations throughout the research seminars and for always being available when asked for help.

I would also like to gratefully acknowledge the gracious financial support of the BK21 plus program, the Korea Institute of Energy Technology Evaluation and Planning (KETEP), the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea, the Ministry of Science, ICT and Future Planning, and the National Research Foundation of Korea (NRF).

Most importantly, I would like to express my profound gratitude and appreciation to my dear parents and family members who have always been a source of constant support, love, and wholehearted care in every situation of my life. They have focused all their attention and resources on me and understandably all their hopes. Without them, this Ph.D. would have been impossible. I wish to be able to meet up to their hopes soon.

I am thankful to my friends here in Korea, back home, and spread throughout the world for their support and well wishes. I am grateful to the many nice Korean men and women who turned Ulsan into a second home for me.

Alexander E. Prosvirin

University of Ulsan, Ulsan, Republic of Korea

February 2021

ABSTRACT**Fault Diagnosis of Rotating Machines Based on Advanced
Signal Processing Methods and Data-Driven Artificial
Intelligence Techniques**

by

Alexander E. Prosvirin**Supervisor: Prof. Jong-Myon Kim**

Submitted in partial fulfillment of the requirements for the Degree of Doctor
of Philosophy (Computer Engineering)

February 2021

Robust condition assessment and predictive maintenance of industrial equipment is a valuable and ever-growing research field. The core of this field is relying on accurate condition monitoring, fault detection, and fault identification in industrial equipment. Fault detection and identification can be applied to a huge variety of industrial components such as rotating machines, blade turbines, industrial robot manipulators, motors, pipelines, gearboxes, spherical tanks, pressure vessels, etc. This work focuses on problems of diagnosing blade rub-impact faults of various intensities in turbines. Rubbing faults, also called in literature as rub-impact faults, represent a class of highly nonlinear and nonstationary mechanical faults that may appear in different rotating machines in general, and turbomachinery, specifically. These faults are caused by the interaction between rotating and stationary parts of rotating machines. The rubbing phenomenon may appear in the industrial system not only as a single isolated fault but as the consequence of the presence of other mechanical faults (mostly, rotor-related faults), which creates challenges for fault diagnosis and analysis. If the presence of the rubbing process in the rotor system is not detected and its severity is not identified in time, it can lead to the failure of the whole mechanical system (i.e., turbine). The failure of the system results in unexpected downtimes and great economic losses. Moreover, any mechanical fault appearing in industrial equipment can

become a threat to the safety of the people working in the facility. The condition monitoring of turbines and the diagnosis of blade rub-impact faults can be achieved using different types of techniques. Therefore, in this dissertation, the approaches based on analyzing vibration signals using advanced signal processing and data-driven artificial intelligence algorithms are developed that can accurately identify and diagnose blade rub-impact faults of various severity levels.

Due to the cost-effectiveness and reliability requirements, fault diagnosis of the industrial equipment using vibration acceleration signals has been a key direction of research over the past several decades that gave us a variety of methods. However, due to the complex nonstationary and nonlinear nature of rubbing faults, the application of the conventional time- and frequency-domain signal analysis techniques for examining and extracting discriminative fault features of rubbing processes appeared to be not effective. Furthermore, another condition that should be considered when diagnosing rub-impact faults is the fact that the rubbing phenomenon can be induced by other mechanical faults appearing in the system, such as bearing failures, shaft imbalance, etc. In these cases, it is favorable to apply the time-frequency analysis (TFA) methods for vibration signal decomposition. Moreover, it is crucial to improve the quality of the rubbing signal decomposition as well as determine, which frequency components of the signal (i.e., sub-bands) are exactly related to the mechanical fault being investigated. For this, the hybrid feature extraction based on the advanced TFA decomposition technique called ensemble empirical mode decomposition (EEMD) in conjunction with the autonomous fusion of optimal intrinsic mode functions (IMFs) containing harmonics of blade rub-impact is proposed in this dissertation. This dissertation also introduces the ratio between degree-of-frequency presence (DFP) of blade rub-impact fault-related harmonics and Kullback-Leibler divergence (KLD) as a specific criterion for autonomous selection of subset containing optimal IMFs related to rubbing faults. The experimental results indicated that the signals reconstructed using selected optimal IMFs related to the rubbing process contained less high-frequency noise and excluded the frequency harmonics not related to rub-impact faults. The hybrid feature pool extracted from the reconstructed signals and used as the input to one-against-all multiclass support vector machines (OAAMCSVM) classifier demonstrated high discriminative properties not only for detecting but also for diagnosing coupling rotor imbalance and blade rub-impact faults of various intensity levels.

Intrinsic mode fusion appeared to be an important procedure that should be used for extracting valuable information about the mechanical faults being investigated when the iterative TFA decomposition techniques, such as empirical mode decomposition (EMD) and EEMD, are used for analyzing complex nonstationary signals. Despite DFP/KLD-based optimal IMF selection algorithm demonstrated its high capabilities of selecting intrinsic modes that contain valuable information about rubbing processes for vibration signal reconstruction, the application of this

method may lead to the selection of non-informative intrinsic components along with informative ones in some extremal cases when computing the objective function value. Therefore, in this thesis, the originally proposed DFP/KLD-based optimal intrinsic mode fusion algorithm has been improved with the adaptive thresholding and objective function normalization techniques to increase the robustness of this selection method and reduce the possibility of selecting non-informative IMF components. The experimental results demonstrated that the vibration signals containing coupling rotor imbalance and blade rubbing faults reconstructed using the set of optimal IMFs selected by the improved selection algorithm include less high-frequency noise in comparison with the originally proposed selection framework while still preserving the valuable features (frequency harmonics) evidencing about rubbing process ongoing in the system. Furthermore, the results show the improvements in terms of fault classification accuracy in comparison with the original technique.

Machine learning (ML)-based condition monitoring and fault diagnosis of a rotary machine, such as turbines, are inevitable given the complexity of many problems. The operation of turbines is dependent on the health condition of its blades that operate in a hostile and high-stress environment. Moreover, due to harsh working conditions and precise requirements for clearance between stator and rotor blades, blade rubbing fault becomes a highly nonlinear type of mechanical faults in which several physical phenomena, such as vibration, friction, thermal effects, and stiffness are involved simultaneously. These factors affect and deteriorate the performance of conventional fault diagnosis models and make the fault diagnosis process more challenging that frequently leads to the incapability of the signal-based fault identification frameworks to maintain their high performance when the operating conditions of rotor systems, such as rotating speed and load, change. In this regard, the system-based data-driven technique from the family of artificial intelligence (AI) methods is presented for detection and identification of coupling rotor imbalance and blade rubbing faults. The data-driven system-based fault identification framework is developed based on approximation of the nonlinear function of the rotor system by deep undercomplete denoising autoencoder (DUDAE) and fault classification by a deep neural network (DNN). The contributions of this work are as follows. First, the proposed DUDAE learns the approximation of the nonlinear function of the rotor system by vibration signals collected under normal operating conditions, i.e., when neither rotor imbalance nor blade rub-impact faults present. Next, the residual signal is computed as the difference between the unknown vibration signal coming from the rotor system and its estimate provided by trained DUDAE. Finally, the DNN is applied to perform fault classification based on the generated residual signals. The experimental results demonstrated that the generated residual signals are sensitive to the rotor system degradation allowing for high fault classification accuracy when they are used as features for determining the state of the system. The obtained results of this study are compared with current state-of-the-art

deep learning-based fault diagnosis techniques, which demonstrates the advantages of the proposed model.

Multivariate signal analysis and data fusion from multiple sensors are important topics in the field of condition monitoring that can significantly increase the performance of fault diagnosis frameworks. The simultaneous analysis of multivariate signal allows for a precise investigation of different processes ongoing in the rotor systems. With the increase of computational complexity of TFA approaches when dealing with multivariate signals, the deep learning-based techniques that perform representation learning drew the attention of researchers and industrial specialists when addressing the problem of fault diagnosis using multivariate signals. In current work, a mechanism has been developed that can recognize and identify coupling rotor imbalance and blade rub-impact faults of various intensity levels by using multivariate vibration signals, i.e., the vibration signal collected by several channels of multiple sensors. To reduce the computational complexity and accelerate fault identification procedure, a fault diagnosis model is developed based on vibration signal resampling concerning fundamental frequency, envelope power spectra analysis of multivariate rub-impact fault signals, and a tiny multivariate-one-dimensional convolutional neural network (ModCNN). The contributions of this work are as follows. First, the multivariate vibration signals in the time domain are resampled with overlap using the fundamental frequency to ensure that these samples contain valuable information obtained during each revolution of the rotor. Second, the envelope power spectra of the resampled signals are computed to create discriminative patterns for the further representation learning task. Finally, ModCNN based on adaptive moment estimation optimization function (Adam) is proposed to extract local features from the resampled multivariate signals, hence reducing the overhead of feature extraction and selection. Adam optimization is computationally efficient and is well suited to problems with big datasets. The proposed framework has been evaluated by the vibration dataset collected using the testbed provided by UIAI Lab. and the obtained results of this study are compared with the current state-of-the-art machine and deep learning techniques used in the field of fault diagnosis of rotary machines, which demonstrate the superiority of the proposed model and of multivariate signal analysis.

Solutions from the field of control theory are widely used in industry for assessing the health condition of engineering systems because, in general, they are more robust and reliable in comparison with data-based (signal-based) techniques, since the operation of those algorithms is based on the system dynamics. For fault detection and diagnosis, first, these solutions require designing the observers which are used for improving the accuracy of system identification. One of the most frequently used observers in the industry is the linear-based observation technique. Linear-based observers have the advantage of the design and implementation simplicity along with

flexibility; however, they suffer the challenge of lack of robustness and reliability, especially, when they are applied to resolve the problems of nonlinear nature. The problems of linear observers can be generally addressed in two ways: 1) design the nonlinear-based observation technique or 2) improve the performance of linear observers using other types of algorithms, such as artificial intelligence-based techniques. Apart from the advantages of nonlinear-based observers such as robustness and reliability, designing these types of techniques is a complex problem. Thus, enhancing the performance of linear observer when applied to resolve a fault diagnosis problem of nonlinear nature by applying the artificial intelligence-based solution, and thus, introducing a hybrid fault identification technique, can be a suitable solution to this problem. In the current work, the hybrid approach for diagnosing blade rub-impact faults (a nonlinear mechanical fault) using a deep learning-based observation technique is introduced for addressing the problem mentioned above. A hybrid fault diagnosis model is developed based on a linear observation technique called autoregressive with eXogenous input Laguerre (ARX-Laguerre) proportional-integral observer (ARXLPIO) that is used for rub signal estimation and scalable deep neural network (S-DNN) which is utilized for improving the fault identification performance of ARXLPIO. The specific contributions of this work are as follows. First, the ARXLPIO is used to estimate the blade rub-impact signals using the principles of system identification and system estimation. Next, the S-DNN the architecture of which is autonomously designed utilizing the novel proposed algorithm is used for improving the fault identification performance of ARXLPIO when applied to nonlinear problems. The experimental results show that the proposed hybrid approach for diagnosing coupling shaft imbalance and blade rub-impact faults demonstrated high fault classification accuracy and outperformed the referenced methodologies used for the comparison. Furthermore, the simplicity of designing the proposed framework makes it favorable for usage in industrial applications.

Contents

VITA	viii
DEDICATION	ix
ACKNOWLEDGMENTS	x
ABSTRACT	xii
Contents	xvii
List of Figures	xxi
List of Tables	xxiv
List of Algorithms	xxv
Nomenclature	xxvi
Chapter 1 Introduction	1
1.1 Motivation	2
1.2 Thesis Outline	6
Part I Fault Diagnosis of Rotating Machines Based on Advanced Signal Processing Methods and Hybrid Feature Extraction	9
Chapter 2 Rub-Impact Fault Diagnosis Using an Effective IMF Selection Technique in Ensemble Empirical Mode Decomposition and Hybrid Feature Models .	10
2.1 Introduction	10
2.2 Proposed Rub-Impact Fault Feature Extraction Technique	14
2.2.1 Data Acquisition	15
2.2.2 Empirical Mode Decomposition and Ensemble Empirical Mode Decomposition	18
2.2.2.1 Empirical Mode Decomposition	18
2.2.2.2 Ensemble Empirical Mode Decomposition	19
2.2.3 IMF Selection Procedure for Rubbing Fault Diagnosis	21
2.2.4 Hybrid Feature Extraction and Configuration of Feature Set	25
2.3 Experimental Results and Discussion	28
2.3.1 Training and Testing Data Configuration	28
2.3.2 Validation of the Optimal IMF Subset Using the Proposed Approach	28
2.3.3 Performance Evaluation of the Proposed Rubbing Fault Feature Extraction Scheme with New IMF Selection Procedure	32

2.4 Conclusions.....	40
Chapter 3 An Improved Algorithm for Selecting IMF Components in Ensemble Empirical Mode Decomposition for Domain of Rub-Impact Fault Diagnosis.....	41
3.1 Introduction.....	41
3.2 Brief Background on the Original DFP/KLD-based Approach for Selecting Optimal IMFs in EEMD	44
3.3 Improved IMF Fusion Algorithm in EEMD for Rubbing Fault Diagnosis.....	46
3.3.1 The Drawbacks of the Original DFP/KLD-based Optimal IMF Selection Approach	46
3.3.2 Objective Function Normalization and Adaptive Thresholding Computation For Improving the Optimal IMF Selection Procedure.....	49
3.4 Experimental Results and Discussion.....	53
3.4.1 Analysis of Signal Denoising Properties.....	53
3.4.2 Fault Diagnosis Performance Analysis	59
3.5 Conclusions.....	61
Part II Fault Diagnosis of Rotating Machines Based on Data-Driven Artificial Intelligence Techniques	63
Chapter 4 Blade Rub-Impact Fault Identification Using Autoencoder-based Nonlinear Function Approximation and a Deep Neural Network.....	64
4.1 Introduction.....	64
4.2. Proposed Methodology.....	68
4.2.1. Signal Resampling	69
4.2.2. Deep Undercomplete Denoising Autoencoder (DUDAE)-based Nonlinear Function Approximation Of The Rotor System.....	70
4.2.3. Residual Signal Generation.....	73
4.2.4. Fault Identification Using Residual Signals and The DNN	74
4.3. Experimental Results and Discussion.....	76
4.3.1. Training, Validation, and Testing Data Configuration.....	76
4.3.2. Choosing the Noise Application Strategy in DUDAE	77
4.3.3. Training the Finalized DUDAE-DNN Model.....	78
4.3.4. Residual Signal Analysis.....	80
4.3.5. Fault Identification Performance.....	82
4.4. Conclusions.....	87

Chapter 5 Intelligent Rubbing Fault Identification Using Multivariate Signals and a Multivariate One-Dimensional Convolutional Neural Network	88
5.1 Introduction.....	88
5.2. Proposed Methodology.....	91
5.2.1. Multivariate Signal Processing and Pattern Preparation.....	91
5.2.1.1. Signal Resampling	91
5.2.1.2. Pattern Preparation.....	93
5.2.2. Rubbing Fault Identification using a Multivariate One Dimensional CNN.....	94
5.3. Experimental Results and Discussion.....	97
5.3.1. Training, Validation, and Testing Subsets Configuration.....	97
5.3.2. Training and Validation Process of the ModCNN	98
5.3.3. Fault Classification Performance Analysis	99
5.4 Conclusions.....	103
Chapter 6 Hybrid Rubbing Fault Identification Using an Artificial Intelligence-based Observation Technique	104
6.1 Introduction.....	104
6.2 Proposed Methodology.....	107
6.2.1. ARX-Laguerre PI-observer (ARXLPIO).....	108
6.2.2. Artificial Intelligence-based Approach for Blade Rub Fault Identification.....	111
6.2.2.1. Signal Resampling Process	111
6.2.2.2. Scalable Deep Neural Network (S-DNN) Design.....	113
6.3. Experimental Results and Discussion.....	117
6.3.1. Training, Validation, and Testing Subsets Configurations	117
6.3.2. Case Study of Building the S-DNN	117
6.3.3. Training and Validation of S-DNN Model	119
6.3.4 Fault Diagnosis Performance Analysis	120
6.4. Conclusions.....	126
Part III Summary and Future Work	128
Chapter 7 Summary of Contributions and Future Work.....	129
7.1 Introduction.....	129
7.2 Summary of Contributions	129

7.3 Future Work.....	133
Publications	136
References	139

List of Figures

Figure 2.1. A block diagram of the proposed fault diagnosis framework including informative IMF selection, fault feature extraction, and one-against-all multi-class support vector machine (OAAMCSVM) – based fault classification.	15
Figure 2.2. The self-designed test rig for rub-impact fault diagnosis.....	16
Figure 2.3. Changes in rubbing intensity validated using a thermal camera.	17
Figure 2.4. The feature space and trend curves formed using the statistical feature parameters extracted from the data samples of a single rotor imbalance and combined blade rub-impact and rotor imbalance faults.....	18
Figure 2.5. Decomposition of a one-second vibration signal containing shaft imbalance and intensive rubbing fault by (a) EMD and (b) EEMD approaches.	21
Figure 2.6. Evaluation of the IMFs by computing the DFP over the frequency of the interest and its harmonics.	22
Figure 2.7. Examples of envelope power spectra of (a) original vibration signal from class #1 (0.0 g), (b) reconstructed signal using selected IMFs for class #1 (0.0 g), (c) original vibration signal from class #10 (2.8 g), and (d) reconstructed signal using selected IMFs for class #10 (2.8 g).	31
Figure 2.8. Examples of envelope power spectra of reconstructed signals obtained using EEMD and the proposed IMF selection technique for classes (a) #1 (0.0 g), (b) #5 (1.6 g), (c)#8 (2.0 g), and (d) #10 (2.8 g), respectively.....	32
Figure 2.9. The boxplots demonstrating the statistics of the FCA metric over 10 experiments.....	34
Figure 2.10. Confusion matrices for classification using (a) proposed, (b) SensIMF+HFM, (c) WPT+MSV, (d) EMD+MSV, and (e) DWT+TDSIG approaches. All results are presented as the average of 10 experiments.	36
Figure 2.11. Feature space projections for (a) proposed, (b) SensIMF+HFM, (c) WPT+MSV, (d) EMD+MSV, and (e) DWT+TDSIG approaches, respectively.....	37
Figure 3.1. The IMFs and their envelope power spectra obtained as the result of EEMD decomposition of the signal containing severe rub-impact fault.	44
Figure 3.2. Block diagram of the (a) original and (b) improved DFP/KLD-based IMF selection techniques for rub-impact fault diagnosis.	47
Figure 3.3. IMF selection results achieved by (a) the original objective values with arbitrary thresholding and (b) the normalized objective values using MinMax scaling with an adaptive threshold value.	53
Figure 3.4. Comparison of original time-domain vibration signals against reconstructed ones using both optimal intrinsic mode selection techniques and their envelope power spectra for the signals of (a) class #1, (b) class #5, (c) class #8, and (d) class #10, respectively.....	54
Figure 3.5. Envelope power spectra of signals reconstructed using both the original and improved intrinsic mode selection techniques for signal (a) class #1, (b) class #5, (c) class #8, and (d) class #10 for frequencies ranging from 0 Hz to 500 Hz.....	55
Figure 3.6. Average values of (a) rSNR and (b) rMSE computed for the reconstructed signals obtained by different IMF selection methods.....	56

Figure 3.7. Original time-domain vibration signals and the envelope power spectra of the reconstructed signals through original DFP/KLD-based IMF selection and the proposed one corresponding to: (a) sample #31 from #1, (b) sample #34 from #5 (c) sample #13 from #9 signal classes.....	58
Figure 3.8. The boxplots demonstrating the statistics of the FCA metric over 10 experiments.....	60
Figure 3.9. The confusion matrix for fault diagnosis results obtained by the SVM machine learning algorithm using the features extracted from signals processed by the proposed methodology. All the results are presented as the average of ten experiments.	60
Figure 3.10. Feature space projections for (a) Proposed + HFM and (b) Original + HFM, respectively.	61
Figure 4.1. The proposed framework for the assessment of the health state of the rotor system.	69
Figure 4.2. The vibration signal resampling process.	70
Figure 4.3. The training and validation loss curves obtained during 10 experimental trials by the proposed technique: (a) training loss curves of DUDAE, (b) validation loss curves of DUDAE, (c) training loss curves of DNN, and (d) validation loss curves of DNN, respectively.	79
Figure 4.4. The original, reconstructed, and residual signal examples corresponding to signal classes of (a) normal operating condition, (b) 1.5 g shaft imbalance condition, (c) 1.7 g shaft imbalance + slight blade-rub fault condition, (d) 2.4 g shaft imbalance + intensive blade rub-fault, and (e) 2.8 g shaft imbalance + severe blade rub fault, respectively.	81
Figure 4.5. The energy feature parameter used for characterizing the residual signals obtained by proposed methodology for the signal classes under normal operating condition, 1.5 g shaft imbalance condition, 1.7 g shaft imbalance + slight blade-rub fault condition, 2.4 g shaft imbalance + intensive blade rub-fault, and 2.8 g shaft imbalance + severe blade rub fault, respectively.	82
Figure 4.6. The boxplots demonstrating the statistics of the FCA metric over 10 experiments.	84
Figure 4.7. The confusion matrices obtained for (a) gnProposed, (b) doProposed, (c) ARXLPIO+DNN, and (d) RAW+DNN methods, and averaged over 10 experiments.	85
Figure 5.1. Proposed intelligent rub-impact fault diagnosis framework.	91
Figure 5.2. The vibration signal resampling procedure with a 15% overlap.	93
Figure 5.3. The envelope power spectra computed from the resampled multivariate vibration signal collected by multiple channels.	94
Figure 5.4. The architecture of the proposed ModCNN.	95
Figure 5.5. The scheme of splitting resampled dataset into training, validation, and testing subsets.	98
Figure 5.6. The (a) training loss, (b) validation loss, (c) training accuracy, and (d) validation accuracy convergence curves obtained during the 10 experimental trials.	99
Figure 5.7. The boxplots demonstrating the statistics of the FCA metric over 10 experiments.	101

Figure 5.8. The confusion matrices obtained for (a) proposed, (b) Kurt+ LeNet5, (c) EPS+CNN, and (d) MSCNN averaged over 10 experiments.	101
Figure 6.1. Proposed hybrid rub-impact fault diagnosis approach.	107
Figure 6.2. The estimated signal resampling process with 15% overlap.	112
Figure 6.3. The frequency power spectra of the resampled estimated signals corresponding to (a) the normal system condition and (b) the severe rubbing fault system condition..	113
Figure 6.4. Validation losses of candidate network architectures.	118
Figure 6.5. The (a) training loss, (b) validation loss, (c) training accuracy, and (d) validation accuracy convergence curves obtained during the 10 experimental trials. ...	120
Figure 6.6. The boxplots demonstrating the statistics of the FCA metric over 10 experiments. (change the legend)	122
Figure 6.7. The confusion matrices obtained for (a) proposed, (b) ARXLPIO+SSAE, (c) ARXLPIO+CNN, (d) RAW+S-DNN, and (e) ARXLPIO averaged over 10 experiments.	123
Figure 6.8. The error signals obtained by the ARXLPIO after signal estimation process.	124

List of Tables

Table 2.1. Data acquisition system specifications.	16
Table 2.2. Time- and frequency-domain statistical feature parameters.	27
Table 2.3. The example of objective values of the proposed optimal IMF selection metric calculated for signal samples collected from different signal classes.	29
Table 2.4. The subsets of optimal IMFs obtained for signal samples presented in Table 2.3.	29
Table 2.5. Experimental results.....	34
Table 3.1. DFP, KLD, and objective values obtained by the original IMF selection metric for the signal sample that contains a severe rubbing fault.....	48
Table 3.2. Original and normalized objective values obtained for the signal sample that contains a severe rubbing fault.	52
Table 3.3. Experimental results.....	59
Table 4.1. The architecture of the proposed DUDAE.....	72
Table 4.2. The architecture of the DNN used for fault identification.	76
Table 4.3. The minimum training and validation loss values obtained for different realizations of dropout and Gaussian noise.	78
Table 4.4. The experimental results averaged over 10 experiments.....	84
Table 5.1. The experimental results obtained for the proposed and referenced techniques.	100
Table 6.1. The example of lookup grid L	115
Table 6.2. The example of search grid G	116
Table 6.3. The Architecture of S-DNN selected using the proposed algorithm.	119
Table 6.4. The experimental results obtained for the proposed and referenced techniques.	121

List of Algorithms

Algorithm 2. 1. EMD Algorithm.....	19
Algorithm 2. 2. EEMD Algorithm.....	20

Nomenclature

AE	Autoencoder
AI	Artificial Intelligence
ANN	Artificial Neural Network
Adam	Adaptive Moment Estimation Optimizer
adTh	Adaptive Threshold
ARXLPIO	Autoregressive Exogenous Laguerre PIO
CNN	Convolutional Neural Network
CSL	Convolutional-Subsampling Layer
DE	Drive End
DL	Deep Learning
DAE	Denoising Autoencoder
DNN	Deep Neural Network
DFP	Degree-of-Frequency Presence
DUDAE	Deep Undercomplete Denoising Autoencoder
DWT	Discrete Wavelet Transform
EMD	Empirical Mode Decomposition
EPS	Envelope Power Spectra/Spectrum
EEMD	Ensemble Empirical Mode Decomposition
FCA	Fault Classification Accuracy
FCF	Frequency Component of FOI
FFT	Fast Fourier Transform
FN	False Negatives
FOI	Frequency-of-Interest
FP	False Positives
GAN	Generative Adversarial Network
GMM	Gaussian Mixture Model
HFM	Hybrid Feature Model
IMF	Intrinsic Mode Function
KLD	Kaulback Libeler Divergence
Kurt	Kurtogram
k-NN	k-Nearest Neighbor
LMD	Local Mean Decomposition
MC	Multiclass
ML	Machine Learning
ModCNN	Multivariate One-Dimensional Convolutional Neural Network

MSCNN	Multi-Sensor Convolutional Neural Network
MSV	Maximum Singular Values
NDE	Non-Drive End
nObj	Normalized Objective Function
NSTFT	Nonlinear Squeezing Time-Frequency Transform
OAA	One-Against-All
Obj	Objective Function
PF	Product Function
PDF	Probability Density Function
PIO	Proportional Integral Observer
PMIO	Proportional Multi Integral Observer
REB	Rolling-Element Bearing
ReLU	Rectified Linear Unit
RFC	Residual Frequency Component
RMS	Root Mean Square
RPM	Revolution per Minute
rMSE	reversed Mean Squared Error
rSNR	reversed Signal-to-Noise Ratio
S-DNN	Scalable Deep Neural Network
SELU	Scaled Exponential Linear Units
SNR	Signal-To-Noise Ratio
SRA	Square Root of Amplitude
Std	Standard Deviation
SVM	Support Vector Machines
TDSIG	Time-Domain Signal
TFA	Time-Frequency Analysis
TP	True Positives
t-SNE	t-Stochastic Neighbor Embedding
UIAI Lab.	Ulsan Industrial Artificial Intelligence Laboratory
var	Variance
WT	Wavelet Transform
WPT	Wavelet Packet Transform

Chapter 1

Introduction

As a definition of the fault in industrial engineering systems, the unwanted deviation of these systems from their standard operating conditions should be understood. For instance, the bearing faults [1], gearbox faults [2], and shaft misalignment [3] are examples of mechanical faults that frequently appear in rotary systems, i.e., rotating machines [4]. These and the other types of mechanical faults may affect the dynamics of a system and the operating conditions that will lead to the health degradation of the system components and, consequently, the whole system itself. Over the years, fault detection and identification in industrial engineering systems became an imperative field of research due to the increasing use of safety-critical systems. This led to the strengthening of safety and reliability requests subjected to the potential faults occurring in the industrial systems and anomalies that appear during the industrial processes. Therefore, the appropriate detection and identification of the potential anomalies and implementation of stable condition monitoring systems are of utmost importance. Such systems not only ensure the reliability and safety of the equipment and people present in the facility but also avoid the economic losses associated with the failure of the equipment [5].

Energy production is essential for the daily life of modern society and the operation and development of the industry. It is known that about 50% of the total energy produced is consumed by heavy industry. According to a study of German energy agency ‘Dana’, about 65% of the total energy consumed by the heavy industrial facilities in the European Community has been consumed by rotating machines [6]. To meet the needs of industry, maintain the energy-efficiency of power generation equipment, and minimize its emissions caused by malfunctions, the condition monitoring and fault diagnosis of machines used for energy production is essential. Gas/steam turbines and compressors are the types of equipment extensively utilized for energy production. Different types of faults that may appear in turbines during the operation have been defined, such as bearing faults, shaft imbalance, blade-related faults, or even coupling faults [7] (i.e., the cases when several mechanical faults are present in the rotor system simultaneously). The operation of these types of turbines is largely dependent on the condition of their blades and is closely associated with high-temperature environments, high load, and severe operating conditions. The blade-related faults are caused tremendous problems for industry in past decades and one of the prevalent reported blade-related faults is blade rubbing fault, also known as a rub-impact fault in literature, which represents 23% of total blade failures, especially in gas and steam turbines [8]. Vibration analysis [9] is the widely used condition monitoring approach for detecting and diagnosing rubbing faults of various intensity levels due to its relatively easy applicability in industry. However, the blade rub-impact faults are recognized as a type of highly nonlinear and nonstationary mechanical

faults [10, 11] so it is difficult to extract the relevant fault features related to rubbing processes using the conventional time- and frequency- domain methods, especially when the blade rub-impact is not a single isolated fault but associated to another mechanical fault present in a system (i.e., the consequence of another fault), such as a rotor imbalance. Therefore, to successfully perform the fault diagnosis of blade rub-impact faults of various intensity levels, different techniques, such as time-frequency analysis-based signal processing, artificial intelligence-based, and hybrid approaches have been used [12–14].

Based on this introduction, the thesis will focus on developing efficient fault diagnosis schemes for the main type of rotary machine used in electricity production – turbines. Specifically, the proposed fault diagnosis solutions for the turbines will address one of the most complex, crucial, and frequently appearing mechanical faults in this type of machine – blade rub-impact faults. The focus will be to develop fault diagnosis schemes that can not only precisely detect coupling rotor imbalance and rub-impact faults in turbines (i.e., differentiate between normal and abnormal operating conditions), but also accurately differentiate the severity (i.e., intensity) levels of these failures.

This chapter provides a concise description of the motivations behind and objectives of the conducted research work which is presented in the dissertation. The research work described in the dissertation has been published/submitted in peer-reviewed journals. Moreover, within each chapter, the necessary motivation, introduction, and background study of the given problem are provided. Nonetheless, the motivation behind this research work is given in Section 1.1 and Section 1.2 outlines the dissertation itself.

1.1 Motivation

Gas and steam turbines are the most important types of rotating machines used in the power generation field and according to the U.S. Energy Information Administration Agency, they comprised 61% of total utilities used for energy generation in 2018. These machines consist of two main parts: 1) stator and 2) rotor. The rotor part of the turbines contains the blades that are utilized for energy extraction in power plants [15]. The small clearance between the rotor blades and the stationary part of the turbines allows to minimize the air reluctance and increase the transmitting torque significantly. The blade-related failures in turbines have caused a lot of problems for maintenance operators in industrial facilities over the years [8]. The failure of even a single blade of the turbine in a powerplant can lead to severe damages, significant financial losses, or even be a threat to the health of the workers. To mitigate the problem, the maintenance of the turbines is compulsory. There are two main scenarios of the maintenance of industrial equipment. The first one is reactive maintenance, whereas the second one is called proactive maintenance [16]. In cases

of reactive maintenance, the damaged components of the equipment are repaired or replaced so that the machine may continue its operation. In this kind of maintenance, there are no preventive measures taken, and the problem is addressed only when it is appeared or is reported. Regarding proactive maintenance, preventive measures are taken in advance to mitigate machinery failure. Thus, proactive maintenance can be divided into two categories: 1) preventive maintenance and 2) predictive maintenance. Preventive maintenance is the type of maintenance of the machine that is performed on a routine basis. Regarding predictive maintenance, this includes the techniques developed to assess the health states of the components and predict when the actual maintenance for the machine will be required. In this way, unnecessary downtime of the machine can be avoided, the economic losses can be reduced, and the overall safety measures within the industry can be improved. So, fault diagnosis can be categorized as a type of predictive maintenance for the turbomachinery, as it detects and predicts its health states.

There are numerous research studies present that have been conducted to develop blade fault diagnosis techniques for rotating machines, including turbines. The conventional fault diagnosis framework consists of three main steps: 1) signal acquisition, 2) signal analysis and fault feature extraction, and 3) decision making.

For the first and second steps, these techniques include vibration signal analysis [9], acoustic emission signal analysis [17], pressure analysis [18], and temperature analysis [19]. However, due to the informativity and the practicality of the application in the real industrial field, the vibration analysis drew its attention as the most widely used method for rub-impact fault diagnosis in turbines. When the rotor-to-stator rub occurs in the system, the rub-impact signal is injected into the collected vibration signal of the system. The key to detecting and diagnosing blade rub-impact faults of various intensities is to extract valuable and discriminative fault information from the vibration signal. Since the blade rub-impact signal is recognized as a highly nonlinear and nonstationary signal, the conventional spectral analysis techniques may result in false information due to limitations of the Fourier transform [20]. To cope with this issue, in recent years, many studies for rub-impact fault diagnosis are employing more informative, but more complex signal processing and feature extraction methods that are called time-frequency analysis (TFA) techniques. These methods mainly include the various types of wavelet transform (WT) [21] and iterative TFA methods such as empirical mode decomposition (EMD) [22], and its derivatives [23–26]. Despite the capabilities of extracting valuable information from nonstationary signals, these techniques have challenges in practical applications. First, the results of wavelet transformation strongly rely on the choice of the wavelet basis function that should correlate well with the signal characteristics [27]. Second, the wavelet transform suffers the problems of border distortion and energy leakage [28] that lead to difficulties in interpreting the transformation results. These

drawbacks increased the number of applications of iterative TFA techniques when addressing the problems of nonlinear and nonstationary signal analysis, such as one containing blade rub-impact faults. Regarding iterative TFA, such as EMD and its derivatives, these methods decompose the vibration signals into series of oscillating modes; however, it appeared that the cardinality of those extracted oscillating modes is large and not all of them are equally important for diagnosing the specific mechanical faults, as well as some of them can be highly contaminated by noise which leads to the problems when trying to interpret the physical meaning of each of the extracted components and extracting valuable features which causes the deterioration of the performance of fault diagnosis models. Furthermore, the common drawback of the TFA techniques is that they require human expertise to identify the health state of the equipment and there exists a lack of actual feature parameters that allow to confidently identify the blade rub-impact fault presented in the rotor system. For the third step, decision making, the traditional techniques for mechanical faults in rotating machines in general and blade rub-impact faults, specifically, mainly utilize the conventional machine learning algorithms; for instance k-nearest neighbors (k-NNs), support vector machines (SVMs), and artificial neural networks (ANNs) [29–31]. Such algorithms are shallow in nature and are inefficient when dealing with the complex signals or noninformative or even false features that are extracted from the signals and are used for making a decision about the state of the system. In such cases, one way is to adapt suitable signal processing techniques and design an appropriate feature model that can be used to extract the fault signatures of the rubbing itself in the system or can detect the existence of coupling mechanical faults that are causing the interaction between rotor and stator. Thus, the work related to diagnosing rubbing faults by the conventional fault diagnosis scenario is presented to explore two key parts: (1) adaptation and improvement of a suitable advanced signal processing technique that can be useful for discovering blade rub-impact fault signatures; (2) designing the hybrid feature model for characterizing these fault signatures of rubbing faults and its evaluation against other feature models proposed for diagnosing the same type of fault.

On contrary to the conventional mechanical fault identification frameworks, the other solution is to develop a fault diagnosis model based on deep neural architectures rather than shallow classifiers that can automatically mine and analyze salient information and can learn the specific high-level features related to the behavior of the system. In recent years deep learning is making its mark in many fields including image processing [32, 33], speech recognition [34], natural language [35] and medical image processing [36], data augmentation [37], feature extraction [38], feature fusion [39], etc. These types of deep architectures have the capability of extracting meaningful information through their specific structures and by using nonlinear transformation in the hidden layers from the provided input data representations, which can be further used as features for decision-making purposes. Furthermore, the recent advances in deep

learning allow researchers and engineers to build end-to-end learning models that autonomously can solve complex problems that involve several steps, such as feature extraction, feature generation, fault identification, multivariate signal-based fault diagnosis, etc. Therefore, in this dissertation condition monitoring and fault diagnosis of rotor systems using deep neural architectures are also carried out. The specific key points investigated are as follows: (1) development of an end-to-end approach that includes discriminative feature generation using nonlinear function approximation of the rotor system and decision making about the state of the system; (2) autonomous fault feature extraction and diagnosis using multivariate vibration signals to resolve the complex problem of blade rub-impact fault identification.

Apart from signal processing- and artificial intelligence-based techniques used for diagnosing faults in complex engineering systems, the control theory algorithms are widely used in industrial scenarios due to their reliability because they are designed based on the actual dynamics of the target system. However, since the design of nonlinear control theory-based algorithms for diagnosing nonlinear faults (such as rub-impact fault) is a complex task with several drawbacks such as a lack of flexibility, one of the common solutions is to design a linear control algorithm first and then improve its fault diagnosis performance. For this, the advantages of methods from other expert fields can be utilized, and hence, a hybrid fault diagnosis approach can be introduced. To explore this scenario, in this dissertation the work related to designing the hybrid technique for diagnosing blade rub-impact faults is carried out. The key points of this work are presented as follows : (1) development of flexible hybrid approach where the fault diagnosis performance of linear control theory-based algorithm is improved using the ideas of artificial intelligence (specifically, deep neural networks) to achieve high classification accuracy when applied to nonlinear mechanical fault problem; and (2) proposing a novel algorithm for designing the specific scalable architecture of deep neural network in an autonomous way to best match the input data complexity and achieving the best classification accuracy on the specific data.

Thus, the objective of this dissertation is to explore two key parts: (1) adaptation and improvement of advanced TFA technique that is suitable for analyzing blade rub-impact fault signals in complex experimental scenarios, including the presence of coupling fault; (2) development of a suitable representation learning, i.e., deep-learning-based and hybrid, solutions that are capable of operating in an industrial environment and the performance exploration of the existing deep networks when applied for diagnosing rubbing faults of various intensity levels which are induced by rotor imbalance faults.

1.2 Thesis Outline

The dissertation is composed of three parts, excluding Chapter 1. Blade rub-impact fault diagnosis using advanced signal processing techniques and hybrid feature extraction is presented in Part I, i.e., Chapters 2 and 3. Part II is about the data-driven artificial intelligence- and hybrid-based methods for fault diagnosis of rubbing faults, i.e., Chapters 4, 5, and 6. Finally, the summary of work, conclusion, and future work is presented in Part III, i.e., Chapter 7.

In Chapter 2 an efficient optimal IMF selection technique in an advanced signal processing approach called ensemble empirical mode decomposition (EEMD) and hybrid feature models are presented for diagnosing coupling rotor imbalance and blade rub-impact fault. The predecessor of EEMD, EMD, is a powerful and convenient algorithm that allows for the analysis of complex and nonlinear multi-component signals, such as rub-impact fault signals, by decomposing them into a finite number of intrinsic mode functions (IMFs); however, this algorithm suffers from the problem called mode-mixing as well as the lack of understanding how to select the optimal intrinsic mode components for efficient fault feature extraction. To address these issues, this chapter describes the application of EEMD for rubbing fault signal decomposition, provides the comparison of the decomposition results obtained by EEMD and by the conventional EMD, and introduces a new efficient criterion for selecting optimal intrinsic modes containing the valuable rub-impact faults signatures based on the degree-of-frequency presence ratio (DFP) and Kullback-Leibler Divergence (KLD). Once the set of optimal IMF components is created, they are used for partial vibration signal reconstruction and creation of a hybrid feature pool that in conjunction with one-against-all multi-class support vector machines (OAAMCSVM) classifier is used for diagnosing rub-impact faults induced by rotor imbalance mechanical fault. The effectiveness of the proposed fault diagnosis framework was tested on the vibration data collected using a rubbing fault simulation test rig provided by UIAI Lab.

Chapter 3 is about the improved algorithm for selecting optimal IMF components in EEMD in the domain of rub-impact fault diagnosis. Despite the efficiency of the selection algorithm presented in Chapter 2 that is used for fusing intrinsic mode components containing valuable information about rubbing faults, this original algorithm appeared to have some drawbacks that in extreme cases may cause the selection of non-informative intrinsic modes. This chapter provides a discussion on these drawbacks as well as introduces the solution to these issues that includes two steps. Specifically, the first step is the normalization of objective function values to cope with the extremal cases that appear while computing the objective values for intrinsic components. The second step is represented by an adaptive thresholding technique to enhance the stability of the optimal IMF selection procedure. The analysis of the vibration rub-impact fault

signals collected using the simulation testbed designed by UIAI Lab demonstrates that the signals reconstructed using the optimal IMFs selected by the improved algorithm contain less high-frequency noise in comparison with the original technique presented in Chapter 2, while preserving the frequency components that are recognized as important ones for the rubbing fault diagnosis.

Chapter 4 proposes a data-driven fault diagnosis framework that is represented as the end-to-end deep learning-based pipeline. Due to the nonlinear behavior of rotor systems with blade rub-impact fault presence and nonstationarity of the vibration signals collected from those systems, there exist problems of applying statistics-based feature extraction techniques without complex signal preprocessing. It is understandable that the values of statistical feature parameters mostly depend on the amplitudes of the signals; however, the amplitude itself is affected by various internal and external disturbances of the rotor system such as environmental noise, physical processes ongoing in the operating machine, or the occurrence of mechanical faults. These issues cause the challenge of fault diagnosis performance deterioration when the approach successfully applied to one rotor system shows degraded performance on another rotor system with a changed environment even when it is applied to identify the same type of mechanical fault. One of the possible solutions to these challenges can be switching from straightforward signal-based conventional fault diagnosis pipelines with feature extraction to system-based techniques that in general are more stable to the external disturbances. Therefore, in this chapter, the data-driven fault diagnosis solution that includes the artificial intelligence-based system identification concept is proposed and its capabilities of estimating the nonlinear function of the rotor system are investigated to address the challenges discussed above.

Chapter 5 addresses a problem of condition monitoring that involves multivariate vibration signals. Since blade rubbing faults in conjunction with rotor imbalance faults are a highly nonlinear type of faults, the complex time-frequency domain signal processing methods are needed to discover incipient fault features; however, there exist challenges of application of these advanced signal processing techniques, especially iterative ones, when the multivariate signals are used to collect the information of the system. This includes the problem of selecting the most informative time sequences collected by different sensors or finding an appropriate way to combine the information accumulated by multiple sensors for fault diagnosis purposes, including the creation of a combined feature pool. Therefore, the data-driven solution for diagnosing blade rub-impact faults of different intensity levels using multivariate signals and the deep learning method is presented to address the aforementioned issues.

Chapter 6 introduces the hybrid data-driven fault identification technique that is represented as a conjunction of control theory-based and artificial intelligence-based approaches. Various methods used for fault identification face limitations in their applicability in real industrial

scenarios due to different issues such as computational complexity, the complicatedness of the design process, the lack of robustness and generalization, and the need for high-level human expertise when applying the designed approach to various systems. One of the possible solutions that can be suitable for addressing the problems mentioned above is to design hybrid adaptive fault diagnosis techniques that are working based on the dynamics of a system. Therefore, in this chapter, the hybrid data-driven fault diagnosis approach using the deep learning-based observation technique is introduced to propose a suitable solution that can be used in industrial applications.

Finally, Chapter 7 concludes the thesis, presents a summary of the contributions, and discusses future work.

Part I
**Fault Diagnosis of Rotating Machines Based on
Advanced Signal Processing Methods and Hybrid
Feature Extraction**

Chapter 2

Rub-Impact Fault Diagnosis Using an Effective IMF Selection Technique in Ensemble Empirical Mode Decomposition and Hybrid Feature Models

2.1 Introduction

Rotating machines such as turbines are widely used for power generation and usually operate under severe operating conditions characterized by high temperatures and high rotational speeds. The goal of the turbine design process is to maintain a small clearance between the rotor blades and the stator to increase torque and reduce air reluctance. Rubbing occurs when the turbine blades interact with the stationary parts, either due to blade expansion resulting from high temperatures or due to faults such as misalignment of the shaft and self-excited vibrations [40]. If faults are not detected at an early stage, rubbing can cause excessive damage to the rotating machine, significantly increasing the maintenance cost. Therefore, the purpose of the rub-impact fault diagnosis is to extract intrinsic information about rub impact and apply the extracted features to a classifier to determine rubbing faults at different intensities.

Rubbing faults are recognized as highly complex non-linear and non-stationary faults [41], [42], which cause a large number of transients to appear in the signal. This makes it difficult to utilize traditional signal processing techniques for feature extraction such as time-domain and frequency-domain fast Fourier transform (FFT) analysis, which cannot efficiently detect transient phenomena in a non-stationary rub-impact signal due to the inherent constraints of these methods. Therefore, the features extracted from rubbing signals by these methods may not be informative and may not reflect the actual rubbing processes but rather the qualities of a particular test rig and testing environment. In general, the quality of such features is not high enough to differentiate the rub-impact faults of different intensities.

To overcome the nonlinearity and nonstationarity constraints in rubbing fault feature extraction and diagnosis, many techniques based on time-frequency domain analysis (TFA) have been introduced over the last decade. Yanjun et al. [43] proposed wavelet packet eigenvalue calculation as a feature extraction technique for rubbing fault diagnosis. In this method, wavelet packet analysis is used to improve the time-frequency resolution of the vibration signal. In addition, the maximal singular values (MSVs) of the coefficient matrix obtained from wavelet packet decomposition are used as feature vectors for a support vector machine (SVM) classifier. Though the features proposed by these authors demonstrated good accuracy, the method may not be appropriate for early-stage rubbing fault diagnosis because the normal system conditions and

different rubbing intensities were not considered in their experiments. Zhihao et al. [44] suggested the use of discrete wavelet transformation (DWT) and time-domain signal extraction at level 3 of decomposition as a feature extraction approach for identifying rubbing faults. To classify signal samples, this extracted datum of the third-level was used as the input to the SVM model. Although the authors claim that the rate of correct rubbing fault identification was more than 92%, this study did not consider rub-impact faults of different intensities and did not suggest any techniques for an adaptive level or sub-band selection to ensure that the presence of rubbing fault in the selected datum is significant. Moreover, the use of time-domain data as a feature vector for SVM may not be suitable for situations when the acquired signals include a large number of time points, even when these signals are supplemented by down-sampling properties of wavelet transformation. Roy et al. conducted a study on the efficacy of various wavelet mother functions and signal-filtering techniques for rub-impact fault detection [45]. This study emphasized the importance of finding the proper combination of wavelet mother functions and filtering techniques, both of which strongly affect the accuracy of rub-impact fault detection. Similar limitations of the wavelet transform were reported by other researchers [46]. It is usually not easy to choose the optimal wavelet function for a complex and nonlinear signal, and since the process of determining a proper wavelet function requires a series of experiments, one can conclude that some subjectivity is present in this process [47, 48]. Deng et al. [26] introduced a combination of local mean decomposition (LMD) and Teager energy kurtosis as a feature extraction technique for rub-impact fault diagnosis. This approach allowed the extraction of numerical features from faulty signals and was capable of distinguishing rub-impact faults. However, it did not consider various rubbing intensities. Also, the Teager energy kurtosis values for the sixth and seventh product functions (PFs), which corresponded to the $1/3X$ fractional frequencies in their study, were not distinguishable between normal and fault conditions even though these fractional frequencies are usually recognized as critical indicators of the rubbing process [49, 50]. Yang et al. [24] proposed an ensemble LMD that improves the ability of LMD to decompose rub-impact signals by calculating the ensemble mean of extracted PFs. This study mostly contributed to the field of signal processing in the areas of rubbing fault detection and the observation of rubbing phenomenon over the frequency spectrum of the extracted PFs; no numerical features that could be used to differentiate between types and intensities of rubbing faults were considered in the experimental part. Another study [51] proposed the nonlinear squeezing time-frequency transform (NSTFT) as a TFA approach for rub-impact detection. This study simulated various rub-impact models and analyzed them using the proposed method, providing detailed information about frequency harmonics and the behavior of instantaneous frequencies when a rubbing fault occurs in a system. However, the focus of the study was the detection of rubbing faults, and no experiments were performed on rubbing faults with varying intensities.

Recently, a state-of-the-art TFA approach called empirical mode decomposition (EMD), introduced by Huang et al. [22], has become widely used for fault diagnosis in rotating machinery such as rotors, bearings, and gears [52]. Moreover, EMD appeared to be an efficient decomposition technique that can be used in other applications [53–55]. EMD itself is a very powerful and convenient algorithm that allows for the analysis of complex and non-linear multi-component signals by decomposing them into a finite number of intrinsic mode functions (IMFs). One of the most important advantages of EMD over its counterparts is that it is an entirely data-driven and self-adaptive signal decomposition technique in which each IMF corresponds to a natural oscillatory component of the original signal and can be treated as a specific frequency sub-band. Due to these advantages, EMD has been successfully used to extract and observe discriminative information from rub-impact fault signals in several studies [56–59]. These studies provide a good insight into the capability of EMD and its IMFs to extract rub-impact information from the vibration signal of rotor systems. However, these studies did not provide any insights regarding the fault feature extraction, while mostly focusing on the visual observation of rubbing faults. Yibo et al. [60] proposed calculating the MSVs from each of the extracted IMFs to create feature vectors for classifying rub-impact fault models using SVM. This feature extraction approach allows differentiation between different rub-impact fault types; however, the efficacy of the proposed features in differentiating normal system conditions when there is no rubbing and various rubbing intensities was not investigated. This approach also ignores the process of selecting meaningful IMFs.

Despite its advantages over its main counterpart, wavelet family decomposition, the traditional EMD algorithm has shortcomings that may cause problems when it is used for feature extraction to differentiate machinery faults. Specifically, the conventional EMD technique can result in abnormalities such as the presence of either disparate oscillations in one IMF or very similar oscillations in different IMFs. This is called the mode-mixing problem and makes it difficult to obtain a clear physical interpretation of each IMF, which is essential for fault feature extraction and diagnosis. Several variants of EMD have therefore been introduced to resolve the mode-mixing problem and improve the traditional approach. The breakthrough in EMD improvement was a method called ensemble empirical mode decomposition (EEMD) [23, 52]. This method is reported to successfully resolve the problem of mode-mixing by performing EMD over an ensemble of the signal with the addition of white Gaussian noise. EEMD decomposition has already been successfully used for rolling-element bearing [61, 62] and rub-impact fault diagnosis [63]. Regarding rub-impact faults, Lei et al. [63] utilized EEMD to diagnose rubbing faults in a power generator and early rub-impact faults in a heavy oil catalytic cracking machine set. These experimental results demonstrated the advantages of EEMD in extracting fault characteristics, even for incipient faults in rotating machinery.

Although EEMD successfully resolved the problem of mode-mixing, one issue that makes it difficult to utilize this approach for fault diagnosis is the adaptive determination of meaningful IMFs that contain specific and valuable information about the mechanical fault. As an EMD-based method, EEMD yields a finite set of IMFs after completing the decomposition process, and the cardinality of the extracted components is very large [64, 65]. In practice, however, the extracted IMFs are not all equally important for fault feature extraction, and they can be either noise-dominant or signal-dominant. Therefore, it is necessary to select a set of discriminative IMFs in which each IMF is highly useful for fault feature extraction and carries essential information about the mechanical fault. Yi et al. [66] proposed the use of steady-state indexes for condition monitoring of railway axle bearings, where they also addressed the problem of selecting the meaningful IMF components obtained by EEMD by picking the single mode with the largest energy. This study demonstrates that the introduced steady-state indices extracted from the single selected IMF are robust features for axle bearing condition monitoring. However, the absence of the adaptive thresholding approach while selecting IMFs based on the energy might be an issue when applying this method for rubbing fault problem where the relation “energy-content” for the extracted modes might not be that well-observable as in the investigated bearing faults.

During this literature review, it was observed that the main directions of current studies are the detection of rubbing faults using different TFA methods, the representation of rubbing faults using various forms of frequency spectra, and the investigation of rubbing itself as a physical phenomenon. Almost no studies have focused on diagnosing rubbing faults of various intensities, and very few actual feature extraction models with numerical values that can be directly applied to classification techniques for rub-impact fault diagnosis have been proposed. Therefore, this chapter proposes a reliable rub-impact fault feature extraction approach that combines EEMD with a new IMF selection procedure and uses the selected components to reconstruct the signal and perform hybrid feature extraction. The signal reconstructed using these selected informative IMFs contains less noise and clear rubbing fault frequency components, which makes the rubbing process easily observable. The proposed hybrid feature model for fault feature extraction allows for the examination of rubbing phenomena in both time and frequency domains and produces well-separable features that can be used for fault diagnosis. It is important to mention, that there exist other metrics that can be used to quantify various repetitive transients and select informative portions (or sub-bands) of a signal, such as a kurtosis and its modification - a spectral kurtosis. These metrics were used in combination with other TFA decomposition techniques; however, these metrics appeared to be not exactly proportional to the degree of defectiveness of the system and the result highly depends on the frequency resolution assigned before the decomposition process [67, 68].

The main contributions of this chapter are as follows:

1. This chapter presents a new informative IMF selection procedure that can be used to select the modes obtained by EEMD in rubbing fault analysis. The proposed method includes a new quality criterion for mode evaluation that combines the degree-of-frequency presence (DFP) of rub impact and the Kullback-Leibler divergence (KLD), a statistical similarity metric. An adaptive selection technique then utilizes a thresholding approach and the aforementioned criterion to adaptively select the most valuable IMFs. The selected informative signal-dominant IMFs carry intrinsic information about the important harmonics of rub-impact faults.
2. Since the selected IMFs are highly effective at detecting rub-impact phenomena, this chapter then extracts features for rub-impact fault diagnosis from the signal reconstructed using the selected components. Thus, a hybrid feature model that well represents rub-impact fault conditions is presented in this study. Our hybrid feature model consists of four features directly extracted from the reconstructed signal in the time domain and three features extracted from the envelope power spectrum of this signal. This hybrid set of features is highly effective for representing each rub-impact fault condition, so these features are further used in a classifier for diagnosing rubbing faults with various intensities.

The remaining sections of this chapter are structured as follows. Section 2.2 presents the proposed rub-impact fault feature extraction methodology, including the adaptive IMF selection procedure and hybrid feature model for diagnosing rubbing faults of various intensities. Section 2.3 provides experimental validation of the proposed methods. Finally, Section 2.4 contains the concluding remarks.

2.2 Proposed Rub-Impact Fault Feature Extraction Technique

The block diagram of the proposed rub-impact fault diagnosis framework is presented in Figure 2.1. The proposed approach consists of four essential steps: data acquisition, signal processing, feature extraction, and fault classification. As shown in Figure 1, after data acquisition, an unknown vibration signal is first decomposed into a finite number of oscillating components by EEMD. Then, the subset of IMFs that are signal-dominant and carry essential information about the rubbing faults is selected using the proposed adaptive selection procedure with our novel IMF evaluation metric. Next, time- and frequency-domain features are extracted from the signal reconstructed using the selected IMFs, which represents a clear rub-impact fault signal. This set of

hybrid features can provide sufficient insight into the rubbing process to classify rubbing faults of various intensities. The fault diagnosis procedure is completed by means of OAAMCSVM classifier [69, 70].

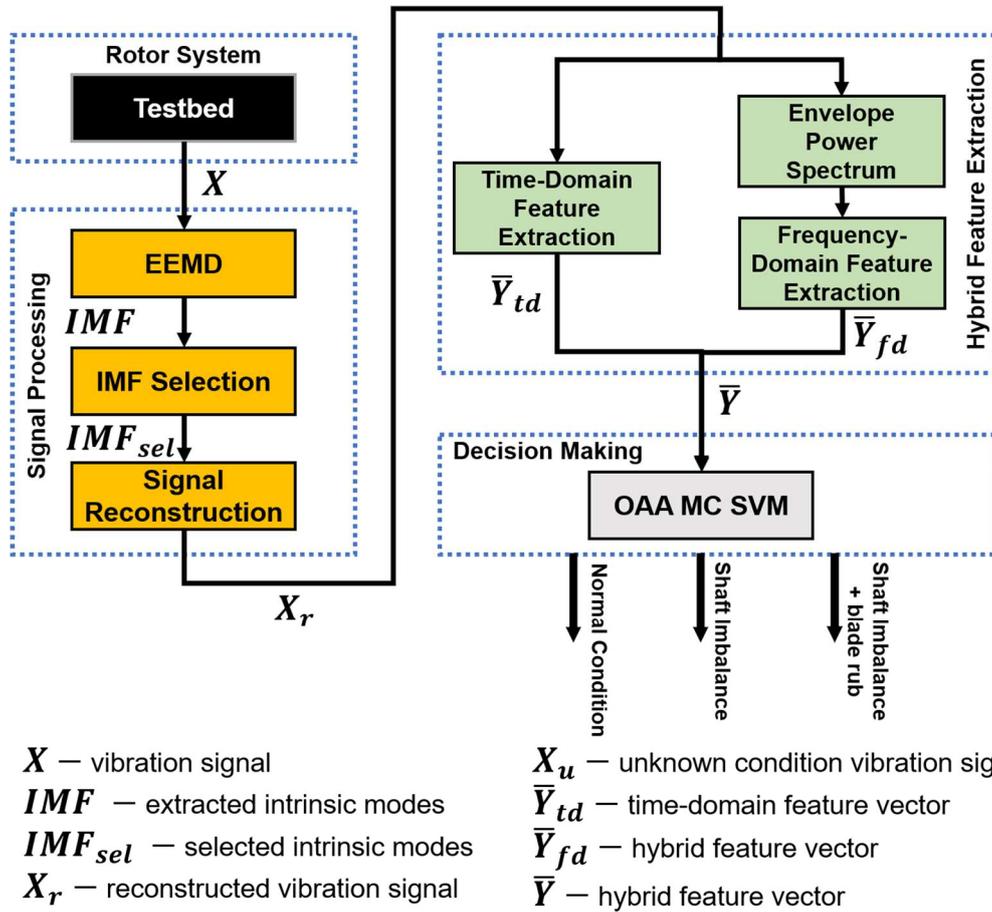


Figure 2.1. A block diagram of the proposed fault diagnosis framework including informative IMF selection, fault feature extraction, and one-against-all multi-class support vector machine (OAAMCSVM) – based fault classification.

2.2.1 Data Acquisition

Figure 2.2 shows the self-designed experimental testbed used to collect rubbing fault data by simulating coupling shaft imbalance and blade rub-impact faults of different intensities. Two sensors were installed at the drive end (DE) of the shaft and the non-drive end (NDE) of the shaft to continuously measure the vibrations of the rotor. Each sensor records the displacements of the 16-blade rotor in both the vertical and horizontal directions, using a different channel for each. Therefore, a total of four channels were used by two vibration sensors to record displacements in the vertical and horizontal directions at both ends of the shaft. A Pulse 3560C device was used to digitize the acquired signal. Details of the data collection system are provided in Table 2.1. The experiment was performed at a constant rotational speed of 2,580 revolutions per minute (RPM), and the signal was sampled at a rate of 65.5 kHz.

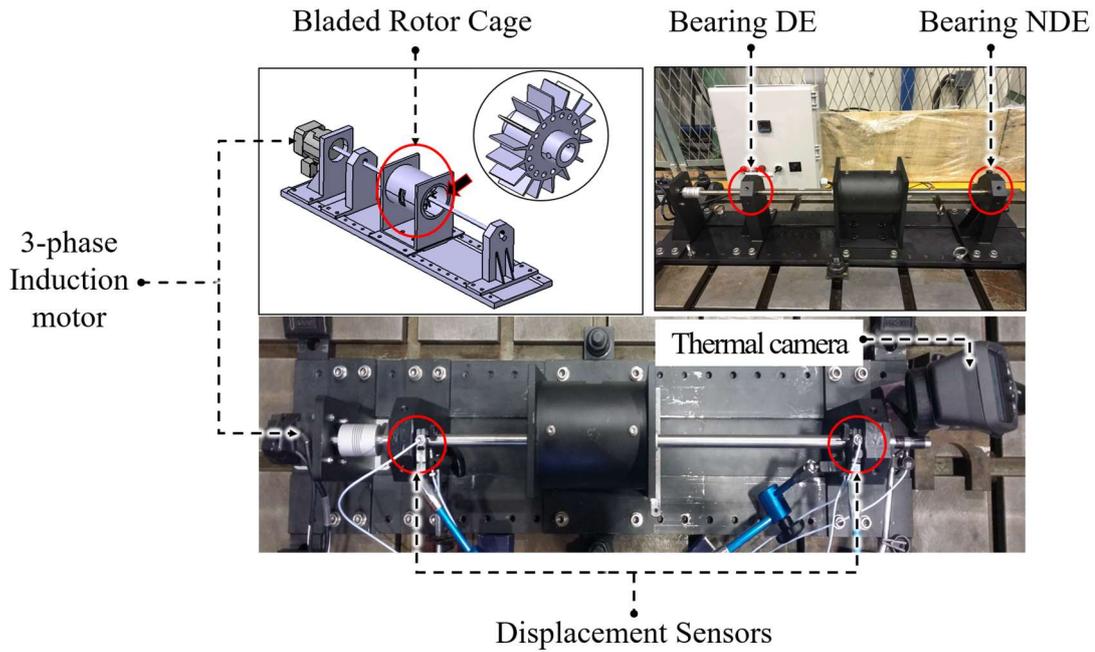


Figure 2.2. The self-designed test rig for rub-impact fault diagnosis.

Table 2.1. Data acquisition system specifications.

Displacement Sensors (3300 XL NSv)	Frequency range: 0 to 10 kHz Sensitivity: 7.87.V/mm (200 mV/mil) +12.5%/-20%
DAQ System (Pulse 3560C)	Generator. Input/Output 4/2 ch. module Input/Output 5/1-ch. controller module Frequency range: 0 to 25.6 kHz

In this dissertation, a coupling shaft imbalance and periodical local blade rub-impact fault was simulated during the experiment. Specifically, to create the interactions between rotor disk blades and rotor cage (i.e., blade rub-impact fault), shaft imbalance fault ($\angle 45^\circ$) has been first simulated by attaching additional weights to the rotor disk. This imbalance in the shaft caused local periodical interactions between the rotor blades and the stationary part. The appearance and intensity of the rubbing process were detected and validated using a thermal camera mounted on the NDE of the shaft. Adjusting the extra mass to the rotor disk resulted in various imbalance and rubbing intensities, as shown in Figure 2.3. The total duration of the recorded signal for each signal class was 59 seconds. However, each signal was divided into slices of 1 second each for signal processing and feature extraction.

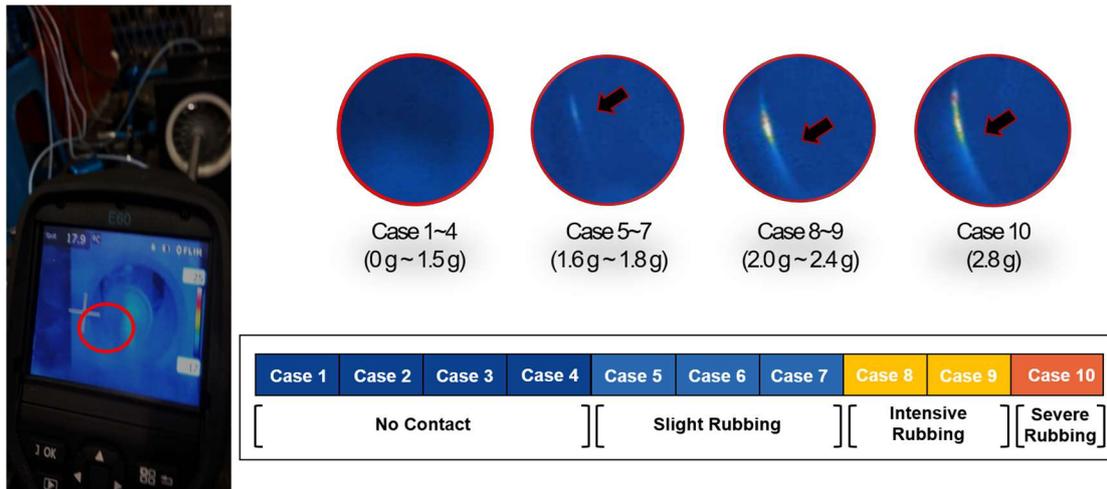


Figure 2.3. Changes in rubbing intensity validated using a thermal camera.

To verify the quality of the collected rubbing fault dataset and its capability of reflecting the degree of wear in a real scenario, an additional session of the experiment has been performed. In this experiment, the data were collected with the dismantled rotor cage while applying the same types of extra weights to the rotor. Thus, during this experiment session, only the rotor imbalance fault could be observed. Next, to compare the signal samples obtained during the two experiments with two different types of mechanical faults, i.e., rotor imbalance fault and blade rub-impact fault caused by rotor imbalance (coupling fault), three simple feature parameters that are sensitive to impulses and are capable of describing the changes in statistical properties of the signals based on their amplitude have been extracted. Specifically, these features are the root mean square (RMS), standard deviation (Std), and 5th normalized moment [71]. The three-dimensional feature space created by these statistical feature parameters is demonstrated in Figure 2.4.

From Figure 2.4, it is seen that the feature values extracted from the signals collected in both of the experiments form clear increasing trend curves that are following the progression of the fault. In the case when only rotor imbalance fault is presenting in the system, the feature values are growing with the increase of the weights applied to the shaft. In the case of the rub-impact fault caused by rotor imbalance, the feature values also produce an increasing trend line and this trend grows with both the increase of the rotor imbalance and the increase of rubbing fault intensity (i.e., the degree of wear). Based on these observations, it can be concluded that the rub-impact fault

simulation performed in this study follows the fault progression and capable of reflecting the degree of wear in real-life scenarios.

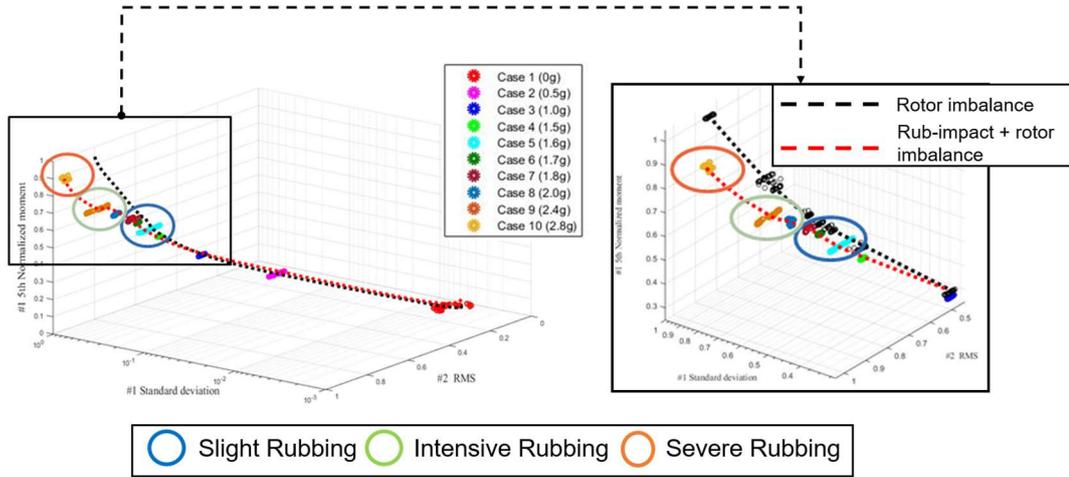


Figure 2.4. The feature space and trend curves formed using the statistical feature parameters extracted from the data samples of a single rotor imbalance and combined blade rub-impact and rotor imbalance faults.

2.2.2 Empirical Mode Decomposition and Ensemble Empirical Mode Decomposition

In this subsection, the necessary background on the original EMD and EEMD algorithms is provided.

2.2.2.1 Empirical Mode Decomposition

EMD [22] decomposes the original signal $x(t)$ into a finite number of oscillatory components using a sifting process. In order to be defined as an IMF, the function should meet the following two conditions [52, 72]: (1) the number of extrema and zero-intersect points are either equal to each other or differ by at most one, and (2) at any point, the mean of the envelopes defined by local maxima and minima is equal to zero. Each IMF can be considered as a specific frequency band of the original signal, where the first IMFs represent high-frequency bands and the last IMFs correspond to lower frequency bands [22]. The details on the conventional EMD decomposition are presented in Algorithm 2.1.

After the decomposition process, the input signal $x(t)$ can be defined as shown below:

$$x(t) = \sum_{i=1}^n IMF_i(t) + r_n(t), \quad (2.1)$$

where n is the total number of extracted IMFs and $r_n(t)$ is the residue of the signal decomposition. The quality of the decomposition can be evaluated by computing the amplitude error between the original signal and the reconstructed one.

Algorithm 2.1. EMD Algorithm.

1. Initialize: $i=1, r_0 = x(t)$
 2. Extract the i th IMF component.
 - 1) Initialize: $h_{i(k-1)} = r_i, k = 1$.
 - 2) Find the local extrema of the signal $h_{i(k-1)}$.
 - 3) Construct the upper and lower envelopes $enmax_n$ and $enmin_n$ by cubic spine line interpolation through the local maxima and minima of $h_{i(k-1)}$, respectively.
 - 4) Compute local average as the envelop mean $enm_{i(k-1)}$ of the upper and lower envelopes of $h_{i(k-1)}$.
 - 5) Compute the candidate IMF as $h_{ik} = h_{i(k-1)} - enm_{i(k-1)}$.
 - 6) If h_{ik} is an IMF, assign it to be $IMF_i = h_{ik}$. Otherwise, repeat the steps 1)-5) with $k = k + 1$.
 3. Define $r_{i+1} = r_i - IMF_i$.
 4. If r_{i+1} contains intrinsic oscillation mode (has at least 2 extrema), then set r_{i+1} as the next input signal and repeat steps 1 to 4. Otherwise, r_{i+1} is the residue of the decomposition.
-

EMD is a powerful tool that is capable of extracting nonlinear and nonstationary parts of the original signal. However, its crucial disadvantage is the mode-mixing problem, which leads to multiple oscillating components being presented in a single IMF or similar oscillating components being split in different modes with smaller amplitudes. This drawback causes difficulties in interpreting the physical meaning of each mode for feature extraction and fault diagnosis.

2.2.2.2 Ensemble Empirical Mode Decomposition

EEMD [23, 52] was introduced to overcome the problems of mode-mixing observed in conventional EMD. The concept of EEMD is to obtain precise IMF components by taking the mean of several EMD trials performed on the original signal, with the addition of various realizations of white noise in each trial. The main advantage of EEMD over conventional EMD is that it significantly reduces the chance of mode-mixing and is capable of decomposing the original

signal more precisely into a set of “true” IMFs. The implementation steps of EEMD are summarized in Algorithm 2.2.

Algorithm 2.2. EEMD Algorithm.

1. Initialize the number of ensembles L , the various realizations of Gaussian noise, and $l = 1$.
 2. Perform the $l - th$ decomposition trial on the signal with added white noise.
 - 1) Generate an artificial observation $x_l(t) = x(t) + \omega_l(t)$, where $x_l(t)$ is a noise-added signal of the $l - th$ decomposition trial and $\omega_l(t)$ is the $l - th$ realization of white noise.
 - 2) Using EMD, described in Algorithm 2.1. , completely decompose the ensemble $x_l(t)$ into K IMFs $IMF_{k,l}$ ($k = 1, 2, \dots, K$), where $IMF_{k,l}$ states for $k - th$ IMF of the $l - th$ trial, and K is the total number of IMFs.
 - 3) If $l < L$, assign $l = l + 1$. Repeat steps 1)-3) with different white noise realizations each time.
 3. Compute the ensemble mean \overline{IMF}_k of the L decomposition trials for each extracted IMF as $\overline{IMF}_k = (1/L) \sum_{l=1}^L IMF_{k,l}$, $k = 1, 2, \dots, K$, $l = 1, 2, \dots, L$.
 4. Assign the mean \overline{IMF}_k ($k = 1, 2, \dots, K$) of each of the K IMFs as the final IMFs.
-

Figure 2.5 shows the IMFs obtained by conventional EMD and EEMD for a signal corresponding to shaft imbalance fault with intensive rubbing condition. Figure 2.5 (a) shows that IMF #8, which corresponds to the fundamental frequency, contained an additional oscillatory component that affects the quality of information carried by this mode, an example of mode-mixing. In this figure, IMF #9 and IMF #10 correspond to the $1/2X$ and $1/3X$ fractional harmonics; however, the behavior of these oscillations looks very similar. Figure 2.5 (b) demonstrates that EEMD better separated the extracted modes, and the IMF of the fundamental frequency (IMF #10) is not affected by other oscillations or noise. Also, EEMD was able to decompose the original vibration signal more accurately into a larger number of modes. IMF #9, the $3/2X$ fractional frequency harmonic, became visible after decomposition, while in conventional EMD decomposition the component with this frequency harmonic was absent. Moreover, better separation of the IMFs was observed in the range of low fractional harmonics such as $1/2X$ and $1/3X$. These results are reasonable because due to its specific features as an improvement to the traditional EMD approach, EEMD can better separate oscillating components and deliver more clear IMFs.

Since it is known that the appearance and behavior of fractional harmonics is an essential feature for rub-impact fault diagnosis [49, 50, 73], the better extraction and separation of the modes containing these harmonics make EEMD preferable to EMD.

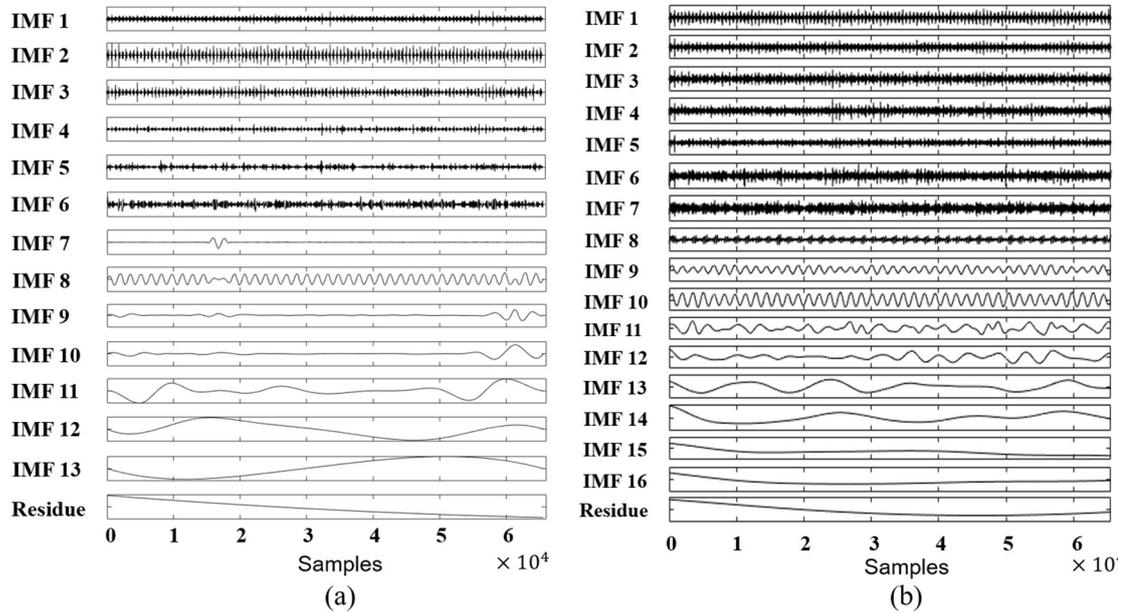


Figure 2.5. Decomposition of a one-second vibration signal containing shaft imbalance and intensive rubbing fault by (a) EMD and (b) EEMD approaches.

2.2.3 IMF Selection Procedure for Rubbing Fault Diagnosis

As discussed in Section 2.1, the extracted IMFs can be either signal-dominant or noise-dominant, so it is crucial to select informative IMFs that contain intrinsic information about rub-impact faults. This chapter presents a novel adaptive selection method for informative signal-dominant IMFs that can be applied to the domain of rubbing signals. The following two procedures must be applied in IMF selection: evaluating the extracted components using certain criteria to determine which candidate IMFs contain the most valuable information and creating a subset that contains those chosen IMFs. To define the objective function for optimal IMF selection, the rubbing phenomenon is carefully analyzed using the DFP of rub-impact and the KLD (a statistical similarity metric) of each extracted IMF. Based on this new criterion, all of the components can be sorted by their relevance to the rubbing process. Once the IMFs are sorted and their quality is determined, the best adaptively selected candidates are used for reconstructing a signal with reduced noise that contains clear rub-impact fault components and can be used for rubbing fault feature extraction.

Whenever rub-impact faults occur in a system, they affect the harmonics of the fundamental frequency and some fractional frequencies, which can be observed in the envelope power spectrum of the signal. DFP aims to detect the presence and power of these expected frequencies and their harmonics in the envelope power spectra of the IMFs and quantifies each IMF concerning the ratio of the presence using the procedure presented in Figure 2.6. To ensure that the DFP can determine informative and meaningful IMF components, the important frequencies of interest and their harmonics were determined based on various studies of rubbing processes [49, 50, 73] to be the following: $1/3X$, $1/2X$, $2/3X$, $1X$, $4/3X$, $3/2X$, and $5/3X$. The detailed DFP calculation procedure is as follows.

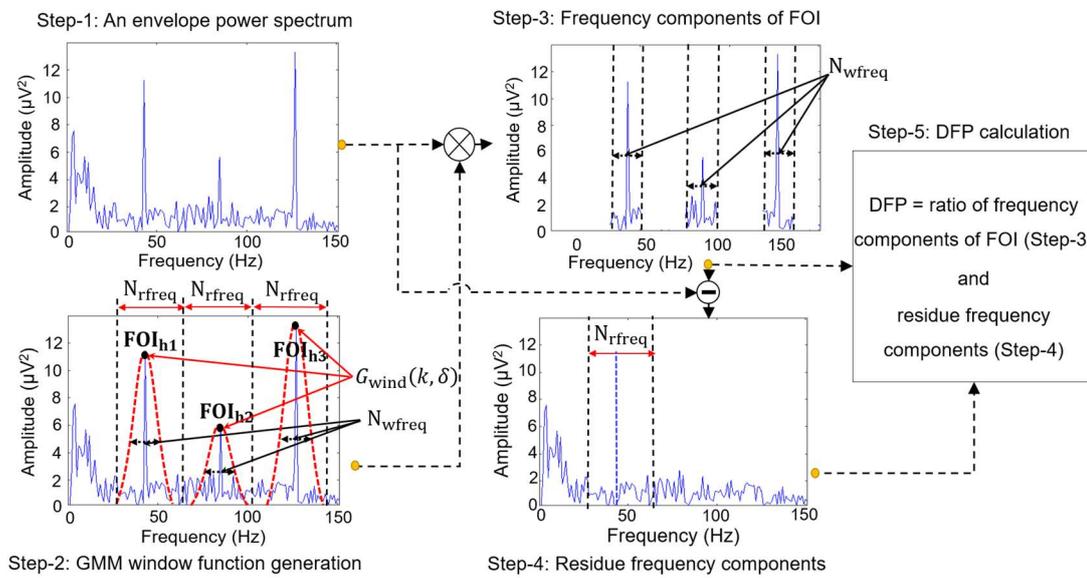


Figure 2.6. Evaluation of the IMFs by computing the DFP over the frequency of the interest and its harmonics.

Step 1: To compute an envelope power spectrum, the analytical signal of the IMF in the time domain is first calculated. For example, if $x(t)$ is the original IMF signal, the analytical signal can be represented as a combination of the original IMF signal and the virtue of the Hilbert transform. The analytical signal can thus be formulated as follows:

$$x^h(t) = x(t) + i\tilde{x}(t), \quad (2.2)$$

where $\tilde{x}(t)$ is the Hilbert transform. Convolution of the original IMF signal with the signal $1/\pi t$ yields the following:

$$\tilde{x}(t) = x(t) \times \frac{1}{2\pi} = \frac{1}{\pi} \int_{-\infty}^{+\infty} h(t) \frac{dt}{t - \tau}. \quad (2.3)$$

The power spectrum of the analytical signal is then obtained by taking the square of the absolute value of the Fourier transform, $|F\{x^h(t)\}|^2$, as shown in Step 1 of Figure 2.5.

Step 2: Since the envelope power spectrum reveals the transient impact of rubbing, a Gaussian mixture model (GMM) window ($G_{\text{window}}(k, \delta)$) is constructed around the peaks of the frequencies of interest and their integer multiples to attain residual components in the frequency domain of the envelope power spectrum. The coefficients of $G_{\text{window}}(k, \delta)$ are computed as follows:

$$G_{\text{window}}(k, \delta) = \sum_{i=1}^n \exp\left(-\frac{1}{2} \left(\delta \frac{(k - \text{FOI}_n)^2}{(N_{\text{rfreq}}/2)} \right)\right) \quad (2.4)$$

$$\text{subjected to } \text{FOI}_n - f_{\text{range}} \leq k \leq \text{FOI}_n + f_{\text{range}}.$$

where FOI_n defines the n_{th} harmonic of the frequency of interest; $n = 3$ was used to calculate the DFP. N_{rfreq} is the number of frequency bins in the range $\text{FOI}_n - f_{\text{range}} \leq k \leq \text{FOI}_n + f_{\text{range}}$, as shown below:

$$N_{\text{rfreq}} = 2 \times f_{\text{range}} / f_{\text{resolution}}. \quad (2.5)$$

The value of N_{wfreq} can be computed (see Step 2 in Figure 2.5) for a GMM-DFP as follows:

$$N_{\text{wfreq}} = ((4/100) \times \text{FOI}) / f_{\text{resolution}}, \quad (2.6)$$

$$\text{subjected to } \text{span}_{\text{FOI}} = (2/100) \times \text{FOI}.$$

Similarly, δ is a Gaussian random variable that is inversely proportional to the standard deviation and can be defined as follows:

$$\delta = (N_{\text{rfreq}} / N_{\text{wfreq}}) \sqrt{2 \ln m}. \quad (2.7)$$

Here, N_{wfreq} is the frequency bin size around the frequency-of-interest components, and the value of m is constant in the range $0 < m < 1$, ($m = 0.1$). A narrow band frequency range (i.e., $f_{\text{range}} = 1/3 \text{ FOI}$) is used for all the frequencies of interest in this chapter.

Step 3: The components of the frequencies of interest are then calculated by multiplying the defined Gaussian window $G_{window}(k, \delta)$ around the frequencies of interest and their harmonics in the attained envelope power spectrum.

Step 4: The residual frequency components are computed by subtracting the frequency of interest component (from Step 3) from the envelope power spectrum. Once the frequency components of FOI and residual components are obtained, the DFP is calculated as the ratio of the frequency-of-interest components and the residual frequency components, as shown below:

$$DFP = 10 \times \log \left(\sum_{n=1}^3 \left(\frac{\sum_{j=1}^{N_{wfreq}} FCF_n^2(j)}{\sum_{j=1}^{N_{rfreq}} RFC_n^2(j)} \right) \right). \quad (2.8)$$

Here, $FCF_{n,j}$ and $RFC_{n,j}$ are the magnitudes of the j_{th} frequency bin for the FOI components and residual frequency components, respectively, around the n_{th} harmonic of the FOI.

Note that a large DFP of an IMF means that the component contains valuable information about the behavior of the fundamental frequency, its fractional harmonics, and other frequencies that can be observed during the rubbing process. A small DFP value indicates the opposite — the current IMF could be noise-dominant and does not provide valuable information that can be used for rubbing fault diagnosis.

Next, the distances between the probability density functions (PDFs) of the extracted IMFs and the original signal are calculated to discover similarities between the hidden structures and determine how much of the original signal information is preserved in the selected component.

The distance function utilized in this chapter, KLD, is one of the most frequently used information-based distance measures and is a member of Shannon's entropy family [74]. The function is defined as follows:

$$KLD_k = \sum_{x \in X} PDF(x(t)) \ln \frac{PDF(x(t))}{PDF(IMF_k(t))}, \quad (2.9)$$

where $PDF(x(t))$ is the PDF of the original signal $x(t)$ and $PDF(IMF_k(t))$ is the PDF of the k^{th} extracted IMF.

Once the DFP and KLD have been calculated, the objective function (*Obj*) can be defined as in Eq. (2.10). This objective function will be used to determine the relevance of the extracted components and then select informative modes:

$$Obj_k = DFP_k / KLD_k. \quad (2.10)$$

Here, DFP_k represents the sum of all degrees of presence calculated for each N_{wfreq} and their three harmonics for the k^{th} IMF, and KLD_k is the measure of how close the PDF of the k^{th} IMF is to the PDF of the original signal.

Finally, the IMF selection procedure is completed using a thresholding technique that compares the objective value Obj_k with a threshold value that is set to 1. The threshold value is assigned to 1 because the relationship between the DFP and KLD for each IMF can be considered as a sort of signal-to-noise ratio (SNR). It is known that when the SNR is greater than 1, the amount of signal in the data is greater than the amount of noise. That is, a small KLD distance value in the Obj_k indicates that the PDF of the mode under evaluation is similar or close to that of the original signal. Therefore, the value of the objective function grows when the DFP (the metric that reflects the presence of important information) is large compared to KLD. This means that the IMF contains a significant amount of valuable information about the rubbing fault and its PDF is very close to that of the original signal, which indicates that the IMF is signal-dominant. In the opposite direction, the objective value decreases when the IMF contains a small amount of valuable information (low values of DFP) and a large KLD (the component appears to be noise-dominant). Therefore, the decision making about IMF inclusion in the optimal subset for signal reconstruction can be formulated as follows: if Obj_k is greater than or equal to 1, the IMF will be included in the optimal subset for signal reconstruction; otherwise, it will not be used.

Using this proposed objective function for assessing the quality of IMF and the threshold-based selection approach, all of the extracted components can be evaluated and the most valuable modes are adaptively selected based on their relevance to rubbing processes and the amount of information presumed from the original signal.

2.2.4 Hybrid Feature Extraction and Configuration of Feature Set

The proposed optimal IMF selection process provides a set of the most meaningful IMFs that carry important information about the rubbing processes present in the original signal. These selected IMFs are next utilized for signal reconstruction (i.e., obtaining a noiseless rubbing signal) and feature extraction. The reconstructed signal can be obtained using the following equation:

$$x_{rec}(t) = \sum_{l=1}^N IMF_l. \quad (2.11)$$

Here, N is the total number of selected IMFs and l corresponds to the order number of each selected mode.

After signal reconstruction, time-domain statistical features in the reconstructed signal and frequency-domain features in its complex envelope power spectrum are extracted. These features are considered to be discriminative because significant changes in the amplitude of the fundamental frequency and its fractional harmonics are usually observed in the envelope power spectrum when a rub-impact fault occurs. The amplitude of the vibration signal in the time-domain also changes due to the increased fluctuations of signal waves depending on the rubbing fault intensity. Thus, the changes in signal behavior and its energy with variations in rubbing intensity levels can be well characterized by extracting dimensional time-domain statistical feature parameters such as the root mean square (RMS), kurtosis, skewness, and the square root of the amplitude (SRA) from the signal reconstructed using the sets of selected IMFs. These features are widely used as health indicators of various systems in fault diagnosis problems and can provide a good insight into rubbing processes in the time domain because they are known to be features sensitive to impulse faults [71]. Here, skewness and kurtosis are the third and fourth central moments of standard deviation. These two features are known as statistical indicators sensitive to the degree of peakedness of the signal that can be used well to characterize the variability of the vibration signal in the time domain when rubbing fault occurs in a system. Also, the RMS and SRA are used in this study to describe the changes in intensity levels of the vibration signal affected by the rub-impact fault. Note that RMS and SRA are usually both thought of as statistical parameters that reflect the behavior of the signal's amplitude in different scales, so these two features may not provide drastically different information, but they complement each other when used simultaneously.

Furthermore, frequency domain analysis is helpful for discovering some information that cannot be observed in the time domain [75]. Thus, features that involve statistical properties of frequency are extracted from the envelope power spectra of the reconstructed signals. The frequency spectra of the rub-impact signals obtained by conventional frequency-domain signal analysis techniques such as FFT fail to present relevant information about rubbing processes. Various studies [56, 59] have shown that the frequency spectra of the original rub-impact signals obtained by FFT usually cannot represent the frequency harmonics relevant to rub-impact faults, although the fundamental frequency and some high-order harmonics still may be present. The frequency-domain features extracted in this way cannot be considered as good features for diagnosing rub-impact faults because the behavior of the fundamental frequency and its higher harmonics can be affected by various mechanical faults in rotational machinery as well as properties of the environment in which the machine is installed. It is therefore difficult to verify whether these extracted features are related to the ongoing blade rubbing faults or not. On the other hand, the frequency-domain features extracted in this study after time-frequency EEMD decomposition of the original vibration signal followed by informative IMF selection are definitely

capable of reflecting rub-impact faults because the fault information is highly detectable and well presented by the envelope spectra of the signals reconstructed using sets of selected IMFs. Moreover, it was observed that the valuable rubbing fault-related harmonics are detectable and well-localized in the envelope power spectrum of the reconstructed signal. Thus, the frequency-domain features extracted in this study are as follows: mean frequency, frequency RMS, and frequency standard deviation (Std). Here, the mean frequency value corresponds to the average frequency value of the envelope power spectrum computed for reconstructed signal, frequency RMS characterizes the intensity of the signal in frequency-domain, and frequency Std describes the deviation of the signal from its main frequency components in frequency-domain.

Due to the extraction of the feature-parameters from both time and frequency domains the proposed feature model can be considered as a hybrid one. The basic idea beyond this model is that each feature extracted from both domains aims to characterize and quantify the specific physical and statistical properties of the denoised vibration signal affected by rub-impact. Furthermore, the features used in this study are well-recognized and frequently used health indicators in other problems of the vibration-based condition monitoring, such as rolling-element bearing fault diagnosis [1] and prognosis [76]. Concretely, frequency-domain features such as mean frequency, frequency RMS, and frequency standard deviation are frequently applied for REB fault diagnosis, whereas the time-domain RMS, kurtosis, and skewness are popular health indicators used for fault prognosis and remaining useful lifetime estimation. All the extracted features are shown in Table 2.2.

Table 2.2. Time- and frequency-domain statistical feature parameters.

Parameters	Equations	Parameters	Equations
Root mean square (f_1)	$\sqrt{\left(\frac{1}{N}\right) \sum_{i=1}^N x_{rec}^2(i)}$	Square root of the amplitude (f_2)	$\left(\left(\frac{1}{N}\right) \sum_{i=1}^N \sqrt{ x_{rec} }\right)^2$
Skewness (f_3)	$\left(\frac{1}{N}\right) \sum_{i=1}^N \left(\frac{x_{rec}(i) - \bar{x}_{rec}}{\sigma}\right)^3$	Kurtosis (f_4)	$\left(\frac{1}{N}\right) \sum_{i=1}^N \left(\frac{x_{rec}(i) - \bar{x}_{rec}}{\sigma}\right)^4$
Mean frequency (f_5)	$\left(\frac{1}{N}\right) \sum_{i=1}^N f(i)$	RMS frequency (f_6)	$\sqrt{\left(\frac{1}{N}\right) \sum_{i=1}^N f^2(i)}$
Frequency standard deviation (f_7)	$\sqrt{\left(\frac{1}{N}\right) \sum_{i=1}^N (f(i) - f_5)^2}$		

where x_{rec} is the reconstructed signal in the time domain, f is the spectral component of the reconstructed signal x_{rec} , and σ is the standard deviation of the reconstructed signal x_{rec} .

2.3 Experimental Results and Discussion

The effectiveness of the proposed rub-impact fault feature extraction technique with its novel IMF selection procedure is presented, along with its possible application to fault diagnosis.

2.3.1 Training and Testing Data Configuration

Appropriate training and testing dataset configuration is important to determine the generalized quality of the proposed fault diagnosis approach. In this dissertation, the blade rub-impact fault was simulated using a shaft imbalance produced by adding extra weights to the rotor disk. In total, 10 different weights (i.e., rubbing intensities) were added to the shaft, namely 0.0, 0.5, 1.0, 1.5, 1.6, 1.7, 1.8, 2.0, 2.4, and 2.8 g, respectively. A 59-second-long vibration signal was acquired for each signal class. Therefore, the created dataset contained 590 signal instances in total. In this chapter, the set of features extracted for the experimental part consisted of a total of $N_c \times N_s \times N_f$ features. Here, N_c is the number of signal classes (conditions) simulated in this study, N_s is the number of data instances for each condition, and N_f is the number of extracted features.

To avoid the influence of randomness on the experimental results, in this chapter, the experiment has been performed ten times. In each experimental trial, the entire dataset was randomly split into training and testing subsets at a ratio of 8:2. Specifically, during each experiment, 47 randomly chosen data instances from each category were utilized for training the machine learning classification algorithm (i.e., SVM in this chapter), whereas the remaining 12 unseen data samples from each class were used to construct the testing subset. Thus, each training subset contained 470 instances and each testing subset consisted of 120 remaining data samples, respectively. However, it is important to keep in mind that some minor deviations between numbers of instances for each category can be observed during the random sampling process.

2.3.2 Validation of the Optimal IMF Subset Using the Proposed Approach

The proposed optimal IMF selection method allows the derivation of the specific objective function values that can be used to evaluate the quality of the components extracted by EEMD and the examples obtained for some signal samples from each signal class are presented in Table 2.3. As a result of the introduced optimal IMF selection procedure, the sets containing the most valuable intrinsic components can be obtained for each vibration signal sample. The examples of IMF sets obtained for the same signal samples belonging to different classes as was presented in Table 2.3 are shown in Table 2.4. In this study, rubbing faults were caused by an imbalance in the testbed shaft, where the imbalance was created by attaching extra weight to the rotor disk. As shown in Figure 2.3 in Section 2.2.1, the increase in weight results in an increased degree of shaft imbalance,

which causes the various rubbing conditions observed during the experiment. It was also observed that some rubbing conditions, which were validated using a thermal camera, had different degrees of imbalance due to the additional weight. This observation explains why our IMF selection method adaptively delivered subsets with various components for each class; however, each subset was also shown to contain IMFs that were common for all classes.

Table 2.3. The example of objective values of the proposed optimal IMF selection metric calculated for signal samples collected from different signal classes.

		Obj _k									
Class #	1	2	3	4	5	6	7	8	9	10	
k	1	2	3	4	5	6	7	8	9	10	
1	0.12	0.23	0.92	1.2	0.75	1.44	1.47	2.24	0.78	1.64	
2	1.66	0.77	2.47	1.8	1.12	3.57	3.9	2.85	2.2	2.98	
3	2.3	3.86	20.6	19.2	2.52	21.9	7.24	28.3	2.44	20.4	
4	0.17	0.09	0.24	0.17	0.55	0.31	0.29	0.34	0.22	0.12	
5	0.38	0.23	0.21	0.17	0.9	0.12	0.32	0.25	0.21	0.32	
6	0.44	0.67	0.62	0.21	0.31	0.71	0.35	0.19	0.21	0.26	
7	1.71	7.48	5.75	3.39	1.34	3.79	6.41	4.46	1.45	12.2	
8	2.93	7.37	25.2	18.7	2.45	30.5	8.2	27.1	2.53	18.9	
9	0.96	12.2	17.5	34.0	6.37	25.8	12.6	51.1	8.12	29.0	
10	1.21	5.1	43.1	40.1	8.45	45.8	10.3	88.1	8.01	70.7	
11	0.61	0.3	0.16	0.32	0.12	0.28	0.17	0.15	0.95	0.43	
12	0.38	0.2	0.18	0.16	0.14	0.19	0.23	0.19	0.34	0.18	
13	0.23	0.26	0.22	0.24	0.25	0.23	0.28	0.17	0.22	0.25	
14	0.23	0.26	0.24	0.3	0.31	0.22	0.23	0.25	0.27	0.24	
15	0.23	0.26	0.27	0.25	0.42	0.30	0.24	0.24	0.31	0.27	
16	0.29	0.24	0.27	0.24	0.23	0.33	0.33	0.24	0.33	0.30	
17	0.23	0.3	0.23	0.24	0.41	0.23	0.28	0.24	0.27	0.29	
18	-	-	-	-	-	-	0.23	-	-	-	

The values highlighted with bold font correspond to the selected valuable IMFs.

Table 2.4. The subsets of optimal IMFs obtained for signal samples presented in Table 2.3.

		Classes									
		1	2	3	4	5	6	7	8	9	10
Selected		2,3,7	3,7,8	2,3,7	1,2,3,7,	2,3,7,	1,2,3,7	1,2,3,7	1,2,3,7	2,3,7	1,2,3,7
IMFs		8,10	9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10

Since the modes obtained by the decomposition are amplitude-modulated, the envelope power spectrum of the reconstructed signal must be computed to verify the quality of the subsets provided by IMF selection. Figure 2.7 demonstrates the envelope power spectra of the original and reconstructed signals using the subsets of selected IMF components corresponding to the intensity

classes produced by 0 g and 2.8 g of extra weight. As shown in Figure 2.7 (a), it was observed that the envelope power spectrum of the original signal when no rubbing occurred mostly consisted of the fundamental frequency and its high-order harmonics. No fractional harmonics were clearly observed in the power spectrum. Also, the peak amplitude of the fundamental frequency harmonic was drastically higher than the peak amplitudes of other harmonics present in the power spectrum. Figure 2.7 (b) shows the envelope power spectrum of the reconstructed signal when no rubbing occurred in the system. This power spectrum includes the fundamental frequency $1X$ and its high-order harmonics, while the majority of fractional harmonics are not present or their amplitudes are very small, which is reasonable for a system in this state. The peak amplitudes of the harmonics present were also relatively small. Figure 2.7 (c) presents the envelope power spectrum of the original signal acquired during a severe rub-impact fault. From this figure, it is seen that the main frequency $1X$ and its multiple harmonics were clearly present in the envelope power spectrum. Moreover, some of the fractional harmonics that are considered evidence of the rubbing process, $1/3X$ and $4/3X$, were barely seen in this figure. Even though these fractional harmonics were present, the problem with extracting features for rubbing fault diagnosis from this kind of power spectrum is that the amplitude of the fundamental frequency is noticeably higher than the amplitudes of all other harmonics. Therefore, the numerical values of features extracted from this original signal mostly reflect the behavior of the main frequency, while the influence of the fractional frequencies will be negligible. It is known that many various mechanical faults in rotating machinery affect the behavior of the fundamental frequency. Therefore, when the only signs of a fault are amplitude changes of the fundamental frequency and its high-order harmonics, it is not easy to determine whether the extracted features actually reflect the rubbing process or whether they are related to other mechanical faults or the environmental features. In contrast, Figure 2.7 (d) shows that the frequencies which appeared in the envelope power spectrum of the reconstructed signal of the severe rub-impact fault contained important information corresponding to the various signal harmonics, such as the fundamental frequency $1X$, its high-order harmonics $2X$, $3X$, $4X$, and $5X$, and the fractional harmonics $1/3X$, $2/3X$, $4/3X$, $5/3X$, $10/3X$, and $9/2X$, which are considered valuable features for rub-impact faults [49, 50, 73]. Furthermore, the amplitudes of the fundamental frequency component and its high order harmonics during the severe rubbing process were higher than those shown in Figure 2.7 (b). These observations reveal that the selected IMF subsets contain meaningful information and that the features extracted from the reconstructed signals well represent rubbing faults of various intensities.

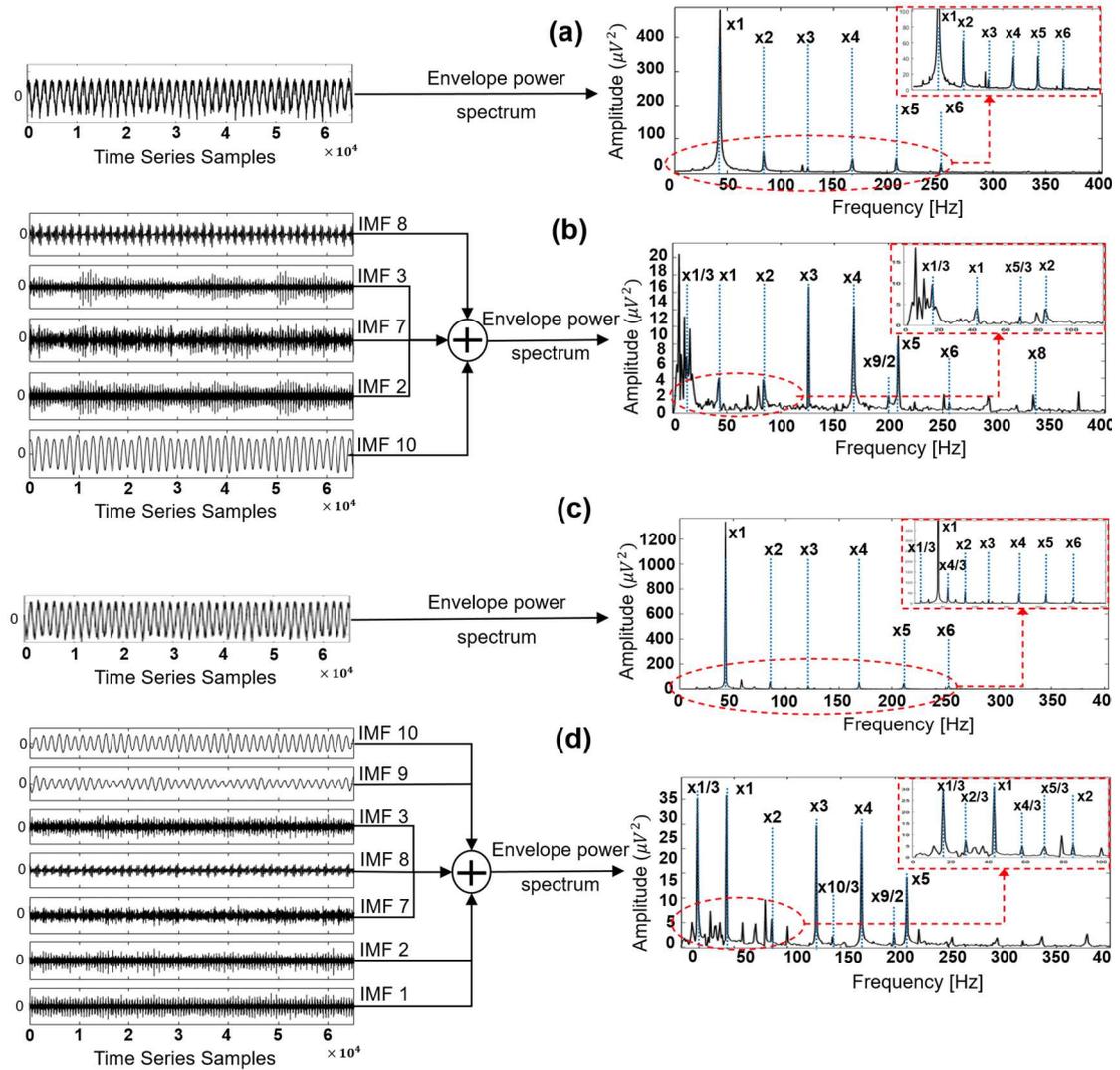


Figure 2.7. Examples of envelope power spectra of (a) original vibration signal from class #1 (0.0 g), (b) reconstructed signal using selected IMFs for class #1 (0.0 g), (c) original vibration signal from class #10 (2.8 g), and (d) reconstructed signal using selected IMFs for class #10 (2.8 g).

Figure 2.8 shows the examples of envelope power spectra obtained after signal reconstruction signal samples corresponding to classes #1, #5, #8, and #10. From this figure, it can be seen that the appearance of fractional harmonics and their amplitude behavior, combined with information about the fundamental frequency and its high order harmonics, allows for the differentiation of rub-impact faults of various intensities, as well as differentiation of the normal state of the system with no rubbing faults.

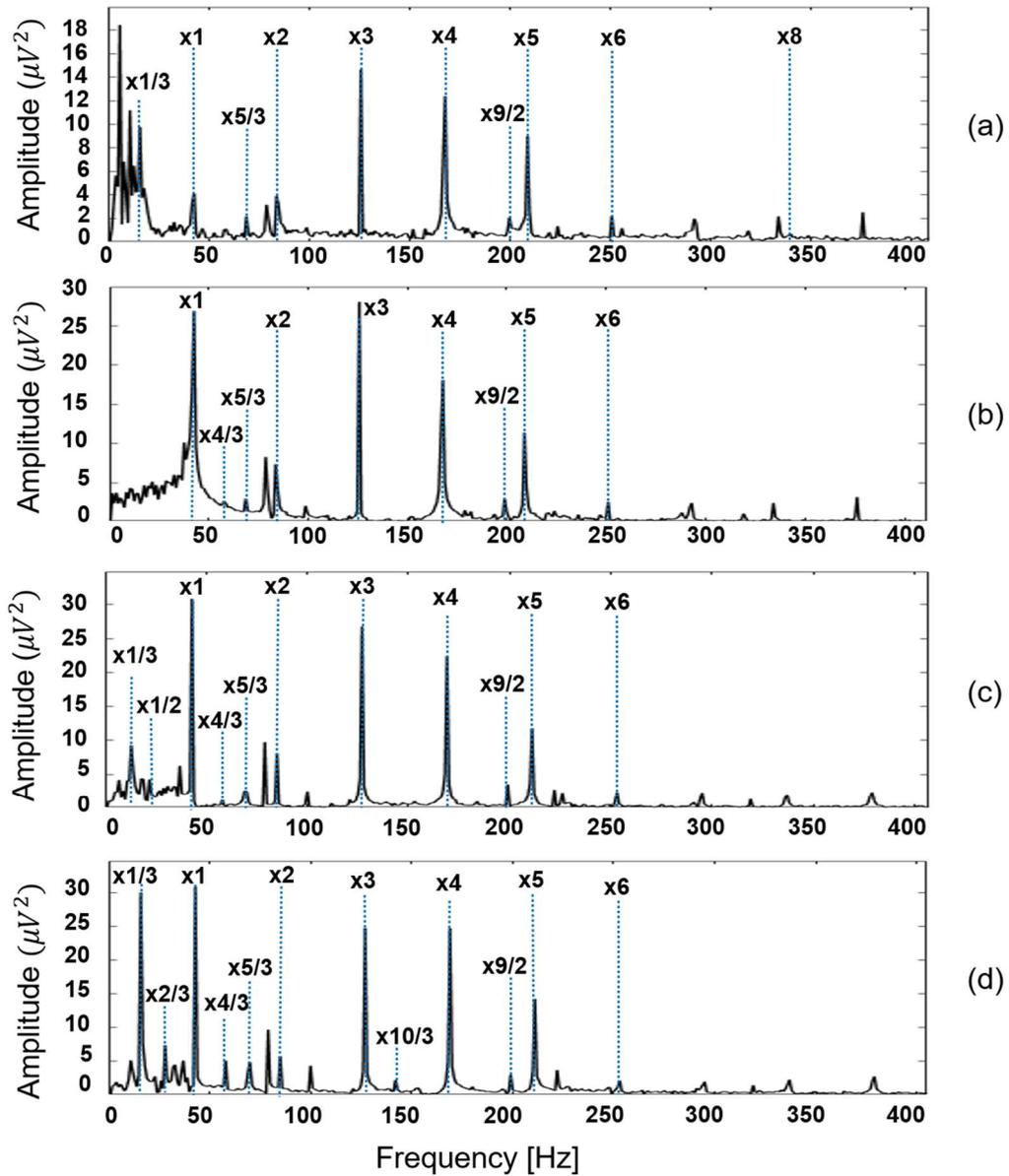


Figure 2.8. Examples of envelope power spectra of reconstructed signals obtained using EEMD and the proposed IMF selection technique for classes (a) #1 (0.0 g), (b) #5 (1.6 g), (c) #8 (2.0 g), and (d) #10 (2.8 g), respectively.

2.3.3 Performance Evaluation of the Proposed Rubbing Fault Feature Extraction Scheme with New IMF Selection Procedure

To evaluate the quality of the features extracted from the reconstructed signal and their ability to provide sufficient information about rub-impact faults with different intensity levels in rotor blades, the proposed feature extraction approach has been compared with three TFA feature extraction methods that use numerical-valued features for diagnosing rub-impact faults. The first TFA method utilizes wavelet packet transform and MSV computation to extract features from signals containing rubbing faults (referred to as WPT+MSV) [43]. The second TFA combines

conventional EMD decomposition and MSV extraction to create a set of features for differentiating rubbing faults (referred to as EMD+MSV) [60]. The third TFA approach applies digital wavelet transform (DWT) to the rubbing fault signal and then uses decomposed sample data in the time domain of the third level of transformation as feature vectors for diagnosing rub-impact faults (referred to as DWT+TDSIG) [44]. Also, for comparison purposes, the proposed hybrid feature model has been applied to signals reconstructed using sensitive IMFs selected by the technique presented in [65] (referred to as SensIMF+HFM). This chapter employs the one-against-all multiclass SVM (OAAMCSVM) classifier [69, 70] to perform the comparison of the above approaches. To ensure the repeatability of the results and exclude the effect of randomness, the experiments were performed 10 times with different combinations of training and testing data.

The fault classification performance for the methods mentioned above is evaluated using the widely micro-averaged forms of widely used metrics [77], such as micro-averaged recall (Rec_μ), micro-averaged precision ($Prec_\mu$), micro-averaged F1-score ($F1_\mu$), and total fault classification accuracy (FCA). It is decided to use the micro-averaged versions of these metrics to address the possible deviations in the numbers of data samples presented in each class in the testing subsets due to the random sampling procedure applied at each trial of the experiment. These metrics are expressed as follows:

$$Rec_\mu = \frac{\sum_{k=1}^K TP_k}{\sum_{k=1}^K (TP_k + FN_k)} \times 100; \quad (2.12)$$

$$Prec_\mu = \frac{\sum_{k=1}^K TP_k}{\sum_{k=1}^K (TP_k + FP_k)} \times 100; \quad (2.13)$$

$$F1_\mu = 2 \times (Prec_\mu \times Rec_\mu) \times 100 / (Prec_\mu + Rec_\mu); \quad (2.14)$$

$$FCA = \frac{\sum_k TP_k}{N} \times 100. \quad (2.15)$$

Here, TP_k , FP_k , and FN_k are the true-positive, false-positive, and false-negative values computed for the data instances of the class k , respectively; N is the total number of data samples in the datasets used for the experiment, and K is the total number of signal classes presented in the datasets. The experimental results expressed in these metrics averaged over 10 experiments are tabulated in Table 2.5. The standard deviations (Std) of the metrics above were also calculated and presented in the results.

Table 2.5. Experimental results.

Methods	Metrics (Std) (%)			
	Rec_μ	$Prec_\mu$	$F1_\mu$	FCA
Proposed	99.2 (0.9)	99.2 (0.9)	99.2 (0.9)	99.2 (0.9)
SensIMF+HFM	97.0 (1)	97.0 (1)	97.0 (1)	97.0 (1)
WPT+MSV	96.4 (2)	96.4 (2)	96.4 (2)	96.4 (2)
EMD+MSV	60.5 (5)	60.5 (5)	60.5 (5)	60.5 (5)
DWT+TDSIG	20.7 (3)	20.7 (3)	20.7 (3)	20.7 (3)

The results presented in Table 2.5 demonstrate that rub-impact fault feature extraction using EEMD with adaptive IMF selection based on a novel presented mode evaluation metric for the domain of blade rubbing faults diagnosis and the proposed hybrid feature model outperforms the reference methods in terms of average fault classification accuracy, with an accuracy of 99.2 % over 10 experiments. This table also shows that the average Rec_μ , $Prec_\mu$, and $F1_\mu$ values of the proposed method over 10 experiments were over 99.2% and the standard deviations of these metrics did not exceed 0.9%.

The statistics of FCA metric distributions over 10 experiments for each method used in the comparison are presented in form of boxplots in Figure 2.9. The black cross in each boxplot corresponds to the mean value of the FCA metric which is presented in Table 2.5.

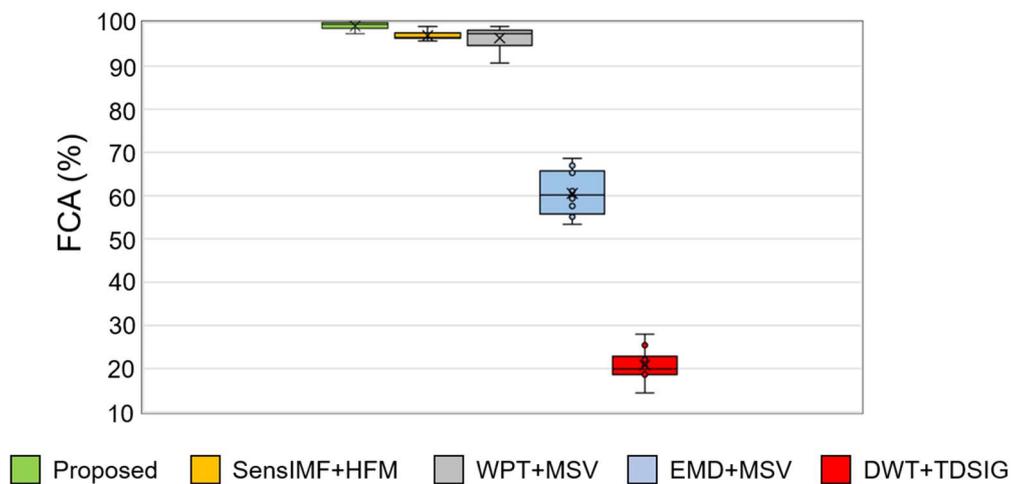


Figure 2.9. The boxplots demonstrating the statistics of the FCA metric over 10 experiments.

From this figure, it can be seen that the FCA values obtained by the proposed method in each experimental trial are distributed with high density around the mean FCA value which demonstrated the repeatability of the results and robustness to random training-testing data permutations. The values of the FCA metric for the SensIMF+HFM technique are also closely distributed around its mean, however slightly lower when compared to the proposed technique. WPT+MSV approach demonstrates very similar results to the SensIMF+HFM technique; however, the length of whiskers of this boxplot is larger which indicates the increasing number of outliers. Overall, it can be observed that the proposed, SensIMF+HFM, and WPT+MSV methodologies demonstrated stable results in terms of FCA that were always higher than 90% during the experimental trial. On contrary, regarding EMD+MSV and DWT+TDSIG referenced methods, their average FCAs are significantly lower than ones demonstrated by the methods mentioned above as well as the length of whiskers in larger which explains the increased values of standard deviations presented in Table 2.5.

The averaged confusion matrices for the proposed framework and the reference methods are presented in Figure 2.10. The confusion matrix is a robust technique that provides a visualization of the performance of a classifier algorithm in terms of the deviation between actual and predicted results [69]. According to the results in Figure 2.10 (a), the proposed method perfectly identified rubbing faults of all intensities with a very low misclassification rate compared to its counterparts: SensIMF+HFM [65] in Figure 2.10 (b), WPT+MSV [43] in Figure 2.10 (c), EMD+MSV [60] in Figure 2.10 (d), and DWT+TDSIG [44] in Figure 2.10 (e).

Additionally, the feature space projections created using t-Stochastic Neighbor Embedding (t-SNE) [78, 79] technique for each approach in this comparison are demonstrated in Figure 2.11. From this figure, it can be seen that the feature projection of the proposed model consisting of IMF selection, partial signal reconstruction, and HFM, demonstrates better feature separability which means that the proposed approach has high classification capabilities in comparison with referenced methods.

These results can be explained as follows. EEMD decomposition is capable of extracting clear ‘true’ IMFs which can be easily associated with frequency bands of the original signal. Importantly, the proposed adaptive IMF selection method for rubbing signals, in conjunction with the novel IMF evaluation technique, enables meaningful intrinsic components to be precisely determined. These components include valuable information about the frequency bandwidths containing specific frequency peaks and their harmonics, which are considered to be evidence of rubbing processes in a signal [49, 50, 73].

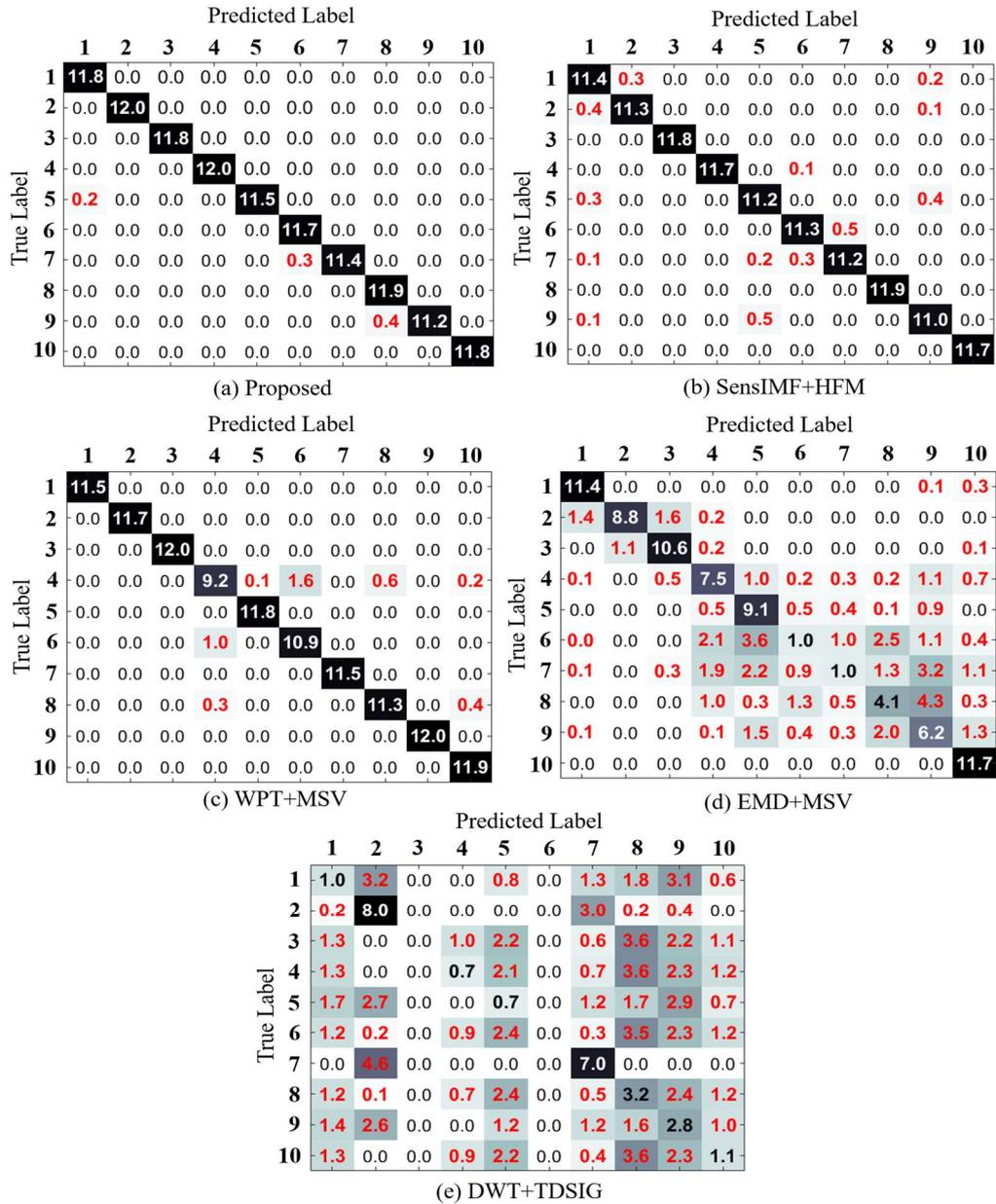


Figure 2.10. Confusion matrices for classification using (a) proposed, (b) SensIMF+HFM, (c) WPT+MSV, (d) EMD+MSV, and (e) DWT+TDSIG approaches. All results are presented as the average of 10 experiments.

The reference feature extraction approaches for rub-impact fault diagnosis resulted in average classification accuracies of 97%, 96,4%, 60,5%, and 20,7%, respectively.

Regarding SensIMF+HFM, Table 2.5 shows that the average FCA value achieved by this method for classifying rubbing faults of various intensities was slightly less than one, made by the proposed methodology. Note that SensIMF+HFM is a synthetic method that combines a sensitive IMF selection method for rub-impact fault diagnosis [65] and the hybrid feature model proposed in this chapter for extraction from signals reconstructed using the selected sensitive IMFs. The results presented in Table 2.5 and Figure 2.10 (b) demonstrate that this combined approach performs well at differentiating various rubbing conditions, but is still slightly worse than the framework proposed in this chapter. The difference between the results of the two methods can be explained as follows. Although both IMF selection methods were introduced for blade rub-impact fault diagnosis, the main concepts behind them are different. The method proposed in this chapter

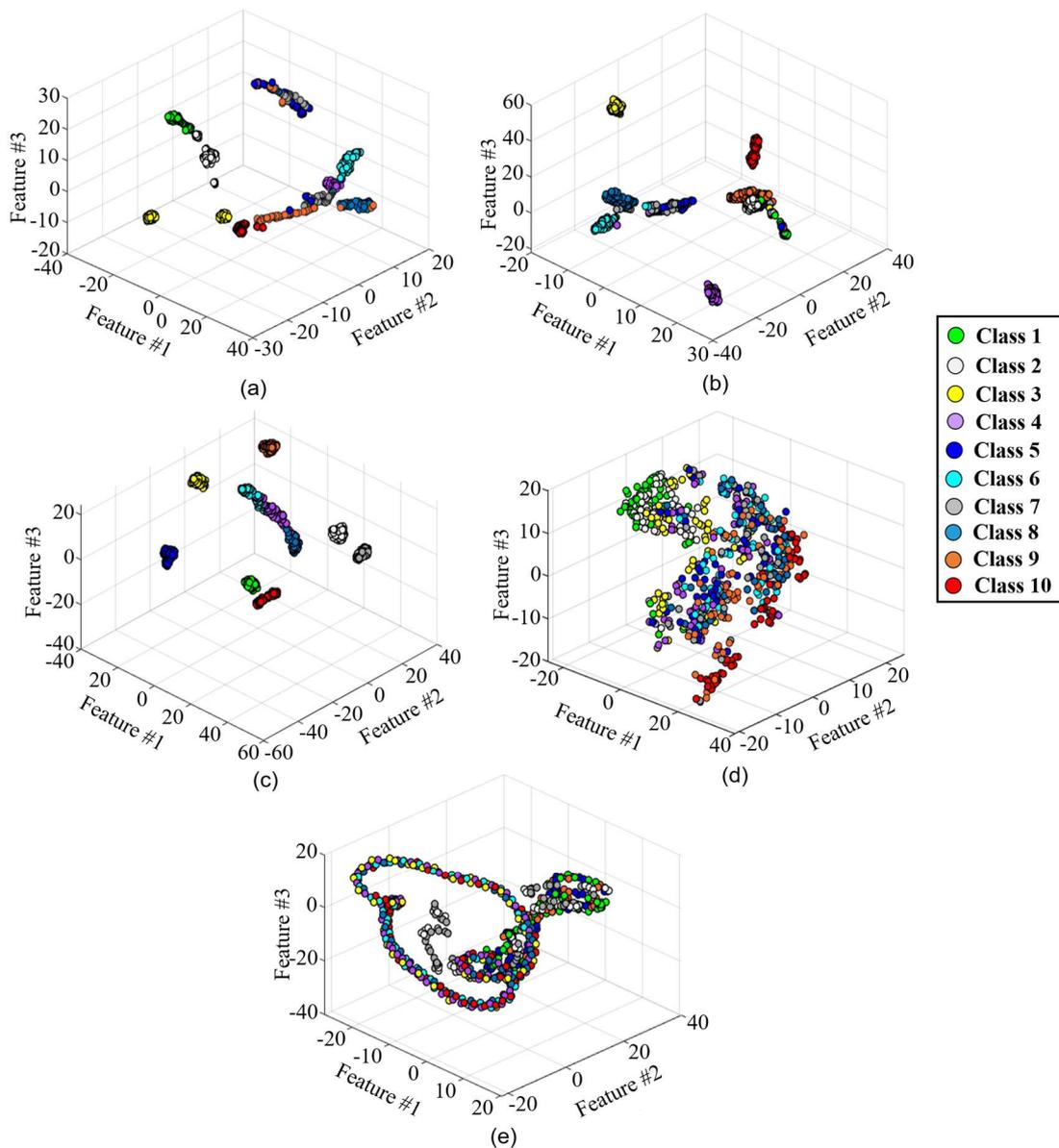


Figure 2.11. Feature space projections for (a) proposed, (b) SensIMF+HFM, (c) WPT+MSV, (d) EMD+MSV, and (e) DWT+TDSIG approaches, respectively.

aims to select the IMFs that are directly related to blade rubbing faults in the rotor system by both searching for the specific rub-impact fault harmonics in the envelope power spectra of IMF candidates and computing the KLD similarity between the extracted components and the original non-decomposed signal. On the other hand, the referenced method relies on computing ‘sensitivity factors’ of IMFs using only the correlation between the extracted components and the original signal, as well as signal acquired under the normal operating conditions of rotating machinery [65]. Considering that both methods utilize the same hybrid feature model for fault identification, it can be concluded that the approach proposed in this chapter can better highlight the important frequency harmonics of rubbing faults than a method that relies only on correlation properties of signals and IMFs without considering the exact information content carried by those modes.

Regarding WPT+MSV, the average FCA was smaller than that of the proposed fault identification pipeline. Table 2.5 and Figure 2.10 (c) show that in general, the WPT+MSV approach deals with differentiating rubbing faults of various intensities well; however, for the rubbing intensities corresponding to extra weights of 1.5g, 1.7g, and 2.0g attached to the rotor, the TP values were relatively low with increased standard deviation. These results can be explained because as mentioned above, rubbing faults in this study were simulated by adding extra weight to the rotor disk. This means that with increasing the degree of rotor imbalance, the intensity of blade rub-impact becomes more severe and the characteristics of the recorded vibration signal change significantly. It is known, that the quality of the wavelet packet transform strongly depends on the choice of mother wavelet function that in an ideal scenario should match the signal properly. Moreover, since this approach is not entirely adaptive and data-driven, the same mother wavelet function simply cannot perfectly match all of the changing types of vibration signals which contain both rotor imbalance and a complex nonlinear fault such as a rub-impact fault.

Regarding EMD+MSV, Table 2.5 and Figure 2.10 (d) show that the average FCA achieved by this method over 10 experiments is low compared to the proposed method, SensIMF+HFM, and WPT+MSV. There might be two main reasons for this. The first is the problem of mode-mixing which is inherent to the conventional EMD approach. The rub-impact fault is known to be a complex fault that induces the nonlinearity in the rotor system, and due to the mode-mixing problem, the quality of the extracted modes affected by this problem is not good enough to extract discriminative features. In the ideal case, when the IMF components are perfectly separated, each extracted component should represent a specific frequency band so that the extracted features can describe the ongoing processes in the rotor system well. However, in this case study mode-mixing causes the mixing of harmonics in the frequency spectra of the obtained IMFs which makes it difficult to clearly understand the physical meaning of each mode, and the quality (i.e., separability) of the extracted features becomes poor. The second reason for the

obtained results is that the EMD+MSV feature extraction approach does not consider the selection of valuable IMFs. The cardinality of extracted IMFs is usually large, and they can be either signal- or noise-dominant. When diagnosing industrial equipment such as rotor systems, it is important to determine which of the extracted components can provide useful and clear information because the features extracted from noise-dominant components can expand the feature models and degrade the overall classification performance by delivering poor-quality or redundant features that are not related to the mechanical fault being investigated.

Also, in both WPT+MSV and EMD+MSV it is difficult to verify whether the extracted SVs as features actually reflect rub-impact faults or whether their values are more closely related to the properties and setup of the system because these environment properties are also reflected in the acquired signal. Due to this concern, the interpretability of the features extracted in these two approaches suffers.

Regarding DWT+TDSIG, its performance for differentiating rub-impact faults of various intensities was the lowest compared to other methods used in this chapter. Table 2.5 shows that this method achieved an average FCA of 20,7% over 10 experiments. Also, the number of TP values was noticeably lower and the standard deviations of the performance metrics were generally higher (equal to 0.03) than those demonstrated by most of the techniques presented in this comparison. The confusion matrix in Figure 2.10(e) shows a similar result. Such poor performance can be explained as follows. First of all, choosing a wavelet function is not an easy task but is important because the quality of wavelet-based decomposition is directly affected by the chosen mother wavelet. Furthermore, the dataset used in this study contains two mechanical faults that affect the rotor system simultaneously: blade rub-impact fault and its cause, imbalance of the shaft. The wavelet function chosen in the referenced method may not match all the signal groups present in the dataset collected. This method also does not propose any technique for the adaptive selection of a proper wavelet level band that contains the most essential information about the mechanical fault after decomposition. Finally, the poor performance seems to be caused by the use of time-domain decomposed sample data as a feature vector to the OAAMCSVM classifier. Although SVM is known to be a robust classifier that is insensitive to the dimensionality of the feature vector [80], the use of a time-domain signal instead of statistical feature parameters creates a feature vector of huge dimensions. This leads the classifier to attempt to create a very complex hyperplane to try to separate the signal samples, causing the final classification accuracy to be poor. Despite, DWT incorporates downsampling while decomposing the original vibration signal, this has benefits only when the signal samples are relatively short (e.g., 2048 sampling points [44]). If the sampled vibration signal is long such as one collected in this experiment that contains 65,536 time

points, even after the downsampling performed by DWT, the sequence obtained at the third level of decomposition is still of huge dimension.

Overall, the proposed feature extraction and fault diagnosis methodology are highly useful for the diagnosing coupling shaft imbalance and blade rub-impact faults of various intensities because of its main concept - optimal IMF selection in EEMD with a novel objective function for assessing the quality of IMFs with a hybrid feature extraction procedure, based on a signal reconstructed using the selected IMFs in which fault symptoms are highly observable.

2.4 Conclusions

This chapter presented a new feature extraction method for coupling shaft imbalance and blade rub-impact fault diagnosis based on EEMD and novel optimal IMF selection procedure along with hybrid feature extraction. The proposed method addresses the problems of both efficient feature extraction and the selection of meaningful optimal IMF components suitable for diagnosing blade rubbing faults. First, EEMD provided well-separated and well-behaving oscillating components of the original vibration signals. Second, subsets of optimal signal-dominant modes were selected using a novel proposed criterion that is based on the DFP and KLD metrics. Finally, the adaptively selected optimal IMFs were used to reconstruct a noiseless vibration rubbing fault signal for hybrid feature extraction. In the experimental part of this chapter, the proposed hybrid feature model was used to differentiate coupling rotor imbalance and blade rub-impact faults. The experimental results demonstrated that the proposed approach is capable of extracting relevant information about rubbing faults of various intensity levels. Furthermore, the proposed methodology outperformed reference methods in terms of fault classification accuracy. Fault classification using the hybrid features extracted by the proposed method achieved an average fault classification accuracy of 99.2% over 10 experimental trials with various combinations of training and testing data. However, the drawback of this proposed IMF selection scheme is the arbitrarily selected threshold value along with the huge difference between the objective function values which can potentially lead to the selection of the intrinsic modes with a small amount of relevant information. This issue can be addressed by introducing the adaptive thresholding computation with some approach for normalizing the statistics of the objective functions. Specifically, the use of adaptive thresholding may reduce the number of selected IMF components for various classes, and hence, possibly lead to a reconstructed signal of better quality. This improvement is introduced in Chapter 3 of this dissertation.

Chapter 3

An Improved Algorithm for Selecting IMF Components in Ensemble Empirical Mode Decomposition for Domain of Rub-Impact Fault Diagnosis

3.1 Introduction

Turbines are one of the most critical rotating machines that are widely used in power plants. Their operation is usually related to severe working conditions with high rotational speeds and high temperatures. Turbine design engineers endeavor to keep a small clearance between the stator and turbine blades to reduce the air reluctance and increase torque. The interaction of the rotor blades with a stationary part causes a rubbing phenomenon. Rubbing can be considered as a separate fault or as a fault that may evidence the presence of other mechanical faults in rotating machines, such as self-excited vibrations, misalignment of the shaft, or blade extension due to the significant increases of operating temperatures [40]. If not detected and diagnosed in the early stages, rub-impact faults can lead to excessive damage of the rotating machines and a significant increase in maintenance costs. Thus, the detection of rubbing faults of various intensities is essential for the health and maintenance of rotating machines.

Rub-impact faults are known as complex non-linear and non-stationary mechanical faults [20]. Furthermore, conventional signal analysis methods that are developed for stationary and linear signals are not useful in diagnosing these faults. Such signal processing methods can be roughly split into time-domain approaches and frequency-domain methods (e.g., the fast Fourier transform).

For different types of complex faults, various types of data-driven [81–83], and model-reference-based [84], techniques can be applied for their fault diagnosis. However, in the case of rubbing fault identification problem, relatively simple, but powerful time-frequency analysis (TFA) methods are found to be very efficient for signal processing and diagnosing rub-impact faults due to the simultaneous analysis of the characteristics of the measured signals in both the time and frequency domains which allows for better discovering the complex and hidden features in signals. One of the widely applied TFA techniques for rub-impact fault signal processing relies on the wavelet transform and its variations [44, 85, 86]. The common advantage of wavelet-based signal processing methods is that they have good time-frequency localization, which allows for the detection of transients that may appear in signals. However, the main drawback of wavelet-based approaches is the need to select an appropriate wavelet kernel function (mother wavelet function), which can directly affect the quality of the signal decomposition [87]. In most cases, the selection

of the mother wavelet function requires a series of experiments and analysis of the decomposition results [47, 48]. This kind of selection procedure is prone to subjectivity. Additionally, the specific frequency band in which the rotating machine fault symptoms are localized needs to be determined for appropriate analysis of the wavelet decomposition results.

Recently, another TFA technique, called empirical mode decomposition (EMD) [22], has been utilized for rubbing signal analysis. EMD is a robust decomposition algorithm that is capable of decomposing complex and non-linear multi-component signals into a finite number of well-behaved intrinsic mode functions (IMFs), in which each of the modes corresponds to a specific frequency band containing the natural oscillatory contents of the original signal. The significant advantage of EMD over wavelet transforms is that EMD is a completely adaptive and data-driven decomposition method that does not require any specific prior knowledge about the nature of the signal. Due to its advantages, EMD has been successfully applied in various fields of rotating machinery fault diagnosis, such as for bearings, gears, and rotors [52]. In several studies [56–59, 88, 89], EMD is applied to observe and analyze valuable information from rubbing fault signals. The studies mentioned above demonstrate the strength and capabilities of EMD applied to complex signal analysis. However, this algorithm suffers from a problem called the ‘mode-mixing’ problem. In the mode-mixing problem, the presence of different oscillating components in one IMF or the appearance of similar oscillations in different modes must be understood. This shortcoming causes difficulties when trying to find a physical interpretation for each of the extracted modes, which is important for solving fault diagnosis problems.

To resolve the problem of mode-mixing, an improvement known as ensemble empirical mode decomposition (EEMD) was introduced [23]. The idea behind this improvement is to obtain ‘real’ IMFs that are the mean of EMD trials run on the signal with various realizations of white Gaussian noise added in each trial. Due to this property, EEMD is also known as a noise-assisted signal analysis method. Thus, EEMD has been employed for fault diagnosis of rolling-element bearings [61] and rub-impact fault diagnosis [63].

However, there remains one issue common to all EMD-based approaches in general, and EEMD specifically, that must be resolved. It is crucial to determine the valuable IMF components that contain important information about the mechanical fault being investigated.

In real scenarios, the cardinality of the extracted IMFs is large, and not all of them are equally important for fault diagnosis. Moreover, these extracted components can be either signal- or noise-dominant. Thus, it is essential to find and select discriminative modes from the set of extracted IMFs that carry the most valuable information and are the most useful for diagnosing mechanical faults. To select sensitive IMF components for rub-impact fault diagnosis, Lei et al.

[65] computed the ‘sensitivity factors’ of the extracted IMFs by using a correlation coefficient between the original signal and extracted modes, and the signal corresponding to the normal operating conditions. This method demonstrates good results for IMF selection. However, selection based on the correlation properties of the extracted components and the signals, without analysis of the information content included in the various IMFs, might not be an appropriate strategy for situations in which the original signal is highly corrupted by noise. In Chapter 2 of this dissertation, the algorithm suitable for IMF selection in the domain of blade for rub-impact fault diagnosis has been introduced. In that chapter, a specific criterion was derived for selecting valuable IMFs that is based on the ratio between the ‘degree-of-frequency presence’ (DFP) of information related to rubbing faults (frequency harmonics) and the Kullback-Leibler divergence (KLD) [90] information-based distance metric. This approach demonstrates good capabilities in selecting informative IMF components for rub-impact fault diagnosis; however, the applied arbitrary thresholding (e.g., 1) for creating a subset of valuable modes is a weak point of this methodology. The subjective IMF selection with a constant fixed threshold value may cause the selection of some noise-contaminated IMFs or the selection of modes with a very small presence of frequency harmonics related to rubbing faults. One of the reasons for this is that the average values of the objective function computed for various IMFs grow with increasing intensity of the rubbing fault signal. Thus, as the intensity of rubbing increases, it becomes easier for noisy or non-informative IMF components to exceed the arbitrary threshold and become a part of a subset containing the selected modes. Another issue may appear while computing the objective function value itself. For instance, when the numerator (DFP) and denominator (KLD) are equal, the objective function becomes equal to 1, which is equal to the previously proposed arbitrary threshold value. Additionally, the threshold value can be accidentally exceeded by the objective function in some extreme cases, such as when the numerator is relatively small (low presence of valuable information about rub-impact) and the denominator value is high (the IMF component is noise-dominant). Therefore, this chapter of the dissertation addresses these issues and presented an improved DFP/KLD-based optimal IMF selection algorithm for EEMD by introducing two modifications. The first is the normalization of objective function values to deal with the extremal cases, which appear while computing the objective values for IMF components. The second technique is the adoption of an adaptive thresholding technique to enhance the selection procedure. The signal reconstructed using the valuable IMFs selected by the improved algorithm contains less high-frequency noise in comparison with the original technique, while preserving the frequency components that are important for the rubbing fault diagnosis.

The remainder of this chapter is organized as follows. Section 3.2 provides a brief background on the DFP/KLD-based IMF selection algorithm for EEMD. Section 3.3 contains the detailed problem statement and the proposed improvements of the original IMF selection

procedure. Section 3.4 presents the experimental verification of the improved algorithm. Finally, Section 3.5 concludes this chapter.

3.2 Brief Background on the Original DFP/KLD-based Approach for Selecting Optimal IMFs in EEMD

The algorithm of EEMD has been introduced in Chapter 2 Algorithm 2.1 of this dissertation and the results of EEMD decomposition of the signal containing severe blade rub-impact fault with rotor imbalance fault are depicted in Figure 3.1. From this figure, it can be seen that the cardinality of the extracted intrinsic mode components is large. However, if observe the amplitude values of the IMFs in the time and frequency domain as well as frequency contents presented in the envelope power spectra of the extracted intrinsic modes it becomes clear that not all of the obtained components can contribute equally to the solution of the blade rub-impact fault diagnosis problem. Thus, in this example, it can be observed that IMFs #9 and #10 carry important information, and their behavior in the time domain can be associated with the rotational frequency. Moreover, the low-frequency contents presented in their envelope power spectra have the highest amplitude values among the other extracted IMFs. At the same time, it can be seen that the amplitudes of the frequency harmonics observed in frequency spectra of the remaining intrinsic

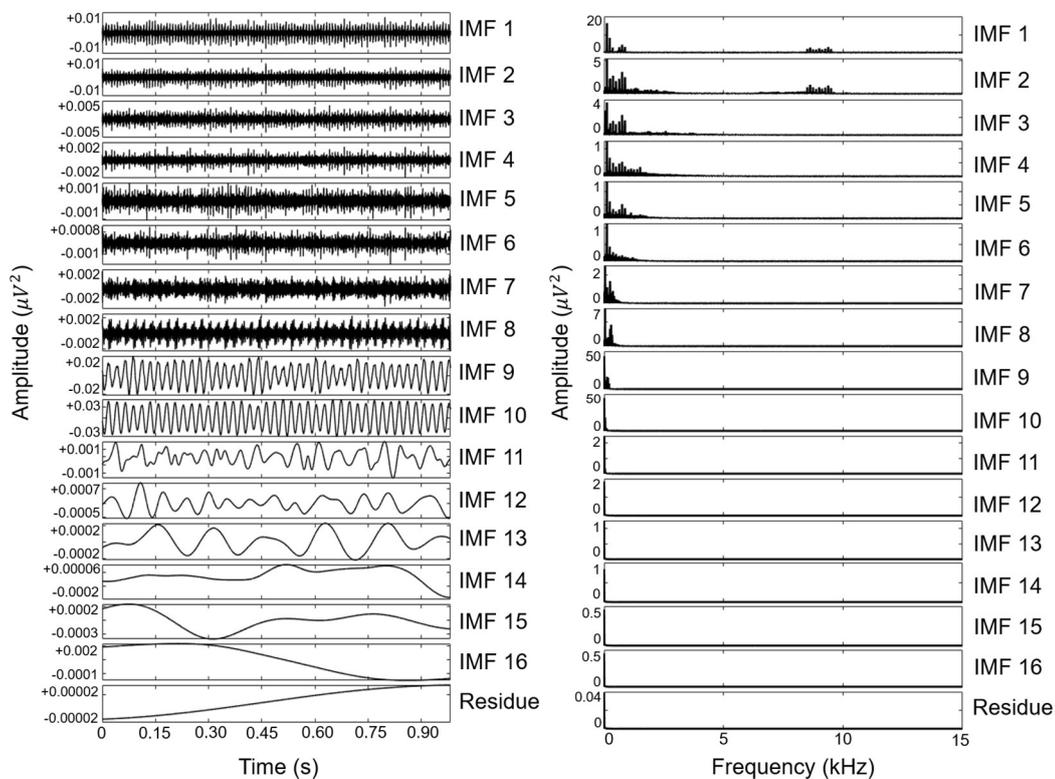


Figure 3.1. The IMFs and their envelope power spectra obtained as the result of EEMD decomposition of the signal containing severe rub-impact fault.

modes are significantly smaller than the ones observed in IMFs 9 and 10. Also, it can be seen that some of the extracted components are containing the presence of the high-frequency noise (mostly can be observed in low-order IMFs). From these observations, it can be concluded that the appropriate analysis and careful fusion is essential for solving the problem of rotating machine fault diagnosis in general and rub-impact fault diagnosis specifically when using EEMD. If the extracted components are not selected carefully, the noise-dominated and not informative IMFs (the ones that are not related to the mechanical fault) can cause the degradation of fault identification performance.

In general, the informative signal-dominant intrinsic modes can be selected by applying a two-step procedure. The first step is to define the evaluation criteria (objective function) to grade each of the extracted components based on their relevance to the fault under investigation, and the second is to select a subset of valuable modes using a threshold applied to the objective function. In Chapter 2 of this dissertation, a ratio between the DFP of rub-impact fault harmonics to the KLD of the IMFs has been introduced as the objective function for grading the extracted components. After obtaining the values of an objective function for each of the extracted modes, those values are compared with the arbitrary threshold to create a subset of optimal IMFs that are further used for a partial vibration signal reconstruction with a low presence of noise frequency harmonics and high presence of harmonics related to rub-impact faults.

The DFP is a first component of the objective function defined in the previous chapter for grading optimal IMFs, due to its property of detecting the appearance and power of the specific frequencies of interest (i.e., rubbing fault frequencies) and their harmonics in the envelope power spectra of the extracted intrinsic modes. Hence, the DFP is used to quantify each of the modes based on the amount of valuable information presenting in them. As specific frequencies (i.e., fault frequencies), the frequency harmonics corresponding to $1/3X$, $1/2X$, $2/3X$, $1X$, $4/3X$, $3/2X$, and $5/3X$, where $1X$ states for the fundamental frequency, are chosen due to their relevance to rubbing processes [49, 50, 73, 91, 92]. The DFP is computed in Eq. (2.8) as the ratio between the components of specific frequencies of interest (FOI) corresponding to rubbing faults and the residual frequency components that are not related to rubbing.

The second component of the evaluation criteria is the KLD [93] computed between the probability density functions (PDFs) of the extracted intrinsic modes and the original signal using the following formulation in Eq. (2.9). The KLD values are varying in the range between zero and one, where values close to zero indicate that the PDFs of the k^{th} IMF and the original signal are close to each other; whereas the values that are close to one state otherwise.

The complete objective function for grading the k^{th} IMF based on its relevance to rubbing faults is defined as a ratio between the DFP and KLD, as shown below:

$$Obj_k = DFP_k / KLD_k. \quad (3.1)$$

Here, DFP_k is the sum of all “degrees-of-presence” computed for each N_{wfr} and their harmonics (up to three) for the k^{th} IMF, and KLD_k is the metric that shows how similar the probability density function of the k^{th} intrinsic mode is to one of the original signals.

Finally, the selection of valuable IMF components is accomplished by comparing the values of the objective functions with the threshold value (arbitrarily assigned to 1 [91]). The IMFs with objective function values less than the threshold are excluded from the partial signal reconstruction of the denoised rub-impact fault signal.

3.3 Improved IMF Fusion Algorithm in EEMD for Rubbing Fault Diagnosis

In Figure 3.2, the block diagrams of the original and improved DFP/KLD-based IMF selection techniques for rub-impact fault diagnosis are presented. As shown in this figure, the improved intrinsic mode selection includes two new steps in comparison with the original algorithm. Specifically, once the objective functions (Obj) are computed, these objective function values are normalized by using MinMax transform, and then, the normalized objective values ($nObj$) are compared with the adaptive threshold ($adTh$) value to create a subset of valuable IMFs. Finally, the new rubbing signal is obtained by partial reconstruction using the subset of chosen modes.

3.3.1 The Drawbacks of the Original DFP/KLD-based Optimal IMF Selection

Approach

The original DFP/KLD-based optimal IMF selection technique, introduced in Chapter 2 and briefly described in Section 3.2, allows for the selection of informative components that contain frequency components related to blade rub-impact faults. However, this approach has a drawback that must be resolved. Specifically, the use of an arbitrary thresholding value cannot always be robust for the selection of valuable components when the rubbing fault intensity changes significantly. When the rubbing intensity of the signal increases, the signal becomes more complex and the energy level of the signal drastically increases. This may lead to an increase in the average DFP values computed from the extracted IMFs. This growth occurs when the intensities of the rotor imbalance and consequent blade rubbing fault are increasing. . While the numerator grows,

the values of KLD do not change significantly, and thus, with the increased values of DFPs, the average value of the objective function for each of the IMFs increases. This also may cause situations in which noise-dominant IMFs or components that do not contain valuable information about rubbing processes can be improperly added to the optimal subset of informative modes when the objective values are compared with the arbitrarily chosen threshold. Additionally, the previously introduced threshold can lead to the selection of IMF components as faults in some ‘exceptional’ cases that may occur during the computation of objective function values.

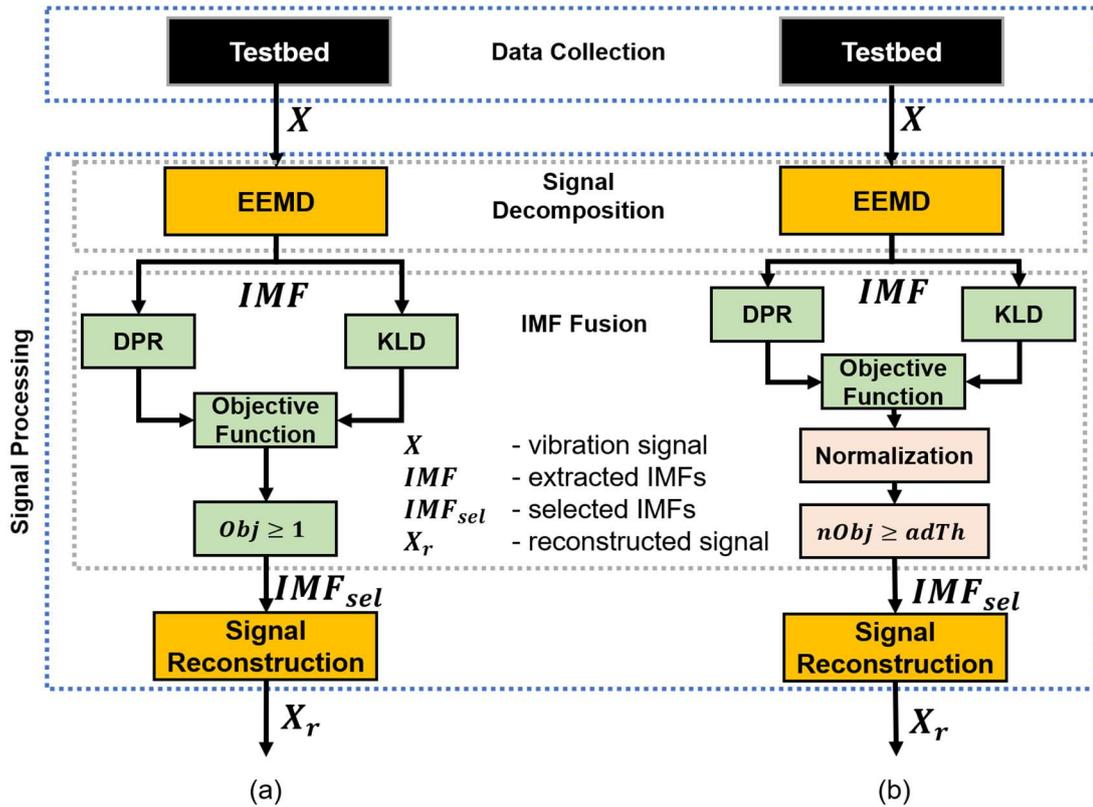


Figure 3.2. Block diagram of the (a) original and (b) improved DFP/KLD-based IMF selection techniques for rub-impact fault diagnosis.

Table 3.1 illustrates an example of the original IMF selection procedure for a signal containing a severe rub-impact fault (class 10 in the collected dataset). This table contains the separate values of the DFP and KLD components of the DFP/KLD-based IMF selection metric as well as the computed values of the objective function for each of the extracted intrinsic modes. According to the original IMF selection procedure, the threshold value for selecting meaningful intrinsic modes is set to 1 in this example.

From Table 3.1, it can be observed that when the thresholding value is equal to 1, the components with order numbers #10, #9, #8, #3, #7, #2, #1, and #6 are selected as valuable IMFs; whereas the components #14, #11, #15, #16, #17, #13, #12, #4, and #5 are excluded from signal reconstruction as modes that do not benefit rub-impact fault diagnosis. However, one can see that among these eight selected components, not all of the selected IMFs are equally important for

Table 3.1. DFP, KLD, and objective values obtained by the original IMF selection metric for the signal sample that contains a severe rubbing fault.

IMF #	DFP	KLD	Obj
10	14.5	0.19	77.0
9	8.6	0.25	34.2
8	17.9	0.8	22.2
3	20.2	0.99	20.3
7	9.5	0.84	11.2
2	3.6	0.8	4.5
1	1.2	0.9	1.2
6	0.2	0.2	1.09
14	0.2	0.5	0.44
11	0.32	0.93	0.33
15	0.24	0.94	0.25
16	0.24	0.99	0.23
17	0.24	0.99	0.23
13	0.23	0.99	0.23
12	0.18	0.96	0.19
4	0.12	0.7	0.16
5	0.09	0.97	0.09

The order numbers of the IMF components in the heading of this table are organized in descending order with respect to the objective function value. The order numbers and objective function values highlighted with bold font indicate the chosen valuable IMF components using the original thresholding method.

rubbing fault diagnosis. Specifically, even though IMF #1 does not contain a considerable presence of relevant information corresponding to the rubbing process and the KLD distance value between this component and the original signal is high, the value of the objective function (the ratio DFP/KLD) became bigger than the value of the arbitrarily assigned threshold in the original selection approach. Moreover, it is known that the low-order modes are usually dominated by high-frequency noise and contain miserable energy content. This is one of the extreme cases that may occur during the computation of the objective function and IMF selection by the originally introduced DFP/KLD-based approach. The second example of another extreme case can be observed in IMF #6. Here, the DFP and KLD values are almost equal, and thus the computed objective function appeared to be slightly higher than the threshold value. However, this intrinsic mode does not contain a lot of meaningful information about rubbing processes according to a low value of DFP. Regarding IMFs #7 and #2, these modes definitely carry some valuable frequency content, but the KLD values are large which means that these intrinsic modes are highly contaminated by noise. In the case of IMF #2, this mode is also a low-order component that might be affected by noise, and, compared to the other selected modes, the presence of important information related to rubbing processes according to the DFP component of this IMF is not high. This IMF includes some valuable information; however, based on KLD, this component appears to be noise-dominant. IMF #7 also appears to be noise-dominant; even its DFP value is quite high compared to IMFs #2, #1, and #6. Of course, from Table 3.1, components #8 and #3 also can be considered as noise-dominant ones, but the strong presence of important information relevant to rubbing (the DFP metric) makes these components valuable for signal reconstruction because the exclusion of them may lead to the loose of valuable rubbing signal contents.

The analysis of the IMF selection results shown in Table 3.1 demonstrates that the use of the original threshold value that was arbitrarily assigned to 1 may not always be appropriate for some extreme cases. Additionally, some cases investigated in this subsection may cause improper IMF selection during the objective value computation when the values of both of the components included in the objective function are nearly equal (IMFs #6 and #1).

3.3.2 Objective Function Normalization and Adaptive Thresholding Computation For Improving the Optimal IMF Selection Procedure

To reduce the influence of the growth of average DFP values on the objective function computation and to cope with the problems caused by the application of the fixed arbitrary threshold for IMF selection, this subsection introduces some modifications for improving the original selection approach. The proposed improvement consists of normalization to stabilize the statistical properties of the objective function values computed for each intrinsic mode function and adaptive thresholding calculation for constructing the subset of optimal IMFs. First, the

objective function values, computed for each of the extracted modes, are normalized using MinMax scaling [94]. Then, the adaptive threshold value is computed using the statistics of the normalized objective values to perform the selection of valuable components by applying the decision rule presented in this subsection.

To compute an adaptive threshold, the normalization of objective values is essential because, with the increased intensities of the rubbing processes, drastic changes in the energies of the acquired signals can be observed. When the signal corresponding to a rubbing fault of high intensity is decomposed into a set of modes, the computed objective functions for grading those IMFs are highly varying in magnitude, as was demonstrated in Table 3.1. The motivation beyond the use of MinMax scaling is to score the appropriate IMF component with the highest non-normalized objective value as 1 (assigning it as a maximum), while the remaining objective values are expressed as a fraction of the best intrinsic mode score. This type of normalization does not disrupt the original order of the extracted modes, while also efficiently neglecting the influence of drastic energy changes in the rubbing signal on the threshold computation. Normalized objective values ($nObj$) are less affected by a sharp growth of the mean value of the objective functions, and thus, the statistical properties of the normalized objective values become smoother, which is favorable for adaptive threshold calculations. The normalization using MinMax scaling can be performed using the formulation shown below:

$$nObj_k = \frac{Obj_k - \min(Obj)}{\max(Obj) - \min(Obj)}, \quad (3.2)$$

where k is the order number of the particular IMF, Obj corresponds to the set of objective values computed for each IMF of a particular signal sample using the original DFP/KLD criterion, and Obj_k is the objective function value of the k -th intrinsic mode component.

Once the normalized objective function values are computed, an adaptive thresholding approach can be used to select the valuable IMF components. Inspired by the concept of the widely applied ‘universal threshold’ from the field of signal denoising, the following formulation is proposed to compute the adaptive threshold value for IMF selection:

$$adTh = var \sqrt{2 \log N}. \quad (3.3)$$

Here, var is the variance of the normalized objective function values and N is the total number of intrinsic modes computed for a particular vibration signal sample. This formulation can match well the purposes of thresholding objective functions since it corroborates the statistical properties of the IMFs delivered for a decomposed signal, such as the variance of normalized objective values

and the total number of objective functions. From (3.3), it is clear that if the objective functions are not properly scaled, very few components with high magnitudes will affect the statistical properties of the total set of extracted components and dominate all the remaining ones which will lead to neglecting the influence of all the extracted IMFs.

The decision on selection of each specific informative mode for further partial signal reconstruction can be performed using a decision-making scheme based on the scaled objective functions and computed adaptive threshold, as follows:

$$\text{SelectIMF}_k = \begin{cases} 1, & nObj_k \geq adTh \\ 0, & nObj_k < adTh \end{cases} \quad (3.4)$$

The improved result of the enhanced IMF selection method including the original and normalized DFP/KLD objective functions computed for each of the extracted IMFs is demonstrated in Table 3.2, which uses a signal sample with a severe rub-impact fault. In this table, unlike Table 3.1, the adaptive thresholding method is used for optimal IMF selection.

From Table 3.2, one can observe that the application of the proposed adaptive thresholding technique with the objective function normalization is capable of reducing the number of selected IMFs by excluding the components that were improperly selected by the original method. Specifically, after the intrinsic mode selection procedure is accomplished by the improved algorithm, only four components are selected for signal reconstruction: #10, #9, #8, and #3. The remaining intrinsic mode components, including the IMFs #6, #1, #2, and #7 that were improperly selected by the original approach are ignored during the signal reconstruction step. Referring to Table 3.1, the IMFs #6 and #1 were mistakenly selected due to the extreme cases that may appear while applying the original threshold value equal to 1 to the not normalized objective function values of the IMF set. Moreover, the components with order numbers #2 and #7, which are contaminated by noise and do not contain much valuable information related to rubbing faults, are also excluded from the signal reconstruction process because their normalized objective function values appear to be lower than a new adaptively computed threshold value. The differences between the selection results achieved using the original objective function with arbitrarily fixed

thresholding and the normalized objective values with an adaptive one are visualized in detail in Figure 3.3.

Table 3.2. Original and normalized objective values obtained for the signal sample that contains a severe rubbing fault.

IMF #	<i>Obj</i>	<i>nObj</i>
10	77.0	1
9	34.2	0.44
8	22.2	0.28
3	20.3	0.26
7	11.2	0.14
2	4.5	0.05
1	1.2	0.015
6	1.09	0.014
14	0.44	0.005
11	0.33	0.004
15	0.25	0.003
16	0.23	0.003
17	0.23	0.003
13	0.23	0.003
12	0.19	0.002
4	0.16	0.002
5	0.09	0.001

The order numbers of the IMF components in the heading of this table are organized in descending order with respect to the objective function value. The order numbers and objective function values highlighted with bold font indicate the chosen valuable IMF components using the new computed adaptive threshold (equal to 0.15 in this example).

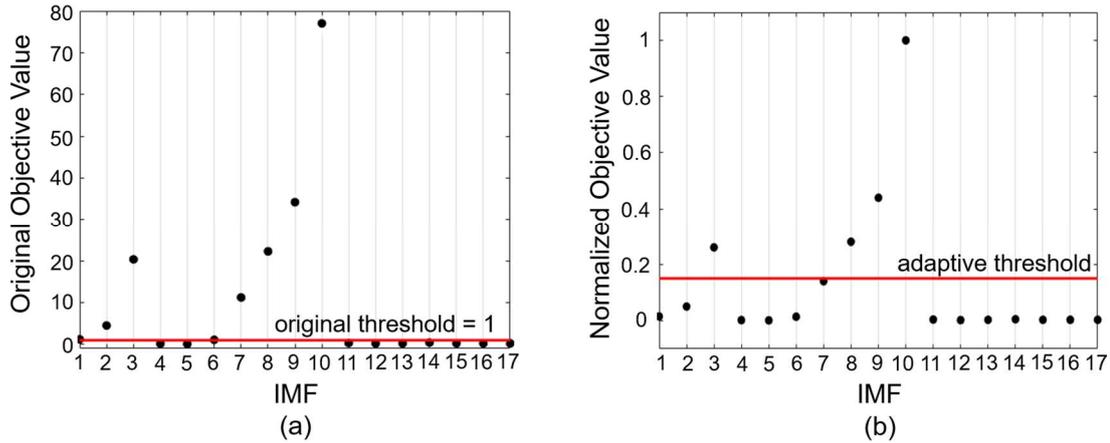


Figure 3.3. IMF selection results achieved by (a) the original objective values with arbitrary thresholding and (b) the normalized objective values using MinMax scaling with an adaptive threshold value.

3.4 Experimental Results and Discussion

In this section, the denoising capabilities and verification of the effectiveness of the proposed improved selection algorithm over its predecessor using a benchmark rub-impact dataset presented in detail in Chapter 2 of this dissertation are investigated.

3.4.1 Analysis of Signal Denoising Properties

The modified IMF selection method is capable of a superior selection of highly informative signal-dominant components that can be used for partial signal reconstruction, allowing for the creation of a new denoised signal, which makes the fault symptoms highly observable. To evaluate the quality of the improved optimal IMF selection approach and the original one, a comparison of the envelope power spectra of the partially reconstructed signals using the components selected by both of the DFP/KLD-based IMF selection techniques is performed. The signals analyzed in this section are acquired from the vibration sensor installed on the DE of the shaft using the channel that collects the displacements of the shaft in the horizontal direction. Figure 3.4 depicts the comparison of the original signals with the reconstructed ones in the time-domain, as well as the comparison of their envelope spectra. The signal plots in this figure correspond to the different rubbing intensity levels caused by ‘0.0 g’ (class #1), ‘1.6 g’ (class #5), ‘2.0 g’ (class #8), and ‘2.8 g’ (class #10) of additional weight attached to the rotor.

Note that in this figure, only some of the data instances corresponding to the classes mentioned above are represented for demonstration purposes. The exact IMF selection results may vary for different samples within the same signal class due to the deviations in objective function values. Figure 3.4 (a) shows that for class #1 (0.0 g of extra weight) when there is no rubbing fault presented in the signal and no shaft imbalance occurred, both selection approaches may deliver modes with similar frequency contents. However, from Figure 3.4 (b), (c), and (d), it is seen that the optimal intrinsic mode selection method modified with objective function normalization and adaptive thresholding delivers a more clearly reconstructed signal, allowing exclusion of the noise in frequency bands higher than 15 kHz and showing a significant reduction of noise in the region around 10 kHz.

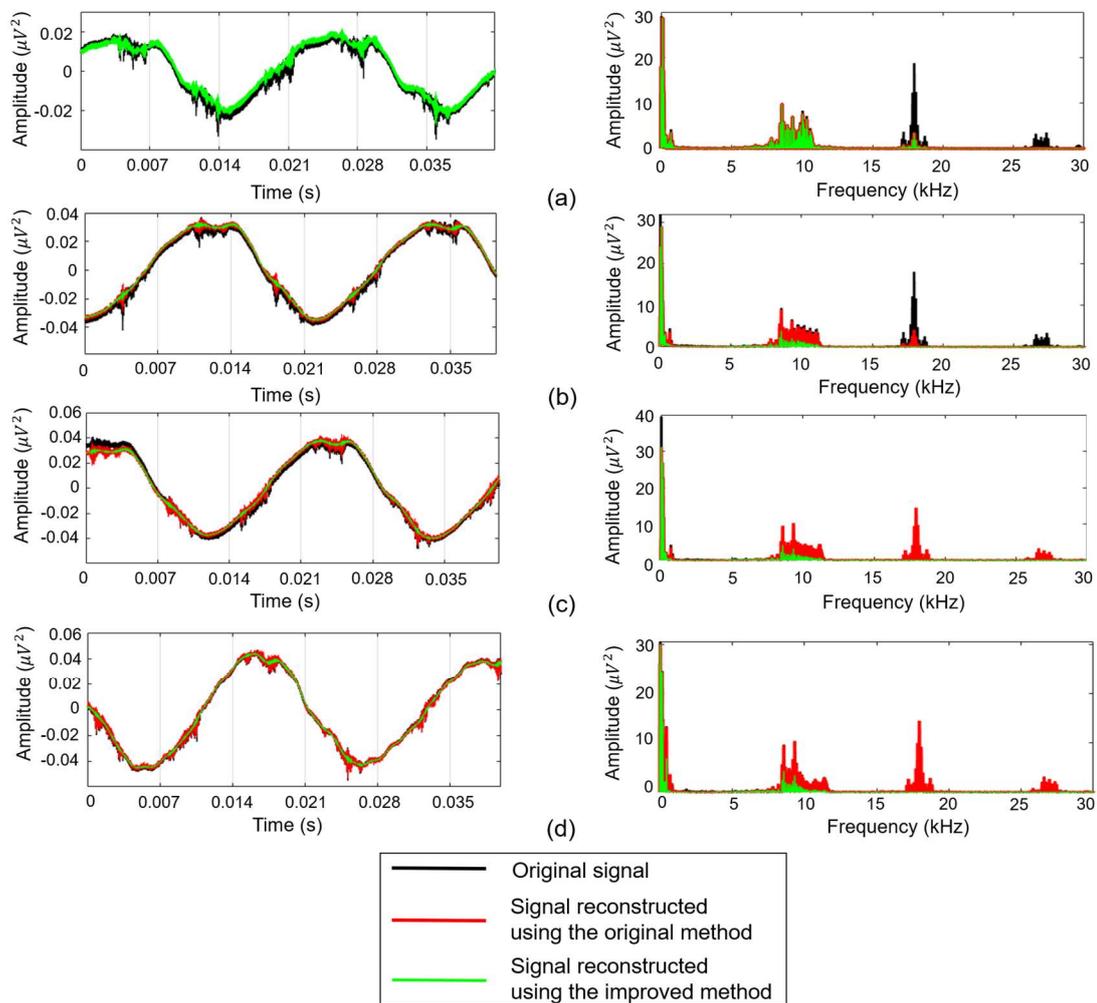


Figure 3.4. Comparison of original time-domain vibration signals against reconstructed ones using both optimal intrinsic mode selection techniques and their envelope power spectra for the signals of (a) class #1, (b) class #5, (c) class #8, and (d) class #10, respectively.

In Figure 3.5, the envelope power spectra scaled in a low-frequency zone (from 0 Hz to 500 Hz), corresponding to the same reconstructed signals as in Figure 3.4, are presented. From Figure 3.5, it can be observed that the signals reconstructed after improved optimal IMF selection preserve the relevant rubbing frequency harmonics, such as $1/3X$, $1/2X$, $2/3X$, $1X$, $4/3X$, and $5/3X$ [49, 50, 73], whereas the amplitudes of these harmonics do not differ significantly from the ones observed in signals reconstructed by the original optimal IMF selection method. The presence of these frequency harmonics is essential for rub-impact fault diagnosis because they are considered valuable features of the rubbing processes.

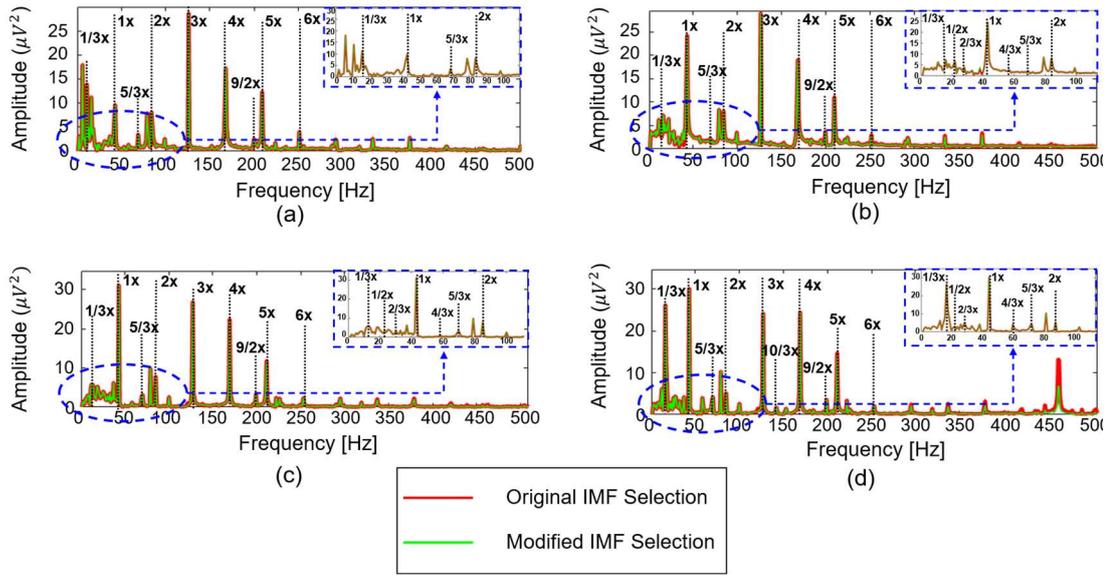


Figure 3.5. Envelope power spectra of signals reconstructed using both the original and improved intrinsic mode selection techniques for signal (a) class #1, (b) class #5, (c) class #8, and (d) class #10 for frequencies ranging from 0 Hz to 500 Hz.

Furthermore, to quantitatively evaluate the quality of signals reconstructed after the improved optimal IMF selection and filtering performance, the reversed signal-to-noise ratio (rSNR) and mean square error (rMSE) metrics [95, 96] were employed in this study:

$$rSNR = 10 \log_{10} \left(\frac{\sum_{n=1}^N (x(n))^2}{\sum_{n=1}^N (x(n) - x_{rec}(n))^2} \right) \quad (3.5)$$

$$rMSE = \frac{1}{N} \sum_{n=1}^N (x(n) - x_{rec}(n))^2. \quad (3.6)$$

Here $x(n)$ is the original noisy vibration signal and $x_{rec}(n)$ is a signal reconstructed using the subset of optimal IMFs. The motivation for using these metrics is as follows: the conventional SNR and MSE computation requires the availability of the original noiseless signal (ground true

signal), which is unknown in most of the real scenarios. However, the noisy vibration signal and the reconstructed signals are available. Thus, by replacing the noiseless signal with an available noisy one, the traditional SNR and MSE metrics can be re-formulated as rSNR and rMSE, respectively. Interpretation of the rSNR and rMSE results is opposite to conventional SNR and MSE: the smaller rSNR and bigger rMSE exhibit a smaller presence of noise and better quality of the reconstructed signal.

Figure 3.6 presents the average rSNR and rMSE values for all the signal groups computed via the signals reconstructed using both optimal intrinsic mode selection methods. From Figure 3.6 (a), it can be observed that for all the signal classes except the first one, in which no rubbing or shaft imbalance was present, signals reconstructed using the components selected by the improved technique demonstrate better results in terms of the average rSNR (a smaller rSNR is better). In particular, this improvement is more significant for the class in which a severe rub-impact fault is observed. From Figure 3.6 (b) it can be seen that for most of the classes, signals reconstructed using the components chosen by the improved IMF selection algorithm demonstrate higher average rMSE values (in terms of rMSE, larger is better) than ones reconstructed using the original approach presented in Chapter 2. However, this behavior pattern is different for the first class, which represents the signals acquired when no extra weight was attached to the shaft. Thus, this state corresponded to normal operating conditions when neither shaft imbalance nor blade rubbing fault could be observed in the rotor system. This result can be explained as follows. Since the criterion developed for selecting optimal IMF components aims to detect the signal-dominant intrinsic modes that contain the most valuable rubbing fault information, this criterion itself might not be completely appropriate for the selection of intrinsic mode components when the system operates under normal conditions, and thus, the cardinality of the selected components may vary significantly from sample to sample.

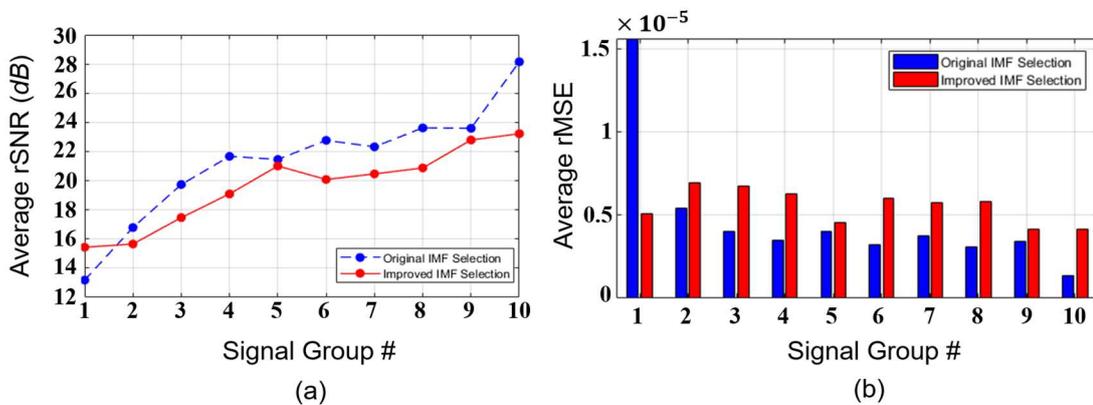


Figure 3.6. Average values of (a) rSNR and (b) rMSE computed for the reconstructed signals obtained by different IMF selection methods.

Furthermore, more details to the results provided in Figure 3.6 are presented to observe the effect of the original IMF selection formulation and the proposed one on the reconstructed signals. According to Figure 3.6 (a), the average rSNR values for classes #5 and #9, which correspond to the slight and intensive rubbing conditions with extra weights equal to 1.6 g and 2.4 g added to the NDE, respectively, are very close to each other. Moreover, for class 1, corresponding to the condition where neither rubbing nor shaft imbalance faults were observed in the system, the average rSNR value is lower for the original method. For the analysis, some of the samples belonging to these troublesome cases are decomposed into a set of IMF components, and both intrinsic mode fusion algorithms are applied to select the best components for signal reconstruction. The samples of the original time-domain vibration signals corresponding to the identified signal groups and the envelope power spectra of the reconstructed signals using the IMFs selected by both approaches are demonstrated in Figure 3.7.

In Figure 3.7 (a), data sample #31 from the signal group, corresponding to a normal operating mode of the rotor system (class #1), as well as the envelope power spectra of reconstructed signals using the optimal subsets of the IMF components selected by both approaches, are presented. From the operating mode, it is known that the rotating machinery was working normally, so the power spectra mostly contain the harmonics corresponding to shaft rotation frequency. In this condition, both the original and the improved IMF fusion methods delivered almost the same modes, except the improved one delivered an additional component – IMF #10. From the envelope power spectra, one can observe that the frequency contents of both reconstructed signals are approximately similar. However, despite the original technique selected only four components, the peak amplitude of high-frequency noise detected around the 8 kHz frequency bin was slightly smaller than that observed in the envelope power spectra of the signal reconstructed using the modes selected by the improved approach. It is clear that the amplitude of the harmonic at the frequency bin corresponding to 7 Hz is drastically higher than all those corresponding to the fundamental frequency or its high-order harmonics. This peak dominates the entire power spectrum, which makes the signal reconstructed via the proposed IMF fusion technique more favorable for feature extraction, especially when the dimensional features are used. The performance is better than the traditional approach despite the intensity of the noise present is slightly higher. Figure 3.7 (b) presents sample #34, corresponding to a slight rubbing fault condition and the power spectra of its reconstructed signals. Compared with the original IMF selection technique, the proposed one delivered a smaller number of intrinsic modes. Despite a smaller number of selected components, the frequency harmonics indicating the presence of rub-impact faults and their higher-order components located in the low-frequency range of the reconstructed signal are equal to that of the original intrinsic mode component selection technique.

The main difference in these presented power spectra is the level of high-frequency noise. After the proposed IMF fusion, the power spectra of the reconstructed signal demonstrate the absence of high-frequency noise in the frequency range between 15 kHz and 20 kHz. Additionally, the peak amplitude of noise detected around the 8 kHz frequency bin is almost 2.5 times smaller than the one observed in the envelope power spectrum of the reconstructed signal after application of the original selection technique. Figure 3.7 (c) depicts the signal reconstruction results for data sample #13, corresponding to an intensive rub-impact fault condition when 2.4 g of extra weight is attached to the rotor of the testbed. The analysis of this case shows similar results, as presented in Figure 3.7 (b): after implementing the improved optimal IMF selection algorithm, less intrinsic modes are selected compared to the original algorithm and the amount of high-frequency noise is

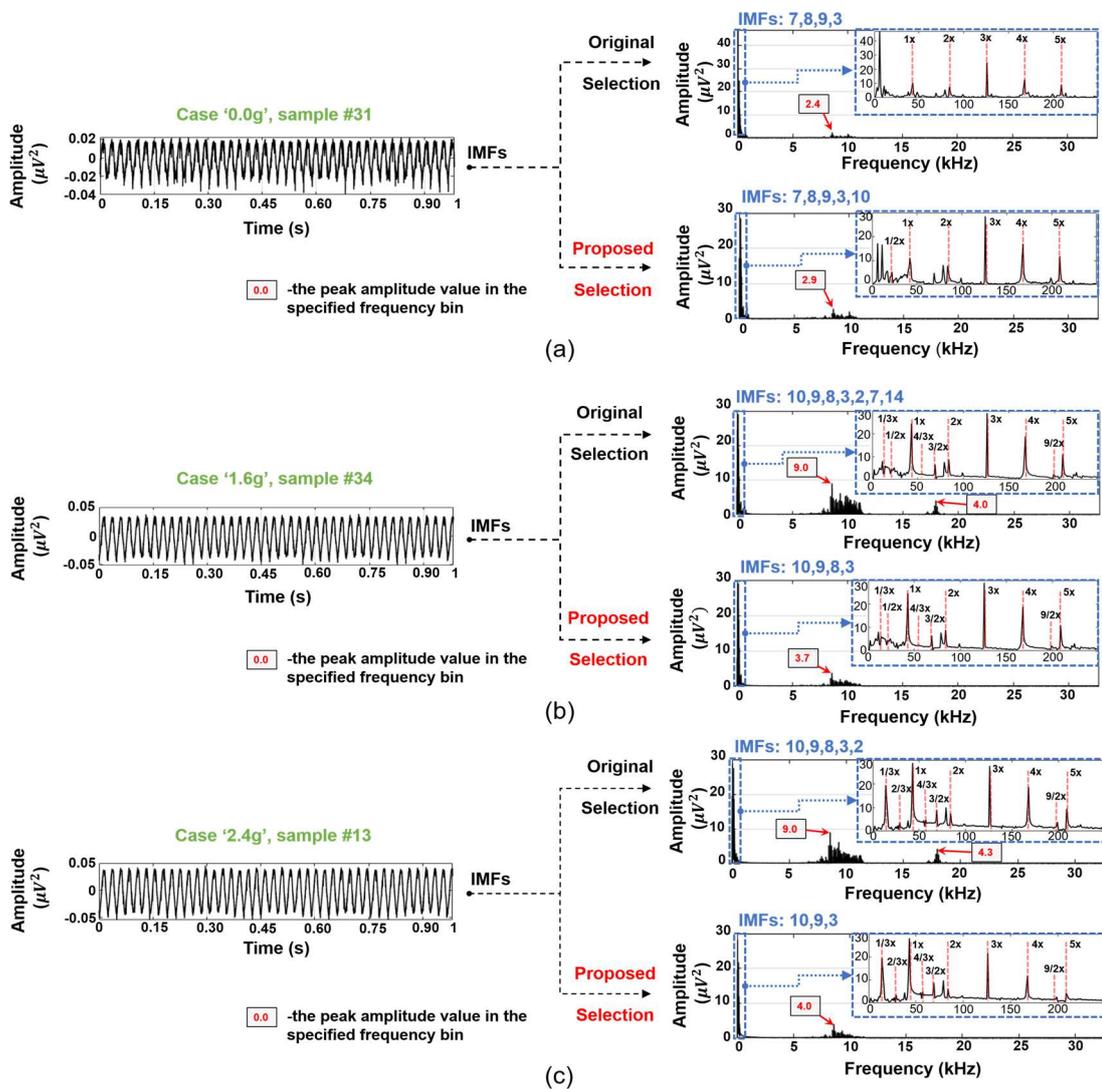


Figure 3.7. Original time-domain vibration signals and the envelope power spectra of the reconstructed signals through original DFP/KLD-based IMF selection and the proposed one corresponding to: (a) sample #31 from #1, (b) sample #34 from #5 (c) sample #13 from #9 signal classes.

significantly reduced. Meanwhile, the valuable harmonics caused by the rubbing faults are still present in a low-frequency range of the power spectrum without drastic changes in the amplitudes of frequency peaks.

3.4.2 Fault Diagnosis Performance Analysis

In this subsection, the improvements of the proposed modified IMF selection algorithm in comparison with the original one in terms of fault diagnosis are investigated. For this, the hybrid feature model proposed for diagnosing rub-impact faults in Chapter 2 of this dissertation is extracted from the signals reconstructed using the optimal IMF components selected by the improved intrinsic mode selection method introduced in this chapter (further referred to as Proposed+HFM) and from the signals reconstructed using the original approach (further referred to as Original+HFM). For a fair comparison, the same machine learning classifier and the same performance metrics as utilized in Chapter 2 are used in this experiment. To ensure the repeatability of the results and exclude the effect of randomness, the experiments were performed 10 times and in each of the experiments, the unique training and testing subsets have been sampled randomly at a ratio of 8:2. The experimental results of fault diagnosis expressed in terms of Rec_{μ} , $Prec_{\mu}$, $F1_{\mu}$, and FCA are averaged over ten experiments and tabulated in Table 3.3.

Table 3.3. Experimental results.

Methods	Metrics (Std) (%)			
	Rec_{μ}	$Prec_{\mu}$	$F1_{\mu}$	FCA
Proposed + HFM	99.8 (0.3)	99.8 (0.3)	99.8 (0.3)	99.8 (0.3)
Original + HFM	99.2 (0.9)	99.2 (0.9)	99.2 (0.9)	99.2 (0.9)

The boxplots containing the statistics of FCA values obtained during the experimental trials by both of the methods are presented in Figure 3.8. The black cross in boxes corresponds to the mean FCA value demonstrated in Table 3.3. From this figure, it can be observed that the proposed improvement of optimal IMF selection algorithm allowed the Proposed+HFM fault diagnosis framework to demonstrate more stable experimental results with very few outliers in comparison to the Original+HFM approach where the previous IMF fusion technique has been used.

Additionally, the confusion matrices containing the detailed results averaged over 10 experiments for both of the techniques are presented in Figure 3.9. The results demonstrated in this figure show that the proposed improved approach allows for reducing the number of misclassified data instances and shortened the list of signal classes that are affected by misclassifications.

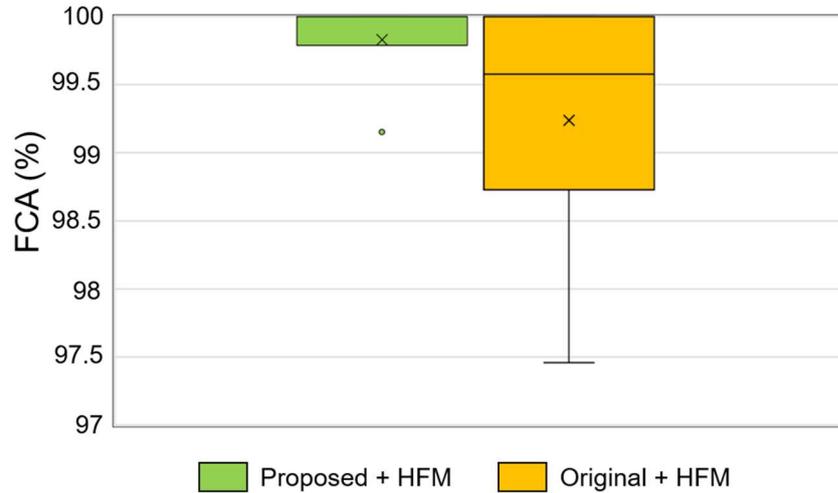


Figure 3.8. The boxplots demonstrating the statistics of the FCA metric over 10 experiments.

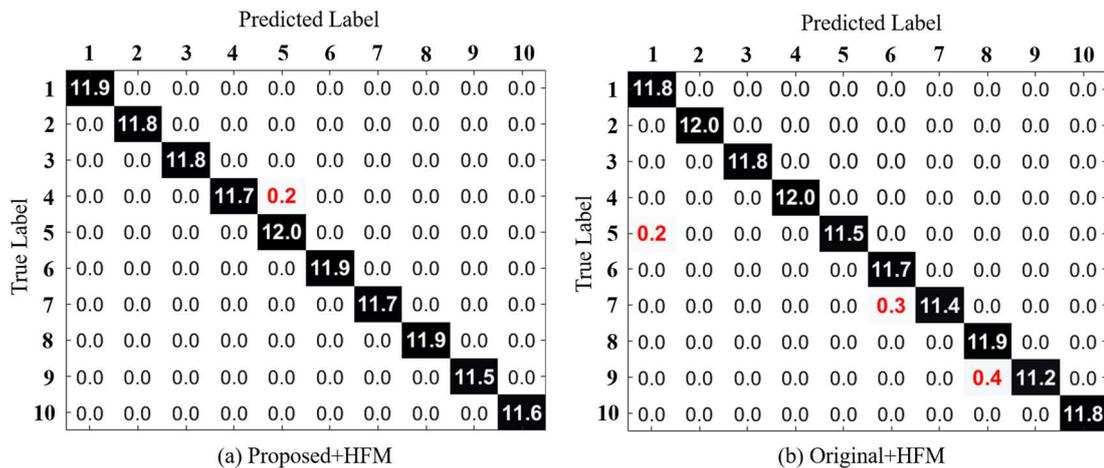


Figure 3.9. The confusion matrix for fault diagnosis results obtained by the SVM machine learning algorithm using the features extracted from signals processed by the proposed methodology. All the results are presented as the average of ten experiments.

Also, the feature space projections created using t-Stochastic Neighbor Embedding (t-SNE)[78, 79] technique for each approach in this comparison are demonstrated in Figure 3.10. From this figure, it can be seen that feature projections of the proposed improved IMF selection approach in conjunction with HFM, introduced in Chapter 2 of this dissertation, demonstrates better feature separability in comparison with the approach presented in Chapter 2, which means that the proposed improved IMF selection leads to improved fault classification capabilities.

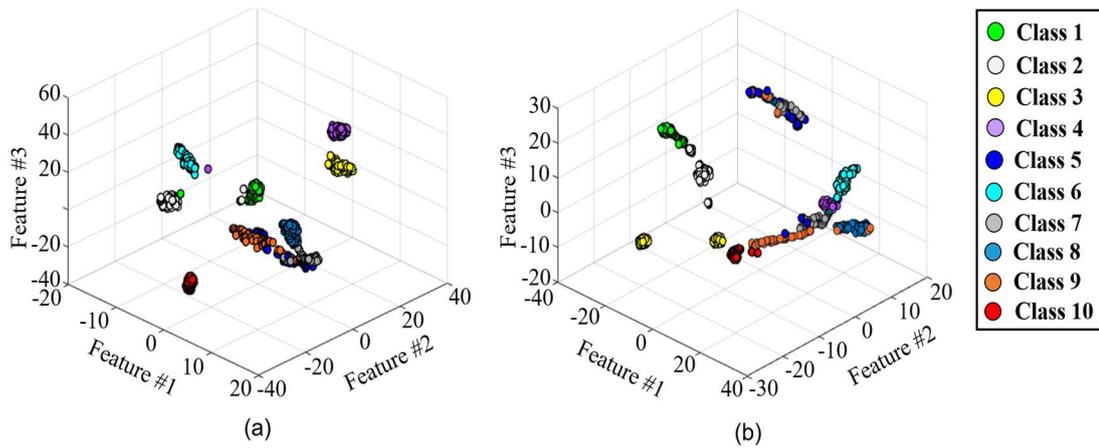


Figure 3.10. Feature space projections for (a) Proposed + HFM and (b) Original + HFM, respectively.

Overall, the results presented in this chapter demonstrate that the proposed improved optimal IMF selection technique for the domain of blade rub-impact fault diagnosis is capable of reducing the presence of high-frequency noise in the partially reconstructed vibration signal while preserving the valuable information content, which consists of frequency harmonics that are considered evident features of rub-impact faults [49, 50, 73]. The presence of these frequency harmonics and the reduction of high-frequency noise is favorable for fault feature extraction and diagnosis because it allows the verification of whether the numerical feature parameters that can be extracted from these reconstructed signals accurately reflect rubbing faults or if they are related to other properties of rotating machinery and the environment in which the machine operates. Furthermore, the proposed rubbing signal processing technique can be highly useful for diagnosing rub-impact faults of different intensity levels when combined with feature extraction and fault classification approaches.

3.5 Conclusions

This chapter introduced modifications for an improvement of the previously proposed DFP/KLD-based optimal IMF selection procedure in EEMD for rubbing fault diagnosis. The proposed improvement used both objective function normalization and the adaptive thresholding computation technique for selecting the subset of optimal IMF components for the diagnosis of blade rub-impact faults. First, the objective function values computed for each of the intrinsic modes were normalized using MinMax scaling for better grading of the modes and to stabilize the statistical properties of a set of extracted IMFs by overcoming the problem of the growth of the average values of metrics with a fault progression. Additionally, this normalization helps to avoid some of the unwanted exceptional cases that might appear while computing objective values. Then, the adaptive threshold was computed and applied to select optimal intrinsic modes. The

experimental results demonstrated that the improved optimal IMF selection approach allows for the better exclusion of the noise-dominant components and the components that do not contain enough essential information about the fault being investigated. This exclusion of the vibration signal modes that are common in normal rotor system operation and are not related to rubbing further enhances the selection of fault components. Moreover, this improvement is capable of reducing the influence of high-frequency noise on the signal reconstructed using the selected optimal IMFs. Thus, this updated technique delivers a rubbing signal with improved clarity that can be efficiently used for fault signal analysis and classification.

Part II
**Fault Diagnosis of Rotating Machines Based on Data-
Driven Artificial Intelligence Techniques**

Chapter 4

Blade Rub-Impact Fault Identification Using Autoencoder-based Nonlinear Function Approximation and a Deep Neural Network

4.1 Introduction

A blade rub-impact fault is a severe type of mechanical fault frequently occurring in rotating machinery, especially in various turbines. The interactions between the blades of the rotor and the stationary parts of rotating machines can be recognized as a separate mechanical fault that can be caused by rotor blade extension due to the high operating temperatures or as a coupling fault where the rub-impact is a consequence (or evidence) of a different mechanical fault. Under the faults leading to blade rub, usually, shaft imbalance, misalignments, excessive self-excited vibrations, or bearing failures are understood [40]. If not detected and identified at the early stages, a blade rub fault may cause the failure of the system and severe economic loss.

Vibration signal analysis [97] is most frequently applied for diagnosing blade rub-impact faults in comparison with other methods, such as acoustic [98], pressure [18], and temperature analysis [19]. The main reason for its application is that performing the vibration signal acquisition in the field is relatively easy compared to other techniques. However, it is known that for proper vibration analysis, the signal processing methods play an important role. A system with a rotor-to-stator rub impact fault is recognized as a complex nonlinear and nonstationary one [99, 100]. This is because, simultaneously, several physical processes are involved in the process of rubbing, such as vibration, thermal effects, stiffness, and friction [20]. These facts cause limitations in the application of the conventional time- and frequency-domain analysis approaches based on statistical feature extraction and Fourier transforms, and force researchers and engineers to utilize complex time-frequency analysis (TFA) methods for extracting valuable information about the mechanical fault and performing fault diagnosis.

In recent years, many studies focused on the extraction of discriminative features from rotor systems using TFA approaches for diagnosing blade rub-impact faults. The most frequently used TFAs are empirical mode decomposition (EMD) [58] and its derivative methods, such as ensemble EMD (EEMD) [92], and wavelet transformation (WT) with its variations [44, 86]. All these methods appeared to be capable of effectively extracting valuable fault features from nonlinear and non-stationary rotor systems in general, and systems with blade rub-impact faults, specifically. However, these methods suffer some drawbacks that lead to difficulties in their application in real industrial fields. First, the EMD method suffers from the problem of mode-

mixing [101], which means that multiple oscillating components are presented in a single intrinsic mode or similar oscillating components are getting split in several modes with disrupted amplitudes. This problem causes difficulties in the interpretation of the decomposition results and affects fault diagnosis accuracy. Second, although EEMD successfully resolves the problem of mode-mixing, it drastically increases the computational complexity of this iterative algorithm. It runs an ensemble of EMD decompositions on a signal to achieve ‘clear’ intrinsic modes. This leads to problems in using this approach in industrial applications, especially when near real-time performance is needed. Finally, talking about the wavelet transform family of methods, they suffer from energy leakage and interference terms that make it difficult to interpret the results of decomposition. Furthermore, the selection of the mother wavelet function significantly affects the results of the wavelet transformation [87]. Unfortunately, to find an appropriate mother wavelet function that correlates well with the signal properties, a series of experiments is needed that leads to subjectivity in this selection process and makes it difficult to make this type of analysis adaptive [47].

For decision making, machine learning algorithms have been extensively utilized in the fields of fault diagnosis and condition monitoring. As examples of the classical widely applied machine learning methods in this domain, k-nearest neighbors [29], support vector machines [102], decision trees [103], and shallow artificial neural network (ANN) [104] architectures can be considered. These algorithms use different concepts for learning how to perform the task of classification, but they have one thing in common: the machine learning algorithms in the field of fault diagnosis are mainly trained on the manually chosen or hand-crafted features to diagnose the mechanical faults. In this case, the fault diagnosis performance of the classical machine learning algorithms relies strongly on human knowledge and expertise, which cannot guarantee that these features are optimal and best characterize the specific fault type being investigated.

Due to the complexity of implementation and the problems of TFA signal analysis approaches for extracting discriminative fault features as well as the problems of the classical machine learning algorithms that are dependent on the feature quality, the other family of algorithms is widely used in the industry for assessing the health condition of rotor systems. These algorithms belong to the family of control theory-based algorithms. Thus, linear-based observers such as proportional-integral observer (PIO) and proportional multi integral observer (PMIO) have been successfully used in several applications [105, 106]. Despite the various advantages of the linear-based observer such as simplicity of implementation in industrial applications and their flexibility, robustness and reliability are the main challenges of this method. To address the issues of linear-based observers, two different scenarios have been defined by researchers: designing the nonlinear-based observer and artificial intelligence-based observation techniques. The nonlinear-

based observation techniques such as sliding mode fault observer, feedback linearization fault observer, and backstepping fault observer are used in several applications. Apart from the various advantages of nonlinear-based observers such as stability, reliability, and robustness, these techniques suffer from a lack of complexity [84, 107]. The second scenario is based on improving the linear-based observer performance using artificial intelligence-based techniques. This scenario also can be implemented in two ways. The first way is to improve the fault diagnosis performance of the conventional linear-based observation techniques with the addition of artificial intelligence techniques into the different stages of the fault diagnosis pipeline [103, 108]. In another way, the nonlinear rotor systems can be diagnosed in a data-driven manner by replacing linear-based observation techniques with artificial intelligence-based approaches, such as using methods from the field of deep learning that are easier to implement and set up compared to the modern nonlinear observation techniques from control theory. This is the scenario that has been chosen in this work.

Recently, the deep learning-based approaches for fault feature extraction and fault classification are drawing attention due to the increased computational power. The deep learning field provides us with a variety of techniques that are capable of learning the discriminative features autonomously or generating them from the given data representations and is actively applied for condition monitoring and predictive maintenance. For instance, talking about supervised learning, convolutional neural networks [109] autonomously extract high-level features from the images [83, 110] and one-dimensional signals [111, 112]. Another technique from the supervised learning family is called deep neural network (DNN) [113] that resembles the conventional ANN with a difference in the depth of the network and number of neurons in its layers, offers better classification capabilities when dealing with high-dimensional input data (can be a one-dimensional signal instead of feature parameters) and uncertainties in it compared to the conventional machine learning algorithms, including shallow structured ANNs. Furthermore, to fully understand the capabilities of deep learning, it is important to discuss the unsupervised methods that are introduced by researchers. One of the great examples of unsupervised deep learning techniques is generative adversarial networks (GANs) [114, 115]. GANs learn the statistical parameters of the input data distribution and are capable of generating new data of similar distribution. In the field of fault diagnosis, GANs are frequently utilized for solving the problem of fault data augmentation before training DNNs for decision making [116]. This application is reasonable and draws a lot of attention in the industry since it is usually not possible or even expensive to collect huge datasets of faulty signals from real applications. However, some challenges might cause difficulties when applying GANs to mimic nonstationary signals of the nonlinear system. First, as was mentioned above, GANs learn the distribution of the input signals during the training so they can generate sequences of similar distribution from the random noise. It is known that in nonstationary signals the statistical parameters of the time sequences tend to

change with the time even when these signals are collected during the same operating condition. Hence, it can be a difficult problem to learn the exact properties of the nonlinear system using GANs. Second, some external disturbances from the environment affect the recorded signals and this unwanted noise will be also included in the distribution learned by GAN. Finally, it is understandable that the learned distribution also depends on the length of time sequence. To address or avoid these problems, the indirect approximation (i.e., the approximation of the system state by learning the specific features of the system instead of signal distribution) of a nonlinear system can be useful. Autoencoders (AEs) [117] is another type of deep learning techniques that can be used to resolve this issue. During the training, AEs learn the specific nonlinear discriminative features (latent coding) of the input data that allow us to use them for feature extraction to determine the state of the system, data reconstruction, data generation, compression, and fault detection in the system [118, 119]. Furthermore, there are extensions of the conventional AEs called denoising autoencoders (DAEs) that enhance the capabilities of AEs to learn the discriminative features of the system by the signals even when those are highly contaminated by noise [120].

Considering the information gathered from the literature review, in this chapter, a novel method based on a deep undercomplete denoising autoencoder (DUDAE) and a DNN is proposed to address the issues of approximating the nonlinear function of the rotor system with coupling blade rub-impact faults and to perform fault identification in a data-driven manner. In this chapter, the replacement of the conventional two-block-based control theory methodology for estimating signal behavior (function approximation of the system and signal estimation) with one block which is represented by the deep learning technique, DUDAE, is attempted. First, the DUDAE is trained using the vibration signals corresponding to the healthy state of the rotor system. During this step, the DUDAE learns the latent coding in its bottleneck layer that represents the nonlinear function of the rotor system under normal operating conditions. Next, the vibration signal corresponding to the unknown state of the rotor system is pushed at the input layer of the DUDAE, where it estimates the signal of the current state using the latent coding learned on signals of normal operating conditions. Then, the residual signal (i.e., error signal) is generated as the difference between the real vibration signal of an unknown system state and the one estimated by the DUDAE. Residue generation is used for enhancing the dissimilarities of the signals corresponding to different classes using the anomaly detection properties of the autoencoder, and hence, generates sequences (residual signals) that are treated as discriminative features capable of improving fault diagnosis performance. At the final step, these residual signals are used as inputs to the DNN to accomplish the task of fault identification of rotating machinery.

The specific contributions of this chapter can be summarized as below:

1. The novel data-driven method for diagnosing coupling rotor imbalance and blade rub-impact faults in nonlinear rotor systems is presented.
2. The deep learning-based system identification approach for approximating the nonlinear function of the system and state estimation has been introduced as a part of the proposed fault diagnosis solution.

The remainder of this chapter is organized as follows. Section 4.2 introduces the proposed methodology for diagnosing the coupling blade rub-impact faults of different severity levels. Section 4.3 provides experimental validation of the introduced framework and discussion. Finally, Section 4.4 contains the concluding remarks.

4.2. Proposed Methodology

A block diagram of the proposed approach for identifying the coupling blade rub-impact faults of various intensities is depicted in Figure 4.1 and consists of three important steps. First, the collected vibration acceleration signals corresponding to the normal operating state when no faults are observed in the system are used to train the DUDAE to create a nonlinear function approximation of a system under normal operating conditions. Then, the autoencoder's property of anomaly detection is used to represent the deviations in the state of the system by generating the residual signal. This residual signal represents the difference (error) between the current vibration signal approximated by the DUDAE using the learned nonlinear function of the normal state of the system and the actual current vibration signal. At the final step, this residual signal is considered a discriminative representation containing fault feature information and describing the current state of the system that is employed as an input to the DNN to accomplish the problem coupling blade rub-impact fault identification.

The dataset containing the coupling shaft imbalance and blade rub-impact fault that is used in this chapter for validating the proposed methodology has been collected by UIAI Lab and described in detail in Chapter 2 of this dissertation.

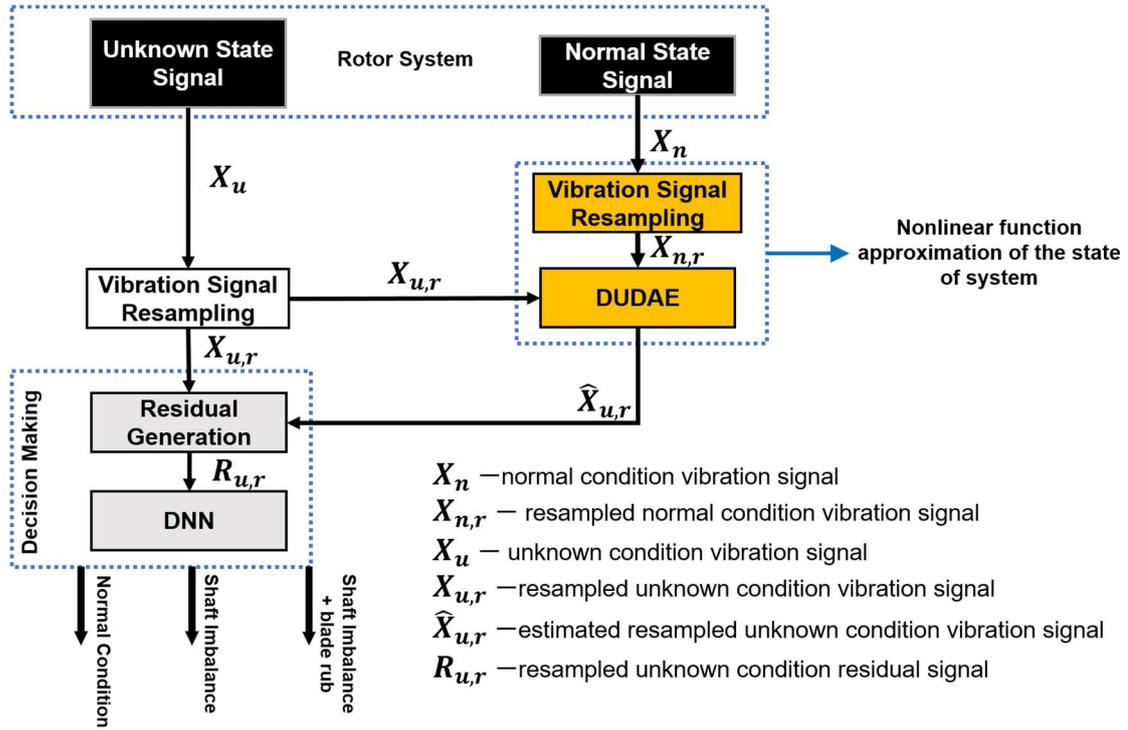


Figure 4.1. The proposed framework for the assessment of the health state of the rotor system.

4.2.1. Signal Resampling

In general, deep learning-based approaches require datasets with a huge number of samples for efficient representation learning. However, it is not always possible and even expensive to collect huge datasets with the samples corresponding to faulty conditions of the system. Furthermore, when the artificial intelligence algorithms are applied to one-dimensional signals, the size of these input signals affects the architecture of the network (i.e., depth of the network, number of nodes, shape) as well as the time needed for learning these representations. To address these issues before creating the autoencoder-based nonlinear observer, resampling of the collected vibration signals corresponding to different states of the system into a series of windows such that each window has a length equal to the number of data points collected during one revolution of the shaft is performed in this chapter.

To resample the vibration signals, the number of revolutions performed in one second (RPS) should be calculated using the following formula:

$$RPS = RPM/60, \tag{4.1}$$

where RPM is the rotational speed used during data recording.

Next, the time needed for one revolution (4.2) and the number of data points (4.3) collected during one revolution of the shaft can be obtained as shown below:

$$TFOR = \frac{1}{RPS}, \tag{4.2}$$

$$w_length = f_{sampling} \times TFOR. \tag{4.3}$$

Here, $TFOR$ stands for the time for one revolution expressed in seconds, w_length corresponds to the length of each window of the resampled signal expressed in a number of data points, and $f_{sampling}$ is the sampling frequency used during the data acquisition.

The computed parameters for resampling the signal into windows are as follows: $RPS \approx 41.6$, $TFOR = 0.024$, and $w_length \approx 1572$, respectively. An example of signal resampling using the achieved resampling parameters is depicted in Figure 4.2.

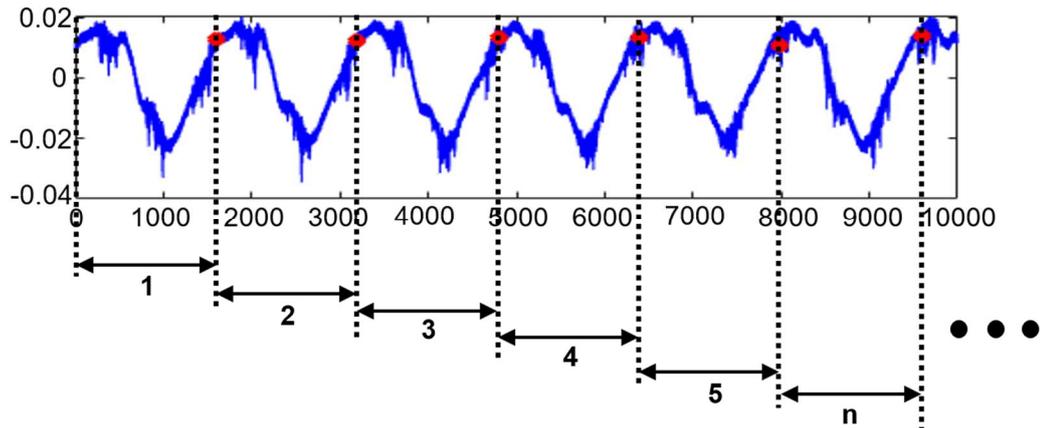


Figure 4.2. The vibration signal resampling process.

After the signal resampling process applied to the dataset described in Chapter 2, the new dataset consisted of 24,190 time-domain resampled vibration signals in total (1,935 resampled signals for each system condition observed during data collection).

4.2.2. Deep Undercomplete Denoising Autoencoder (DUDAE)-based Nonlinear Function Approximation Of The Rotor System

Autoencoders are a type of deep neural networks that are widely used for problems where manifold learning is required. The most common tasks that are solved by autoencoders are feature learning [121], feature extraction [122], and feature selection [123]. However, since autoencoders are deep neural networks with a symmetric structure, they can successfully utilize the properties of neural networks to learn and discover complex nonlinear relations of the input data (i.e.,

nonlinear function approximation) and successfully utilize them for the input data reconstruction, which is the purpose of the autoencoder in this chapter.

The simple undercomplete autoencoder consists mainly of three layers that are trained in an unsupervised manner. The first layer of the autoencoder is called the input layer. It receives the input data and pushes it to the further layers. The hidden layer after the input layer with lower dimensionality is called a bottleneck layer. It is used to extract the latent coding, i.e., the high-level representative features of the input data. The dimensionality of the latent codes is equal to the number of nodes in the bottleneck layer. The last layer, called the output layer, is used to decode the obtained latent codes and reconstruct the original input data. In summary, the autoencoder performs two tasks: 1) it encodes the input data into the latent coding, and 2) it decodes the latent coding to reconstruct the original data. The operation of the autoencoder can be summarized as follows:

$$\begin{aligned} e: x &\rightarrow F \\ d: F &\rightarrow x' \\ e, d &= \operatorname{argmin}(x - x')^2 \end{aligned} \quad (4.4)$$

As mentioned above, the simple undercomplete autoencoder has only one hidden layer, the bottleneck layer. During the encoding stage, the autoencoder receives the input data x of the dimensions R^m and nonlinearly maps the input data to the latent coding F with the dimensions R^n . The encoding process can be presented as below:

$$F = f(Wx + b), \quad (4.5)$$

where F is the latent coding, W represents the weight matrix, b stands for the bias, and f is a nonlinear activation function. The decoding process of the autoencoder is described by:

$$\hat{x} = f'(W'F + b'). \quad (4.6)$$

Here, \hat{x} is the reconstructed output that resembles the input data, and W' , b' , and f' are the weight matrix, bias vector, and activation function of the decoder, respectively.

To perform the training of the autoencoder, the mean squared error (MSE) loss function should be calculated between the original input data and the reconstructed data using the following equation:

$$L(\boldsymbol{\theta}) = \frac{1}{N} \sum_{n=1}^N (x - \hat{x})^2, \quad (4.7)$$

where L stands for the MSE loss function, θ is a set of model parameters, and N is the dimensionality of the input data, i.e., the number of nodes in the input layer of the autoencoder.

In this chapter, the DUDAE is utilized to approximate the nonlinear function of the normal state of the rotor system. The detailed architecture of this autoencoder is presented in Table 4.1. Unlike the simple three-layer undercomplete autoencoder, the proposed DUDAE is a deep autoencoder (emphasized by the first ‘D’ in the abbreviation) that has more than one hidden layer, as can be seen in the table. However, the basic idea described in equations (4.4), (4.5), (4.6), and (4.7) pertains to the DUDAE, with the only difference that during the encoding and decoding phases more nonlinear data transformations are done concerning the increased number of hidden layers. From the same table, it can be seen that the size of the encoding layers is smaller than that of the input layer, which means that the structure of the proposed autoencoder is ‘undercomplete’ (highlighted by the ‘U’ in the abbreviation). This is needed to force the autoencoder to learn a more compact representation (i.e., nonlinear function) from the input data. To increase the tolerance to the noise of the autoencoder used for approximating the nonlinear function of the normal operating state of the system, the dropout [124], with a rate equal to 0.1, is added to the input layer in which the input signals are received. This makes the proposed autoencoder belong to a type of denoising autoencoders (this property is expressed as the second ‘D’ in the abbreviation).

Table 4.1. The architecture of the proposed DUDAE.

Layer #	Type of Layer (Purpose)	Node #	Activation	Dropout
#1	Input (<i>Encoder</i>)	1572	-	0.1
#2	Hidden (<i>Encoder</i>)	790	SELU	-
#3	Hidden (<i>Encoder</i>)	395	SELU	-
#4	Hidden (<i>Encoder</i>)	128	SELU	-
#5	Hidden (<i>Encoder</i>)	64	SELU	-
#6	Hidden (<i>Encoder/Decoder</i>)	32	SELU	-
#7	Hidden (<i>Decoder</i>)	64	SELU	-
#8	Hidden (<i>Decoder</i>)	128	SELU	-
#9	Hidden (<i>Decoder</i>)	395	SELU	-
#10	Hidden (<i>Decoder</i>)	790	SELU	-
#11	Output (<i>Decoder</i>)	1572	SELU	-

As the activation function for the hidden and output layers of the DUDAE, the scaled exponential linear units (SELU) function is chosen in this study. There are a few main reasons for employing this activation function: 1) the input vibration signals collected by the sensors contain both the positive and negative values, hence a possibly non-saturating activation function that supports these types of inputs is needed, 2) the specific formulation of the SELU activation

prevents the vanishing gradient problem that may be faced in deep architectures, as well as avoids the situations when the neuron can die during training, and 3) the SELU activation function speeds up the training process and convergence of the deep neural network due to its normalization properties [125]. The formulation of the SELU activation function is shown in equation (4.8):

$$f_{selu}(x) = \lambda \begin{cases} \alpha e^x - \alpha & x \leq 0 \\ x & x > 0 \end{cases} \quad (4.8)$$

where $\lambda \approx 1.05$ and $\alpha \approx 1.6731$ are the coefficients predetermined by the inventors of SELU activation [125].

Glorot uniform weight initialization [126] was chosen as the initialization strategy of the weights in the hidden layers of the proposed deep undercomplete denoising autoencoder.

As the optimization algorithm for training the deep denoising undercomplete autoencoder to estimate the non-linear function of the normal system state using backpropagation, an adaptive moment estimation (Adam) optimization algorithm [127] is used in this chapter. The Adam optimizer simultaneously tracks the exponentially decaying average values of the past gradients and past squared gradients of the loss function, which is categorical cross-entropy in this study. The flow of the Adam optimization method can be summarized using the following equations:

$$\begin{aligned} s_T &\leftarrow \beta_1 \cdot s_{T-1} + (1 - \beta_1) \cdot \nabla_{\theta} Loss_T(\theta_{T-1}), \\ z_T &\leftarrow \beta_2 \cdot z_{T-1} + (1 - \beta_2) \cdot \nabla_{\theta}^2 Loss_T(\theta_{T-1}), \\ \hat{s}_T &\leftarrow s_T / (1 - \beta_1^T), \\ \hat{z}_T &\leftarrow z_T / (1 - \beta_2^T), \\ \theta_T &\leftarrow \theta_{T-1} - n \cdot \hat{s}_T / (\sqrt{\hat{z}_T} + \varepsilon). \end{aligned} \quad (4.9)$$

Here, s_T and z_T are the estimates of the first moment and second raw moment, respectively; n stands for the learning rate (equal to 0.001); β_1 and β_2 are the momentum decay hyperparameters; T corresponds to the order number of training iterations; θ is a set of model parameters such as weight and bias values of the kernels, and ε corresponds to a smoothing term used to prevent dividing by zero.

4.2.3. Residual Signal Generation

The main purpose of the autoencoder (DUDAE) in this work is to learn the nonlinear function of the system under normal operating conditions. Once the training is completed, this

trained model is used to give its estimate of the current system state by attempting to reconstruct the signal previously unseen during the training (i.e., a signal corresponding to the unknown state of the system). Next, the residual signals are generated as a difference signal between the real unknown vibration signal and the estimate of this signal delivered by the DUDAE. These residual signals are used at the next step as the input for the DNN to perform fault identification, and can be computed as below:

$$r_{\hat{x}}(n) = x(n) - \hat{x}(n), \quad (4.10)$$

where $r_{\hat{x}}(n)$ is the residual signal, $x(n)$ stands for the original vibration signal in the time domain, and $\hat{x}(n)$ is the signal reconstructed by the autoencoder using the latent coding learned while training on signals corresponding to the normal operating state of the system (i.e., when no imbalance and no blade rub fault are observed).

The purpose of computing the residual signals is as follows. Since the DUDAE is trained using only the data collected under normal system operating conditions, it learns how to reconstruct this data by using the learned nonlinear function, i.e., latent coding. However, it cannot accurately reconstruct the data that have not been used during the training process. That is, if the DUDAE is applied to reconstruct the signals not observed during training and that significantly deviate from the signals corresponding to the normal system state, it will inevitably lead to a reconstruction error. Furthermore, when a shaft imbalance or a coupling imbalance and blade rub fault appear in the system, the values of the statistical parameters of the vibration signals increase with the increase of their amplitude. This means that errors between the real signals corresponding to abnormal conditions of the system and the ones estimated by the DUDAE will increase too. This allows for detection of the current state of the system and residual signals computed by Eq. (4.10) can be used as discriminative features to perform fault identification of coupling blade rub faults of various intensity levels.

4.2.4. Fault Identification Using Residual Signals and The DNN

Even though the DNN is a variation of the conventional ANN, which was first introduced a long time ago, due to the higher dimensionality and the number of hidden layers, it became one of the most powerful and widely applied decision-making algorithms for a huge variety of problems. Furthermore, DNNs became the main core of recent trends in the field of artificial intelligence algorithms, such as deep representation learning.

The general DNN architecture resembles the architecture of an ANN and consists of an input, output, and a sequence of hidden layers. The generalized formula of the m^{th} hidden layer operation can be summarized as follows:

$$x_m = f(W_m x_{m-1} + b_m), \quad (4.11)$$

where x_m is the output of the m^{th} hidden layer after applying the nonlinear activation function f ; x_{m-1} is the output of the previous hidden layer after application of the activation function; and W_m and b_m are the weight matrix and bias vector of the m^{th} hidden layer, respectively.

In this chapter, the DNN is used to perform the task of blade rub-impact fault identification using the residual signals computed using Eq. (4.10). The exact architecture of the DNN used for fault identification is presented in Table 4.2. As can be seen from the table, the architecture of the proposed DNN is like the encoder part of the DUDAE described in Subsection 4.2.2. However, there are two differences that are discussed below.

The first is the way the dropout regularization has been applied. Unlike the autoencoder, where the dropout was applied only to the input layer to increase its robustness to the noise in the data, in the DNN it is used for fault identification and a dropout rate of 0.1 is applied to hidden layers #2, #3, #4, and #5 to avoid overfitting of the data. If the DNN overfits the training data, it might fail to generalize the validation and testing data (the data unseen during the training process) which will lead to a decrease in the fault classification performance. It cannot be seen from the table, but along with dropout regularization, an early stopping procedure is applied during the training of the DNN to reduce the chance of overfitting. The idea of early stopping is to interrupt the training process once the validation error stops decreasing or starts increasing with some tolerance level during a defined number of epochs.

The second difference is the activation function of the output layer. To solve a multiclass classification problem, SoftMax activation is employed in the output layer of the DNN. The SoftMax activation function is given as follows:

$$\hat{P}_k = \exp(\mathbf{s}_k(x)) / \sum_{i=1}^K \exp(\mathbf{s}_i(x)), \quad (4.12)$$

where K is the total number of classes and $\mathbf{s}(x)$ is a vector with the scores of every available class for the specific data sample x . The input data sample is assigned to the class with the highest estimated probability \hat{P}_k (i.e., the class that has the highest computed score for this instance).

To train the DNN to perform blade rub fault identification using the residual signals, the categorical cross-entropy loss function is used with the outputs of the SoftMax activation of the output layer to perform decision making about the state of the system. The categorical cross-entropy loss can be formulated as below:

$$Loss(\boldsymbol{\theta}) = -\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K y_k^i \log(\hat{P}_k), \quad (4.13)$$

where $\boldsymbol{\theta}$ is the set of model parameters and y_k^i and \hat{P}_k are the target and estimated probabilities that the i^{th} data sample belongs to the class k , respectively. The same optimization algorithm used for training the autoencoder, Adam (Subsection 4.2.4, Eq. (4.9)), is used for training the DNN by computing the gradients of the categorical cross-entropy loss function concerning model parameter $\boldsymbol{\theta}$.

The remaining parameters of the machine learning model, such as the weight initialization algorithm, the learning rate of the optimization algorithm, and other parameters of the network remain the same, as described in Subsection 4.2.2.

Table 4.2. The architecture of the DNN used for fault identification.

Layer #	Type of Layer	Node #	Activation	Dropout
#1	Input	1572	SELU	-
#2	Hidden	790	SELU	0.1
#3	Hidden	395	SELU	0.1
#4	Hidden	128	SELU	0.1
#5	Hidden	64	SELU	0.1
#6	Hidden	32	SELU	-
#7	Output	10	SoftMax	-

4.3. Experimental Results and Discussion

4.3.1. Training, Validation, and Testing Data Configuration

In this subsection, to investigate the fault identification capabilities of the proposed approach, the two experimental datasets were constructed from the newly obtained dataset after signal resampling presented in Subsection 4.2.1.

The first dataset consisted of all the resampled time-domain vibration signals corresponding to the normal state of the system (2,419 resampled signals), i.e., when neither imbalance nor coupling imbalance and blade rub faults were observed (this dataset is further referred to as dataset #1). This dataset is needed to train the DUDAE to reconstruct the input data using the learned latent coding and to derive the residual signals that are further used for fault identification by the DNN. For training the DUDAE, dataset #1 was randomly split into training and validation subsets at a rate of 8:2. Thus, 1,935 resampled signals corresponding to normal system conditions were used as a training subset for the DUDAE; whereas the remaining 484 signals comprised the validation subset used to measure validation error.

Once the autoencoder was trained, it was used to generate the residual signals using the whole 24,190 original resampled vibration signals. The data were used further to accomplish the task of fault diagnosis. For this, the dataset of residual signals (further referred to as dataset #2) was first randomly split into training and testing subsets at a ratio of 8:2. Then, the obtained training subset was randomly split again at the ratio 8:2 to get a validation subset. Thus, the obtained training subset from dataset #2 consisted of 15,481 residual signals, the validation subset contained 3,968 samples, and the remaining 4,838 residual signals previously unseen by the DNN were used as a testing subset for evaluating fault diagnosis capabilities of the proposed framework.

To ensure the reliability of the proposed methodology and exclude the effect of randomness, the experiments for the proposed and referenced methods will be performed 10 times with different training, validation, and testing subsets randomly sampled at each trial.

4.3.2. Choosing the Noise Application Strategy in DUDAE

Before training and validating the proposed methodology, in this subsection, the noise application strategies for improving the robustness of DUDAE towards the noise presence in the data are evaluated to choose the most suitable one for the finalized DUDAE-DNN model. In denoising autoencoders, the noise corruption to the data is usually applied in two different ways: 1) application of dropout technique on input layer of an autoencoder and 2) mixing the additive Gaussian noise into the input signals. To select the best scenario for current work, the DUDAE was subsequently trained and validated using dataset #1 during 400 epochs with different rates of dropout applied to the input layer of DUDAE and with Gaussian Noise generated using various variance values. The specific rates of dropout investigated in this subsection were within 0.1 and 0.55 with a step equal to 0.05. Regarding the additive Gaussian Noise, the variance values were varying between 0.1 and 1 with a step equal to 0.1.

The obtained minimum training and validation values of MSE loss function for different realizations of dropout and additive Gaussian noise in the proposed autoencoder architecture are presented in Table 4.3.

From the results presented in the table it can be seen that based on validation loss values, the best candidate strategies for applying noise corruption to the input data in the DUDAE model are dropout with a rate of 0.1 and additive Gaussian Noise with parameters $N(0,0.1)$. Thus, the final fault identification performance of the proposed DUDAE-DNN model will be evaluated using both strategies for the sake of consistency.

Table 4.3. The minimum training and validation loss values obtained for different realizations of dropout and Gaussian noise.

	Value	Minimum Training Loss	Minimum Validation Loss
Dropout Rate	0.1	5.1449e-06	4.7204e-06
	0.15	5.4662e-06	4.8044e-06
	0.2	5.4428e-06	5.4887e-06
	0.25	5.9627e-06	5.4831e-06
	0.3	6.0753e-06	5.5354e-06
	0.35	6.0566e-06	6.2125e-06
	0.4	6.3664e-06	5.9437e-06
	0.45	6.3820e-06	6.0325e-06
	0.5	6.5031e-06	6.1219e-06
	0.55	6.7455e-06	6.6316e-06
Additive Gaussian Noise	0.1	4.4639e-06	2.6747e-05
	0.2	4.3942e-06	7.8743e-05
	0.3	4.5247e-06	1.3001e-04
	0.4	4.3222e-06	1.5928e-04
	0.5	4.4410e-06	1.6850e-04
	0.6	4.4685e-06	1.9876e-04
	0.7	4.5786e-06	1.9855e-04
	0.8	4.4306e-06	2.0628e-04
	0.9	4.6818e-06	2.1095e-04
	1.0	5.1233e-06	2.2212e-04

4.3.3. Training the Finalized DUDAE-DNN Model

Before validating the capabilities of the proposed framework to identify blade rub-impact faults of various intensity levels, the modules of the proposed framework, the DUDAE and the DNN, should be trained. Furthermore, they should be trained in a pipeline (i.e., sequential order). Thus, first, the training and validation subsets of dataset #1 containing the time-domain resampled vibration signals corresponding to the normal condition are used to train the DUDAE. Next, the training and validation subsets of dataset #2 (consisting of residual signals obtained after training the DUDAE) are utilized to train the DNN to perform fault diagnosis. For training both parts of the model, data batches with 64 data samples each were utilized. The number of training epochs for the DUDAE was assigned to be equal to 600 while the training of the DNN model was completely left under the control of an early stopping algorithm that stops the learning process once the validation accuracy stops improving and restores the model parameters that demonstrated the highest fault classification accuracy on the validation subset. Once the training epoch number of DUDAE and training scenario for decision-maker (DNN) are fixed, the training-validation procedure is repeated 10 times to observe the behavior of training-validation loss curves and generalize the conclusions on the convergence of the proposed methodology. For simplicity, in

this subsection the dropout of rate 0.1 was applied to the input layer of DUDAE as a noise corruption technique; however, in the next subsection, where the fault diagnosis results are presented, the fault classification performance of the DUDAE-DNN model with both strategies used for noise corruption is presented. The training and validation loss curves obtained during 10 experiments are presented in Figure 4.3.

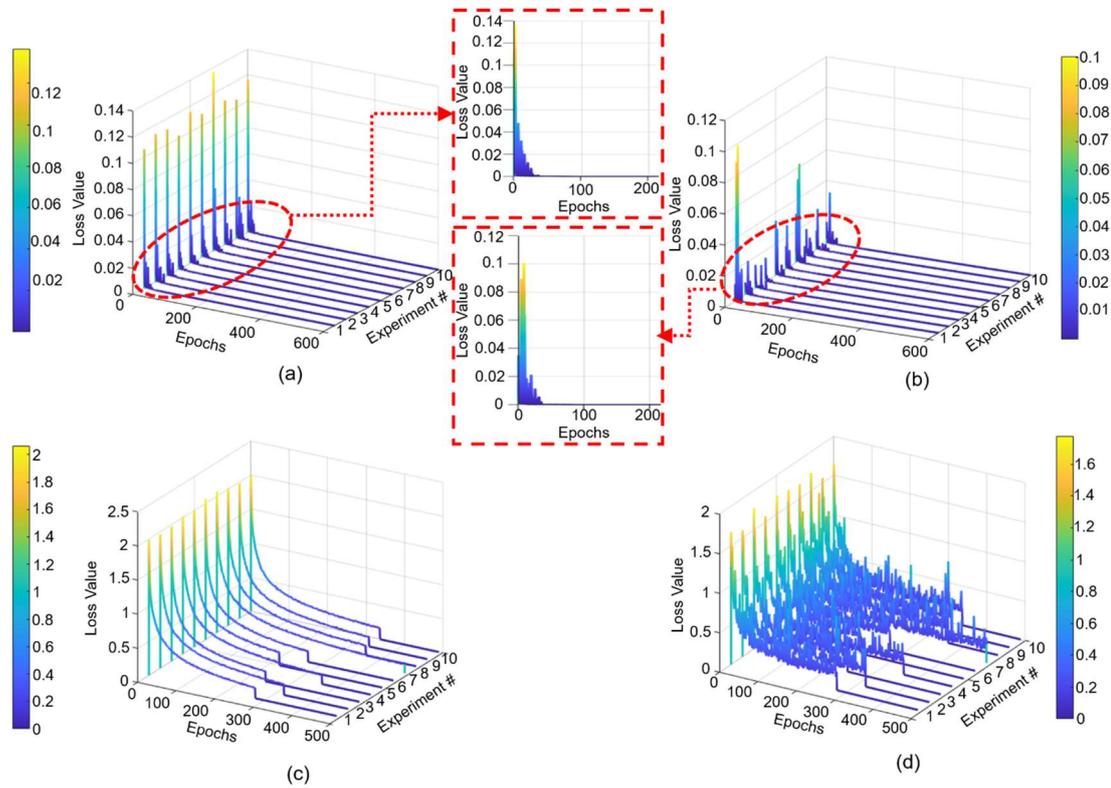


Figure 4.3. The training and validation loss curves obtained during 10 experimental trials by the proposed technique: (a) training loss curves of DUDAE, (b) validation loss curves of DUDAE, (c) training loss curves of DNN, and (d) validation loss curves of DNN, respectively.

The training and validation curves corresponding to DUDAE are demonstrated in Figure 4.3 (a) and (b). From these figures, it can be seen that the values of loss functions during ten experiments first demonstrated sharp descent during the first 40 epochs of training and then continued decreasing towards zero steadily. Despite in all experimental trials DUDAE has been trained during 600 epochs, from Figure 4.3 (c) and (d) one can observe that the training process of DNN has been stopped by an early stopping algorithm at different moments before 400 epochs in all trials except experiment #7, where the training of DNN has lasted for 488 epochs (the longest result). From Figure 4.3 (d) and its color bar it can be concluded that in all experimental trials the validation loss curves of DNN demonstrated similar descending patterns and at the moment when the training procedure was stopped, they were oscillating around the value of 0.2.

Overall, it can be concluded that the proposed methodology demonstrates repeatable results in terms of convergence under various training and validation subset permutations. However, it can be also seen that there is an open direction for improvement of the part related to the decision making in the proposed framework because, despite a good convergence of DUDAE under various data permutations, the loss functions of DNN saturated at a certain level without moving closer to zero.

4.3.4. Residual Signal Analysis

In this subsection, the analysis of residual signals obtained after the DUDAE was trained on signals corresponding to the normal system state is provided. From the previous subsection, it was concluded that when the number of training epochs of the DUDAE is equal to 600, the obtained residual signals that are used as input to the DNN for decision making on the state of the system lead to the highest classification accuracy on the validation dataset. The main point of this is that the well-trained DUDAE delivers residual signals of a small magnitude oscillating around zero (i.e., small reconstruction error) for the signals that correspond to the normal state of the system or for the signals that resemble those signals. On the other hand, when the imbalance and blade rub-impact fault appear in the rotor system, the vibration signals start deviating from the ones corresponding to a normal operating state. Hence, with the increase of rub-impact fault intensity, the reconstruction error increases as well, which leads to residual signals of higher magnitudes and higher deviations from zero. The examples of residual signals computed after the trained DUDAE for different states of the system are depicted in Figure 4.4.

As can be seen from this figure, the magnitudes of residual signals and their shapes change with the progression of the fault. Furthermore, it can be seen that MSE values computed between the original and reconstructed signals also increase when the signals in the input of the trained DUDAE deviate significantly from the signals corresponding to the normal system condition when neither shaft imbalance nor blade rub faults was observed.

Figure 4.5 illustrates the energy of residual signals generated by the proposed methodology for five signal classes, namely normal system condition (class #1), shaft imbalance fault (class #4), shaft imbalance + slight rubbing fault (class #6), shaft imbalance + intensive rubbing fault (class #9), and shaft imbalance + severe rubbing fault (class #10), respectively. The signal groups presented in Figure 4.5 are the same as those demonstrated in Figure 4.4 for the sake of consistency. In the proposed methodology, DUDAE extracts the function of the dynamic behavior of the normal signal (the rotor system is under the normal operating condition when no faults are observed) during its training. However, in abnormal conditions of the system, the behavior of the signal is utterly different from its behavior in the normal state of the system. Regarding Figure 4.5, it can

be seen that the accuracy of the dynamic behavior estimation for the signals belonging to different classes is satisfactory, especially for class #1. The reason for this observation is that the residual signal itself is a type of error signal that is computed between the actual vibration signal and one estimated by DUDAE. That is, since the DUDAE has been trained on signals belonging to normal conditions, it is capable of accurately estimating the unknown signals when their dynamic behavior is close to the ones it has been learned on. Furthermore, it can be seen from Figures 4.4 and 4.5 that when DUDAE estimates the unknown signal dynamic behavior of which drastically differs from ones collected under normal operating conditions, the estimation error (residual signal) between the actual and estimated signal is increasing. In Figure 4.4 this difference can be observed in the deviation of residual signals from zero-mean along with the increasing value of MSE metric, while in Figure 4.5 this difference is highlighted by growing values of energy features extracted

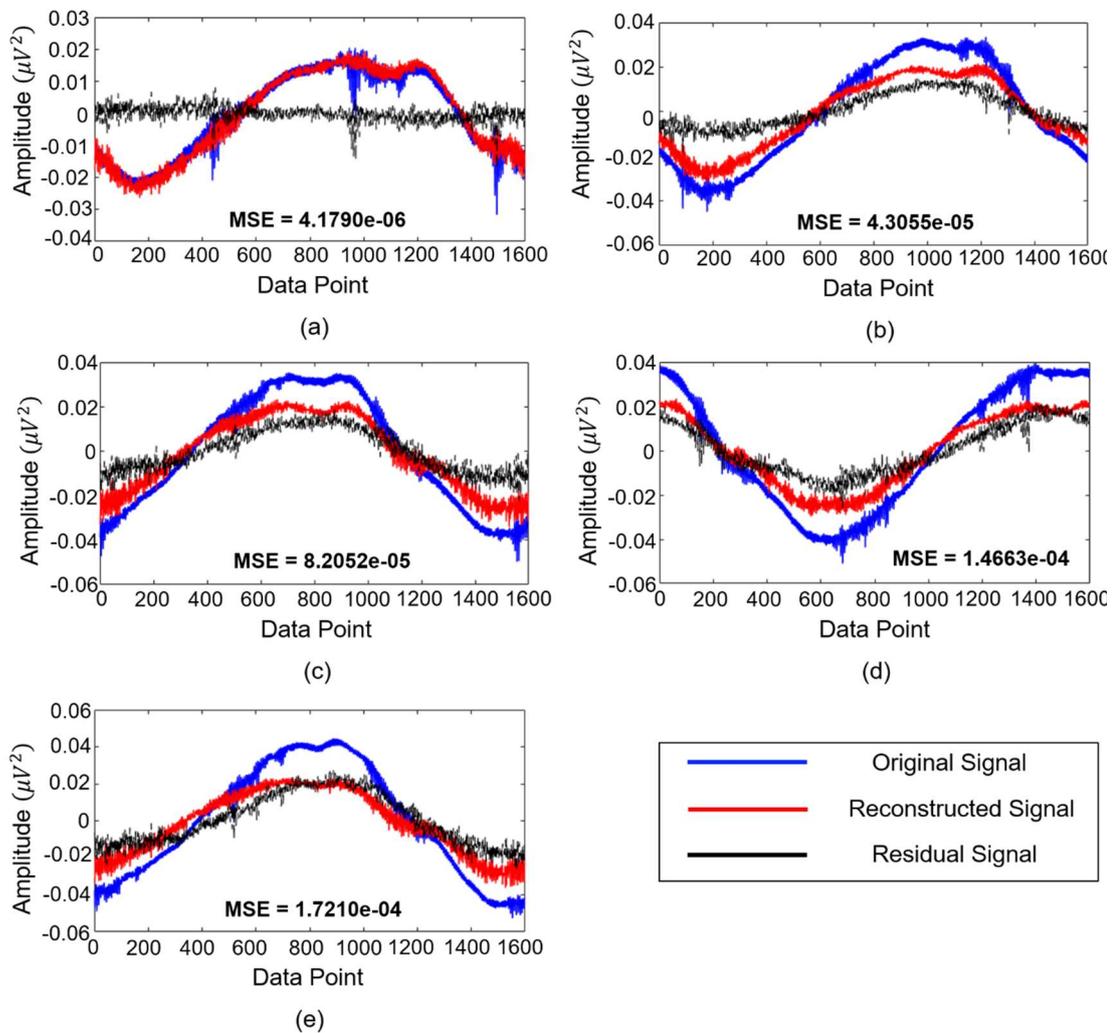


Figure 4.4. The original, reconstructed, and residual signal examples corresponding to signal classes of (a) normal operating condition, (b) 1.5 g shaft imbalance condition, (c) 1.7 g shaft imbalance + slight blade-rub fault condition, (d) 2.4 g shaft imbalance + intensive blade rub-fault, and (e) 2.8 g shaft imbalance + severe blade rub fault, respectively.

from those residual signals. Based on energies of the residual signals presented in Figure 4.5 it can be concluded that the obtained residual signals are sensitive to the degradation of the system which means that these residual signals can be used as discriminative features itself for fault classification or for feature extraction in conjunction with feature-based machine learning classifiers for diagnosing faults. Thus, the more discriminative the residual signals are, the easier it is for the classifier to perform fault identification. However, some overlap can be observed when the intensity of rub fault increases, such as in classes #9 and #10. Therefore, to improve the potential fault classification accuracy, the DNN with the residual signals as input features is recommended in this work instead of conventional amplitude-based statistical feature extraction and fault classification schemes the performance of which can be affected by the overlap of the extracted feature parameters.

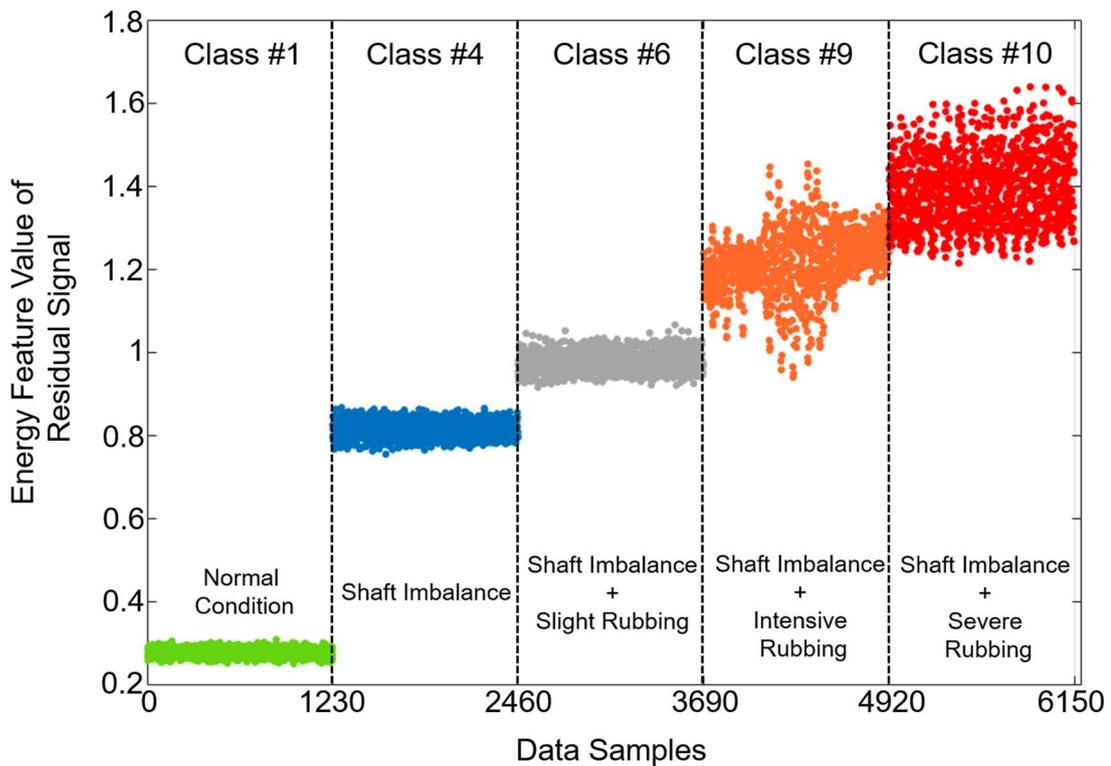


Figure 4.5. The energy feature parameter used for characterizing the residual signals obtained by proposed methodology for the signal classes under normal operating condition, 1.5 g shaft imbalance condition, 1.7 g shaft imbalance + slight blade-rub fault condition, 2.4 g shaft imbalance + intensive blade rub-fault, and 2.8 g shaft imbalance + severe blade rub fault, respectively.

4.3.5. Fault Identification Performance

To evaluate the fault identification capabilities of the proposed framework (further referred to as doProposed), this technique is compared with three counterpart methods. First, for a better understanding of DUDAE properties, the model which comprises of the proposed DUDAE and

DNN with the only difference in how the denoising properties of the autoencoder were elaborated is created. In this new model, instead of dropout-based data corruption at the input of DUDAE, the additive Gaussian noise with the commonly used parameters of $N(0,0.1)$ was mixed into the signals corresponding to the normal operating condition of the system while training the DUDAE (further referred to as gnProposed). Furthermore, since the proposed model is a pipeline process containing two main steps, nonlinear function approximation of the system state and decision making, for fair comparison it is reasonable to fix the decision-making approach (i.e., DNN) and vary the methods at the first step to observe whether the proposed pipeline influences the fault identification abilities or not. Thus, the second method used for the comparison is directly applying the DNN to resampled signals in the time domain (further referred to as RAW+DNN). This approach allows us to investigate the improvement in classification performance of the proposed method where nonlinear function approximation by the DUDAE is utilized in comparison to when no function approximation is used. In the third approach used for the comparison, a widely used state-of-the-art linear observation method from the field of control theory, autoregressive with external input ARX-Laguerre proportional-integral observer (PIO) (ARXLPIO) [103], is utilized for estimating the nonlinear blade rub-impact fault signals. The residual signals computed as the difference signals between the original raw signals and ones estimated by ARXLPIO are used as the inputs to the DNN to accomplish the task of fault diagnosis. This method will be further referred to as ARXLPIO+DNN. The architecture of the DNN employed in the comparison approaches matches the one used in the proposed DUDAE + DNN model. Note, in this comparison, the modern control theory algorithms, such as nonlinear observation techniques, are not utilized. The main reason for this, as was discussed in the introduction part of this chapter, is the complexity of the design process of these approaches in a real industrial environment as well as the need to re-design the nonlinear observation technique whenever the system changes.

The fault classification performance for the methods mentioned above is evaluated using the micro-averaged forms of widely used metrics [77], such as micro-averaged recall (Rec_μ), micro-averaged precision ($Prec_\mu$), micro-averaged F1-score ($F1_\mu$), and total fault classification accuracy (FCA). It is decided to use the micro-averaged versions of these metrics to address the possible deviations in the numbers of data samples presented in each class in the testing subsets due to the random sampling procedure applied at each trial of the experiment. These metrics are defined in detail by Eq. (2.12-2.15) in Chapter 2 of this dissertation. The experimental results expressed in these metrics averaged over 10 experiments are tabulated in Table 4.4.

Table 4.4. The experimental results averaged over 10 experiments.

Methods	Metrics (Std) (%)			
	Rec_{μ}	$Prec_{\mu}$	$F1_{\mu}$	FCA
gnProposed	95.64 (0.3)	95.64 (0.3)	95.64 (0.3)	95.64 (0.3)
doProposed	95.11 (0.3)	95.11 (0.3)	95.11 (0.3)	95.11 (0.3)
ARXLPIO+DNN	92.19 (0.7)	92.19 (0.7)	92.19 (0.7)	92.19 (0.7)
RAW+DNN	83.68 (7)	83.68 (7)	83.68 (7)	83.68 (7)

As can be seen from the table, both types of the proposed framework (i.e., gnProposed and doProposed) outperformed the counterpart methods with the highest average FCAs of 95.64% and 95.11%, respectively. The counterpart approaches, ARXLPIO+DNN and RAW+DNN, achieved FCAs of 92.19% and 83.68%, respectively. Furthermore, the standard deviation of FCA values demonstrated during 10 experimental trials is the lowest for both variants of the proposed method. More details regarding fault classification accuracies can be observed in Figure 4.6 where the boxplots with distributions of accuracy values obtained during 10 experiments are presented. The black cross in the boxes belonging to different methods in Figure 4.6 corresponds to the average classification accuracy values presented in Table 4.4.

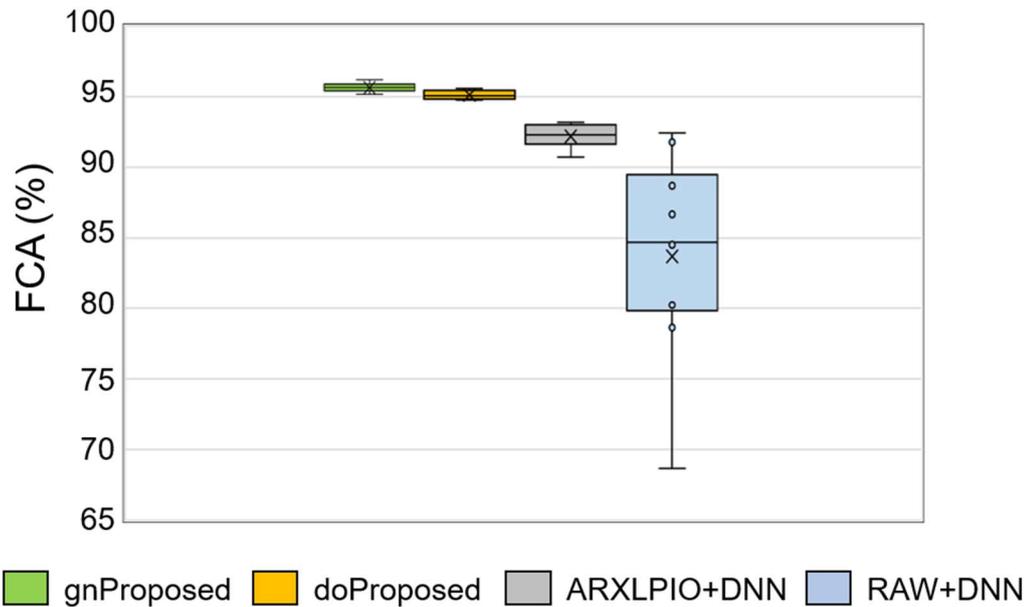


Figure 4.6. The boxplots demonstrating the statistics of the FCA metric over 10 experiments.

As can be seen from the boxplots demonstrated in Figure 4.6, the classification accuracy values did not deviate significantly from the mean and median values during the experiments for the variants of the proposed method which ensures the repeatability of the results. For ARXLPIO+DNN it can be seen that the deviation of accuracy values also not very significant with

outliers not laying far from the box; however, all the accuracy values distributed lower than the results of the proposed technique. Unlike the proposed method where AI-based system identification has been used and ARXLPIO+DNN where the linear observer has been utilized, one can see that the box corresponding to the RAW+DNN method is wider with a long whisker laying towards the outlier of 68.1%. From this figure it can be concluded that the proposed DUDAE used for nonlinear function approximation and ARXLPIO observation technique both can improve the fault diagnosis stability; however, the DUDAE helps to increase the average classification performance when applied to nonlinear rubbing signal in comparison with linear observation technique.

Figure 4.7 presents the confusion matrices averaged over 10 experiments to provide more details on the fault diagnosis performance. From this figure, it can be seen that the variants of the proposed technique demonstrated the lowest numbers of misclassifications in conditions where the nonlinearity of the rotor system increases in comparison with referenced techniques, especially, with the RAW+DNN where the DNN has been applied directly to nonstationary vibration signals.

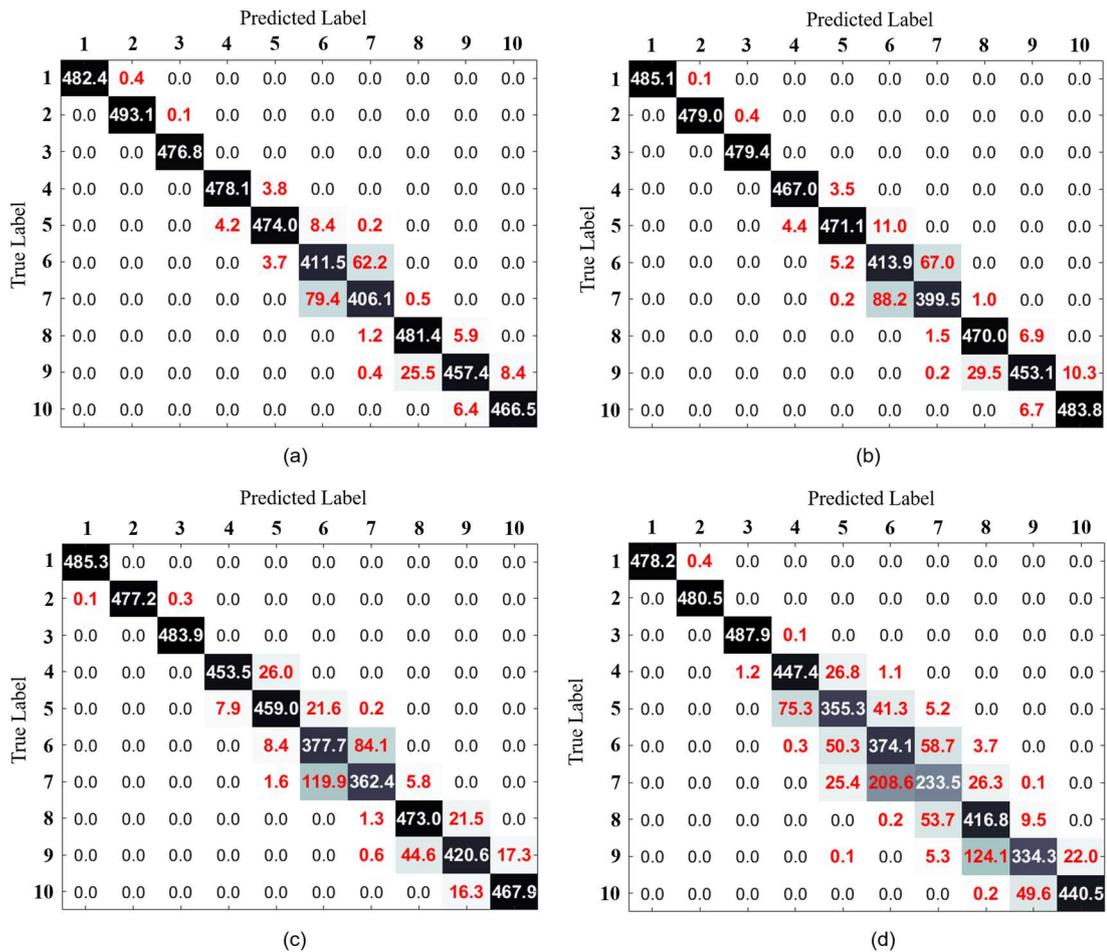


Figure 4.7. The confusion matrices obtained for (a) gnProposed, (b) doProposed, (c) ARXLPIO+DNN, and (d) RAW+DNN methods, and averaged over 10 experiments.

These conditions corresponded to the appearance of coupling imbalance and slight blade rub-impact faults (classes #5-7) and imbalance with intensive blade rubbing faults (classes #8 and #9).

The main reason that the ARXLPIO+DNN method shows degraded performance in comparison with the proposed methodology is in the nature of the ARXLPIO method. When a linear observation technique, such as ARXLPIO, is applied to approximating the nonlinear function of the system state when dealing with nonlinear and nonstationary signals, it inevitably leads to the appearance of estimation errors. This is because the uncertainty term of the nonlinear and nonstationary signal (i.e., blade rub-impact fault signal) modeling cannot be estimated properly by the linear observation technique. Even though the residual signals obtained after the ARXLPIO observation technique appear to be more discriminative as features in comparison with the original raw time-domain signals, the degradation of classification performance in comparison with the nonlinear observation techniques from the field of control theory or the artificial intelligence-based techniques is expected.

The RAW+DNN method demonstrated the lowest FCA in comparison to other techniques presented in Table 4.4. In this approach, the DNN utilized the raw resampled time-domain vibration signals as the inputs to perform the task of fault identification. It mainly demonstrated lower accuracy in comparison with the proposed approach due to the complexity of the blade rub-impact fault signal. Due to its non-stationarity, the statistical properties of time-domain samples may vary over time even when they belong to the same signal class, which leads to the problem that time-domain vibration signal patterns are not discriminative enough and may lead to the failure of the DNN to adjust its weights during training to reach a good level of generalization.

Overall, it can be concluded that the proposed data-driven framework consisting of the DUDAE-DNN model is suitable for diagnosing blade rub-impact faults of various intensity levels with high fault classification accuracy in comparison with the other referenced methods. From the experimental results, it can be seen that the application of the DUDAE for approximating the nonlinear function of the nonlinear rotor system state improves the fault diagnosis capabilities of the DNN in comparison with the state-of-the-art linear observation techniques frequently used in industry and the situations when no signal observation is used. Another important advantage of the proposed methodology is that its structure is pipeline-shaped which supports modifications of current architecture as well as applicability to other systems since the residual signals used as features in this study are generated based on the ideas of system identification. Furthermore, during the experiment, two types of noise corruption (i.e., dropout-based and Gaussian noise-based) in DUDAE were evaluated and the results obtained by both variants did not differ significantly from each other while demonstrating high classification accuracy in comparison with counterpart approaches. However, from the results, it is also can be seen that the proposed methodology for

diagnosing coupling blade rub-impact faults still should be improved to increase its classification performance when dealing with vibration signals corresponding to increasing nonlinearity of the rotor system. Furthermore, to accomplish a comprehensive investigation of the robustness and reliability it is important to test the proposed methodology on the datasets with varying operating conditions, such as varying rotating speed and varying load.

4.4. Conclusions

In this chapter, a novel data-driven method for diagnosing complex coupling faults consisting of shaft imbalance and blade rub-impact faults of different severity levels is introduced. In the proposed fault diagnosis technique, the input time-domain vibration signals are first resampled concerning the fundamental frequency of the rotating machine. Then, the nonlinear function approximation of the system state under normal operating conditions is accomplished by training the deep undercomplete denoising autoencoder on the resampled signals corresponding to the state of the system when neither imbalance nor blade rub-impact faults were observed. Next, the residual signals are computed as an error between the unknown vibration time-domain signals and their estimates by the autoencoder. Finally, these residual signals were used as the inputs to the deep neural network to perform the decision making about the current state of the rotor system. The series of experiments show that the proposed fault diagnosis model demonstrated stable convergence behavior under different training-testing data permutations and outperformed other methods used for the comparison in terms of the micro-averaged performance metrics.

Chapter 5

Intelligent Rubbing Fault Identification Using Multivariate Signals and a Multivariate One-Dimensional Convolutional Neural Network

5.1 Introduction

Rotating machines are widely used in various industrial fields. One of the vital fields for modern society and manufacturing is electricity production, and turbines are the main type of rotating machines that are utilized in power plants. The working conditions of the turbines are usually extreme and associated with high rotating speeds and high-temperature levels. To maximize the efficiency of turbines, increase the torque, and reduce air reluctance, it is important to maintain a small clearance between the stationary and rotary parts of the turbine. When the rotor blades start interacting with the stationary part of the turbine the rubbing phenomenon appears. It usually happens due to the presence of different types of mechanical faults appearing in rotating machines, such as shaft misalignment, self-excited vibrations, rotor blade extension (due to the high operating temperatures), and other faults [40]. If not detected in the early stages, rubbing faults in turbines can cause excessive damage to the equipment of the power plants and significantly increase the expenditures required for maintenance or even cause a blackout.

The conventional way to perform fault diagnosis on rotating machines includes two main steps: fault feature extraction and decision making. Due to the complex nature of rubbing faults, they are recognized as a non-stationary and non-linear type of mechanical fault [41]. Analysis of these mechanical faults requires complex solutions which are usually based on time-frequency analysis (TFA) methods, i.e., the methods which allow for simultaneous analysis in the time and frequency domains. The popular TFA approaches applied for extracting discriminative features of rubbing faults include the wavelet transform and its variations [43–45], empirical mode decomposition [128], ensemble empirical mode decomposition [63], local mean decomposition [129], variational mode decomposition [25], time-frequency distribution decomposition [130], and others [131–134]. These techniques have various advantages. For example, wavelet-based approaches are known for providing a good time-frequency resolution, which allows precise analysis of the decomposed signals. EMD, EEMD, LMD, and similar iterative techniques are data-driven methods that can decompose the time frequencies into the intrinsic oscillating components corresponding to the specific frequency bands of the original signals. Despite the advantages and insights that the aforementioned approaches can provide during the analysis of non-stationary and non-linear signals, these techniques have different drawbacks that can cause difficulties for

utilizing these TFA methods in real industrial applications. Specifically, the techniques from the wavelet family are strongly dependent on prior signal filtering using the particular wavelet mother function. Unfortunately, there is no uniform methodology for selecting these wavelet functions. To find an appropriate one, a series of experiments or thoughtful analysis of the signal before its decomposition are required [47]. Regarding the iterative decomposition approaches, these techniques do not require any prior knowledge about the signal. However, these iterative approaches are very expensive in terms of computational time needed to perform the decomposition of the signals. The analysis of the extracted intrinsic mode functions also requires a solid level of expertise. Furthermore, there are different challenges which can arise when utilizing these approaches for multivariate signals analysis. These challenges include finding matching intrinsic mode components extracted from multivariate signals collected by different sensors, finding an appropriate sub-band for each signal collected using various sources, as well as the growing complexity of computations. These issues make it difficult to directly utilize the aforementioned techniques for analyzing multivariate signals in industrial applications.

Once the fault features are obtained, the decision-making process should be undergone to determine the state of the system. For this purpose, approaches based on machine learning algorithms that utilize the extracted features to separate the signals corresponding to different states of the system are widely applied. These techniques include the solutions where k-nearest neighbors [29], support vector machines [89], support vector regression [135], artificial neural networks [136], and classification and regression trees algorithms [103] are applied for resolving the problems of fault diagnosis and remaining useful life estimation. However, with the increase of computational capabilities of modern devices, another set of methods from the field of artificial intelligence is drawing attention. The set of these approaches is called ‘representation learning techniques’. The main difference between them and the conventional machine learning methods is that these methods are capable of extracting the hidden patterns and features autonomously from the given data. One of the state-of-the-art techniques from the family of representation learning approaches is called convolutional neural networks (CNN) [137]. This network type was originally invented for image processing [109, 138]. It proved its capabilities of successfully extracting discriminative local features from the representation patterns and performing classification. Taking into account this information, researchers successfully applied CNNs for health assessment of rotating machinery [83, 139–141]. Thus, the fast kurtogram representation of the acoustic emission signals with a modified LeNet-5 CNN has been applied to diagnose rolling element bearing faults in [83]. Liu et al. [139] introduced the dislocated-time series two-dimensional CNN to accomplish the task of induction motor fault diagnosis. The authors in [140] suggested slicing the vibration signal into windows and stacking them one after another to create an image pattern for the

application of a two-dimensional CNN to diagnose rolling-element bearing faults. Wu et. al. [141] applied a one-dimensional CNN to detect rub-impact faults in the hydro turbine.

From the literature review, it was observed that most of the TFA approaches designed for analyzing complex mechanical faults such as rub-impact faults are ‘heavy’ in terms of algorithm complexity. Thus, there might be challenges in utilizing them for multivariate signal processing for diagnosing mechanical faults in real scenarios. On the other hand, the majority of the intelligent methods proposed for diagnosing mechanical faults in general and rubbing faults specifically, do not consider the employment of multivariate signals. Instead, those works utilize the data collected by an arbitrary selected single source (i.e., a sensor or a channel of a sensor), even when multiple sources of signals are available. In the opinion of the authors, the use of multiple sources of data in conjunction with intelligent solutions, such as CNN, that can solve both the feature extraction and decision-making tasks can be favorable for industrial applications and represent robust fault diagnosis frameworks. Additionally, to the best of our knowledge, almost no works have been reported on applying CNNs to perform fault identification of the rub-impact faults of various intensity levels.

To address these issues, in this chapter a new intelligent technique for diagnosing rub-impact faults of different intensities using multivariate signals and a multivariate one-dimensional convolutional neural network (ModCNN) is proposed. Specifically, first, the multivariate vibration signal collected using multiple sources (in this study, the channels of the sensors) is resampled with overlap into windows of the specific size dependent on the operating speed of the rotating machine. Next, the envelope power spectra of these windows corresponding to the resampled multivariate signals are computed and used as the inputs to the ModCNN. The proposed ModCNN consists of the set of convolutional-subsampling layer pairs for extracting discriminative local features from the windows of the multivariate vibration signal. The number of these pairs is equal to the number of sources used for multivariate signal acquisition, where each pair extracts the features from the signals acquired from that particular source. Finally, the extracted features are merged in a fully connected layer and the fault diagnosis task is accomplished in the output layer of the ModCNN.

The remainder of this chapter is organized as follows. Section 5.2 provides insights into the proposed fault identification methodology. Section 5.3 presents the experimental validation of the introduced techniques and a discussion on the obtained results. Finally, Section 5.4 contains concluding remarks.

5.2. Proposed Methodology

As shown in Figure 5.1, the proposed intelligent approach for rubbing fault diagnosis consists of three essential steps: multisensory data collection, signal processing, and fault identification. Specifically, first, the set of multivariate vibration signals is collected using the available sensors mounted on the testbed. Next, the collected signals are resampled to expand the number of data instances and their envelope power spectra are computed. Finally, the power spectra of the resampled signals collected from multiple vibration sensors are used for the representation learning by the ModCNN to accomplish the task of rubbing fault identification.

The original dataset used in this chapter for validating the proposed methodology comprises of signals collected from the testbed where the rubbing faults were simulated by adding extra weights to the shaft to create a shaft imbalance that caused local interactions between the blades of the rotor and a rotor cage. This dataset is described in detail in Chapter 2 of this dissertation.

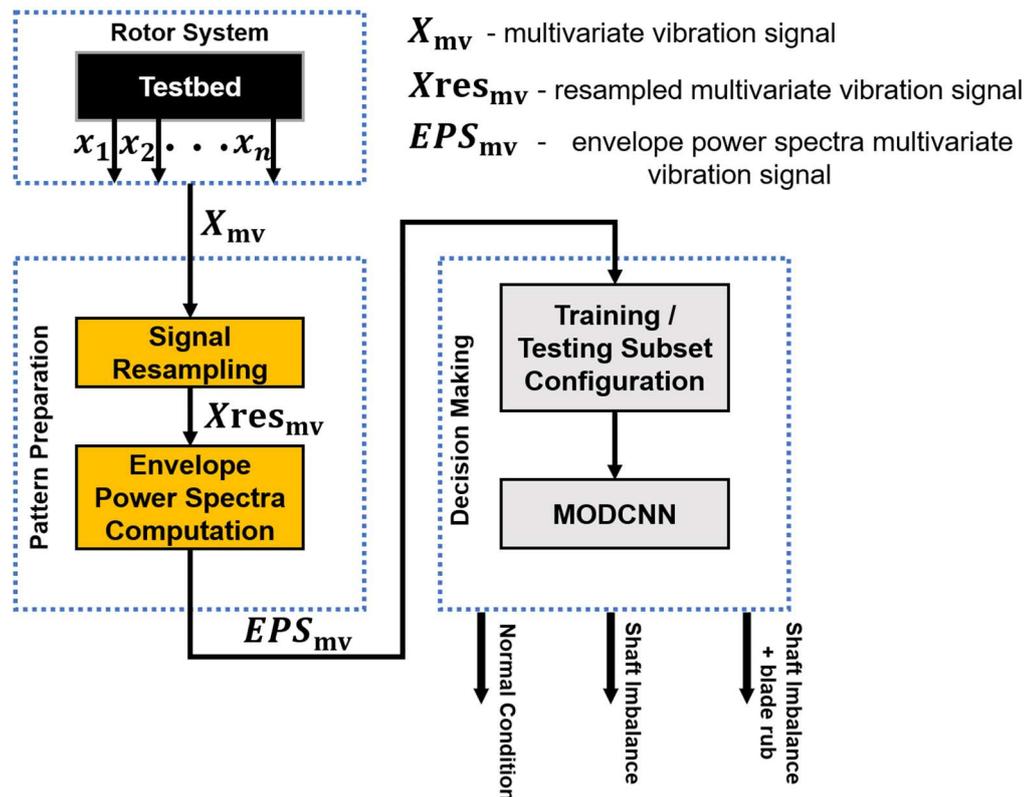


Figure 5.1. Proposed intelligent rub-impact fault diagnosis framework.

5.2.1. Multivariate Signal Processing and Pattern Preparation

5.2.1.1. Signal Resampling

In general, to achieve a high generalization capability, the CNNs require huge training datasets. However, in real industrial applications, it is not always possible to collect huge datasets. This is especially so when talking about collecting data samples corresponding to faulty conditions of the system. At the same time, the sampling frequencies used for data collection are quite high, which makes it difficult to use them directly as the inputs to CNNs, but allows us to expand the number of samples by resampling the originally collected data instances. In this work, we performed the resampling of the originally collected signals corresponding to various severity levels of rubbing faults into a series of windows, where each new sample has the length of the time-domain samples corresponding to one revolution of the shaft.

To resample the initially collected signal, the number of revolutions performed in one second (*RPS*) should be calculated using the following formulation:

$$RPS = RPM/60, \quad (5.1)$$

where *RPM* is the rotational speed in RPMs that is used during data collection. After computing the number of revolutions per second, the time (5.2) and the number of signal points (5.3) needed for one revolution can be obtained as shown below:

$$TFOR = \frac{1}{RPS}, \quad (5.2)$$

$$w_length = f_{sampling} \times TFOR. \quad (5.3)$$

Here, *TFOR* stands for the time of one revolution expressed in seconds, *w_length* corresponds to the length of each new instance of the resampled signal expressed in a number of data points, and *f_{sampling}* is the sampling frequency used during data collection.

The values of the *RPS*, *TFOR*, and *wsize* computed using (5.1), (5.2), and (5.3) for the signals of the dataset used in this study were approximately equal to 41.6, 0.024, and 1572, respectively

To ensure that each of the new resampled instances has a relation with the previous one, we cut the resampled signal into windows of *w_length* with a 15% overlap, as shown in Figure 5.2.

After resampling with the overlap procedure described above the resampled version of the dataset for the experiment had a size of 28,320 instances in total with 2,832 samples per class, respectively.

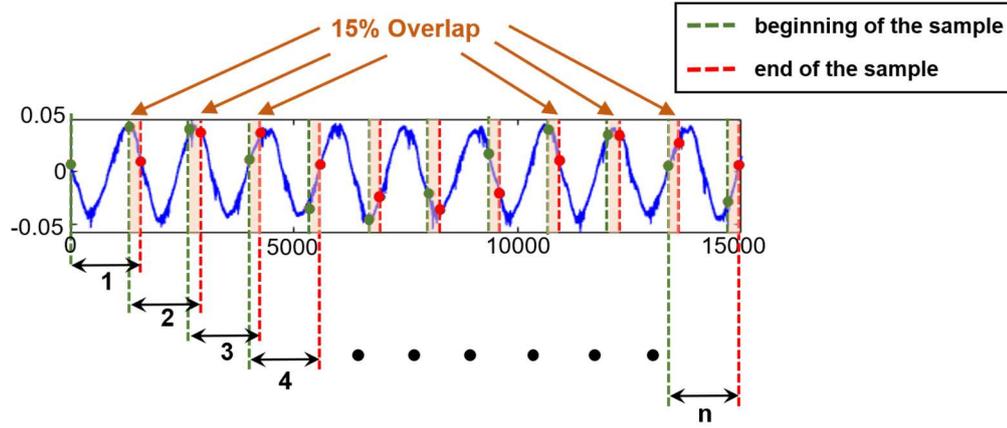


Figure 5.2. The vibration signal resampling procedure with a 15% overlap.

5.2.1.2. Pattern Preparation

In this work, the rub-impact fault signals acquired from the multisensory testbed were represented as envelope power spectra computed from the simultaneously recorded vibration time-domain signals after their resampling. To represent rubbing signals as envelope power spectra, the analytical signals of the resampled signals in the time-domain are first calculated. For instance, if $x(t)$ is the original signal, the analytical signal can be represented as a combination of the original signal and the virtues of the Hilbert transform. The analytical signal can thus be formulated as follows:

$$x^h(t) = x(t) + i\tilde{x}(t), \quad (5.4)$$

where $\tilde{x}(t)$ is the Hilbert transform. Convolution of the original signal with the signal $1/\pi t$ yields the following:

$$\tilde{x}(t) = x(t) \cdot (1/2\pi) = (1/\pi) \int_{-\infty}^{+\infty} h(t)(dt/(t - \tau)). \quad (5.5)$$

The power spectrum of the analytical signal is then obtained by taking the square of the absolute value of the Fourier transform, $|F\{x^h(t)\}|$.

The examples of the envelope power spectra of the resampled rubbing fault signals corresponding to the normal state of the system from both collected datasets are presented in Figure 5.3. The representation of the collected signals in the form of envelope power spectra is useful since this type of representation successfully demodulates the rubbing signals. Furthermore, as can be seen from Figure 5.3, this type of representation serves as a sort of normalization for the input data. Thus, the frequency power spectra are all positive values (i.e., the amplitude of frequency harmonics is equal or higher than zero), which provides useful information about the scales and

form of the input data for the selection of the activation function used in the layers of the ModCNN. Additionally, after removing the mirrored parts of the power spectra, the dimensionality of the resampled signals is significantly reduced. This affects the speed of feature extraction and fault diagnosis processes performed by the ModCNN.

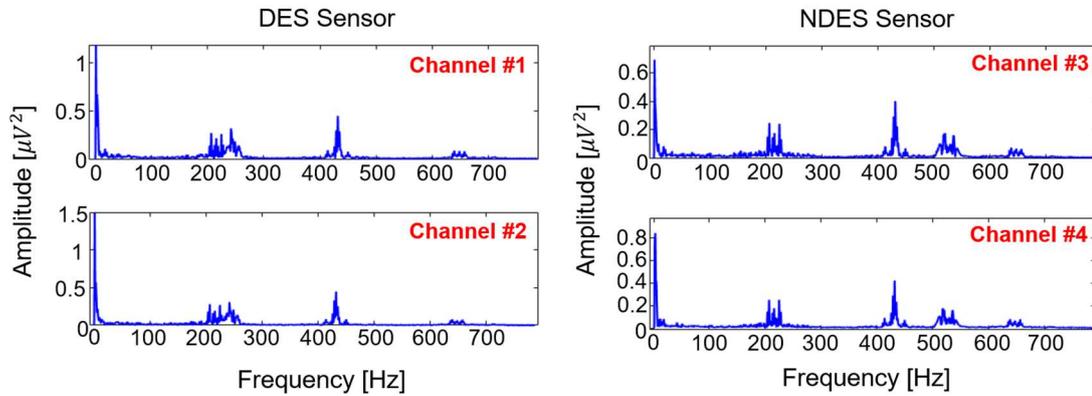


Figure 5.3. The envelope power spectra computed from the resampled multivariate vibration signal collected by multiple channels.

5.2.2. Rubbing Fault Identification using a Multivariate One Dimensional CNN

To accomplish the task of rubbing fault diagnosis, the ModCNN is proposed in this study. The main concept of the ModCNN is that several pairs of convolutional-subsampling layers (CSL) are used simultaneously to extract discriminative local features from the multivariate signal, i.e., when the process is characterized by the time-sequences incoming from several sources. After the CSL pairs, the local features extracted from multiple inputs by these pairs are merged in a common fully connected layer and are pushed towards the output layer to perform decision making about the state of the system. It is known that due to the low computational requirements and shallow architectures of the small one-dimensional CNNs, they are capable of fast operation and relatively easy to implement in comparison with the conventional CNN architectures. Hence, a set of these tiny CNNs can be applied for processing multivariate signals in industrial applications where rapid analysis and operation are essential, which is one of the advantages of the proposed methodology. The overview of the proposed ModCNN is depicted in Figure 5.4. As can be seen from this figure, the proposed architecture of the ModCNN consists of several CSL pairs, which is equal to the number of sensors/channels available for multivariate signal collection in this study (denoted as ‘N’ in the figure for a general case), one common fully connected layer, and one output layer where the decision about the state of the system is made. The stride step in both the convolutional and subsampling layers is equal to two, and no padding is used in the proposed network architecture.

In a one-dimensional convolution layer applied to the pattern represented by the envelope power spectrum in this study, there are several convolutional kernels. The corresponding output,

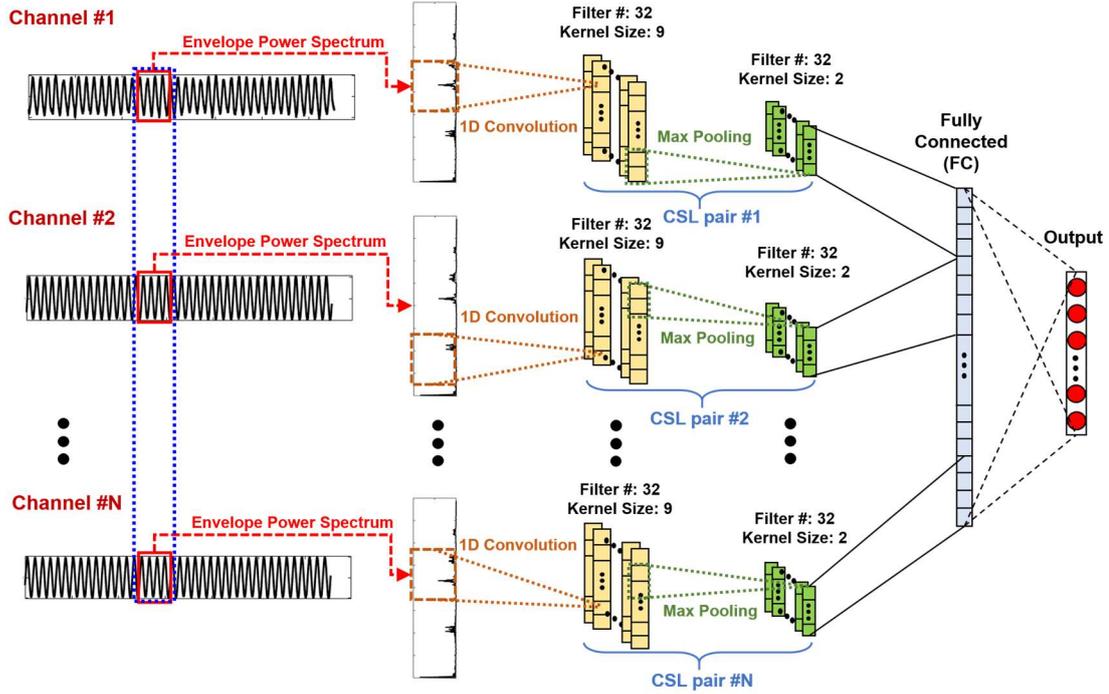


Figure 5.4. The architecture of the proposed ModCNN.

obtained after the convolutional operation is applied to the one-dimensional pattern in different CSL pairs can be formalized as below:

$$out_{k,p}^{conv} = f(\sum X_p^{inp} * W_{k,p} + B_{k,p}). \quad (5.6)$$

Here, k stands for the kernel number, p denotes the number of CSL pairs in the ModCNN, $out_{k,p}^{conv}$ is the output (feature map) of the k -th kernel corresponding p -th CSL pair, $W_{k,p}$ is the k -th convolutional kernel corresponding to the p -th CSL pair, $B_{k,p}$ is the bias vector of the k -th convolutional kernel in the p -th CSL pair, X_p^{inp} is the input of the p -th CSL pair (i.e., the envelope power spectrum computed from the signal collected by p -th source), $f(\cdot)$ is the nonlinear activation function, which is chosen as a Rectified Linear Units (ReLU) function in this study, and $*$ stands for elementwise multiplication.

In each CSL pair, the one-dimensional convolutional layer is followed by a pooling layer. Pooling layers are widely used for dimensionality reduction of the feature maps delivered by the convolutional layers by selecting the most important features from local sub-regions. In this study, the max-pooling operation is applied in the subsampling (pooling) layer which selects the maximum value in each sub-region. The feature maps, which are the outputs of the subsampling layer, are given as follows:

$$out_{k,p}^{ss} = f(mB_{k,p}down(out_{k,p}^{conv}) + B_{k,p}), \quad (5.7)$$

where $out_{k,p}^{ss}$ is defined as the k -th output feature map of the subsampling layer in the p -th CSL pair, $mB_{k,p}$ is multiplicative bias, $B_{k,p}$ defines additive bias, $down(.)$ stands for the pooling function, and $f(.)$ corresponds to the nonlinear activation function. Glorot Uniform Distribution [126] is used to initialize the weights of the kernels of convolutional and subsampling layers in this chapter. The values of bias were initialized at zero before the training process.

After the features are derived by all CSL pairs, i.e., from signals collected by all the vibration sensors or channels, these features are merged and weighted by a fully connected layer which provides the output through the ReLU activation function. The fully connected layer is followed by an output layer with the SoftMax activation function. SoftMax activation is a popular choice of activation functions for output layers of neural networks where multiclass classification problems need to be solved. SoftMax allows for estimating the probability that the particular data instance belongs to a specific class by computing the scores of each class for this data instance. The formulation of this activation function is provided below:

$$\hat{P}_m = exp(s_m(x)) / \sum_{i=1}^M exp(s_i(x)), \quad (5.8)$$

where M is the total number of classes and $s(x)$ is a vector with the scores of every available class for the specific data sample x . The input data sample is assigned to the class with the highest estimated probability \hat{P}_m (i.e., the class that has the highest computed score for this instance).

The categorical cross-entropy loss function is used with the outputs of the SoftMax activation to perform decision making about the state of the system. The categorical cross-entropy loss can be formulated as below:

$$Loss(\theta) = -(1/n) \sum_{i=1}^n \sum_{m=1}^M y_m^i \log(\hat{P}_m), \quad (5.9)$$

where θ is the set of model parameters and y_m^i and \hat{P}_m are the target and estimated probabilities that the i -th data sample belongs to class m , respectively.

This chapter applies the same optimization algorithm for adjusting the hyperparameters of the ModCNN as was described in Chapter 4, Subsection 4.2.2: adaptive moment estimation (Adam) optimizer. The Adam optimizer simultaneously tracks the exponentially decaying average values of the past gradients and past squared gradients of the loss function, which is categorical

cross-entropy in this study. The flow of the Adam optimization method in detail is presented in Chapter 4, Subsection 4.2.2, Eq.(4.9).

To prevent overfitting of the ModCNN, the dropout and early stopping approaches are utilized in this study. Specifically, dropout regularization with a rate of 0.1 is applied to the common fully connected layer where the local features extracted from different channels of the vibration signal are stacked together. This means that during the training process, at each learning iteration, 10% of the features stacked in the fully connected layer are ‘dropped out’ randomly and do not participate in the computation of the loss function gradients during this step. Regarding early stopping regularization, the training process stops once the validation loss stops decreasing or starts increasing after a pre-defined standby time. The method above significantly improves the capabilities of deep learning approaches to generalize previously unseen data instances and reduce the chance of the model overfitting the training data.

The very shallow architecture of the proposed ModCNN in combination with the resampling signal into small overlapping windows and its conversion into discriminative patterns (envelope power spectra) allows for reasonably fast extraction of local features from the multivariate signal, as well as high-quality fault classification of the rub-impact faults of various intensities. Furthermore, due to these advantages, this approach can be employed in real industrial applications where strict requirements for signal processing and decision-making procedures exist.

5.3. Experimental Results and Discussion

5.3.1. Training, Validation, and Testing Subsets Configuration

To evaluate the fault diagnosis capabilities and applicability of the proposed approach for diagnosing rubbing faults of different intensity levels, both of the resampled datasets were randomly split into training, validation, and testing subsets. The complete process of creating these subsets is depicted in Figure 5.5. As can be seen from the figure, the resampled dataset was first separated into training and testing subsets at a ratio of 8:2. Then, the obtained training subset was again split into training and validation subsets using the same proportion as in the previous step. Specifically, 18,124 samples were used as training samples, and 4,532 data instances were used as a validation subset for training the proposed ModCNN, whereas 5,664 data samples were used as

a testing subset to evaluate the capabilities of the learned DNN to correctly classify rubbing faults of various intensities.

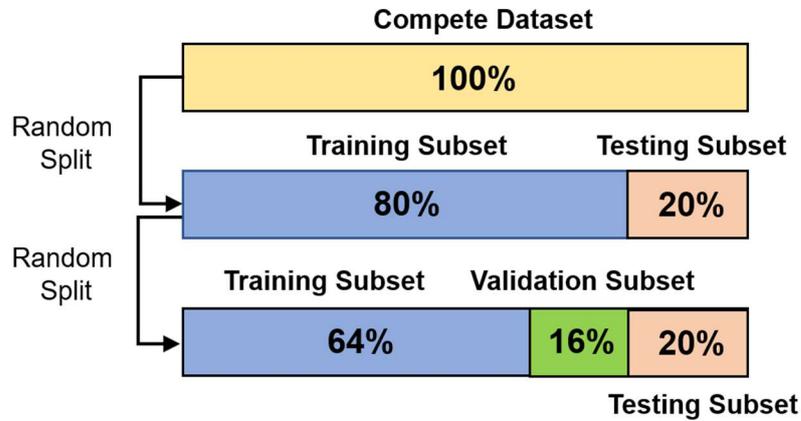


Figure 5.5. The scheme of splitting resampled dataset into training, validation, and testing subsets.

5.3.2. Training and Validation Process of the ModCNN

Prior to performing the fault diagnosis and evaluating the classification capabilities of the proposed methodology, the ModCNN should be trained. In this subsection, the network is trained with the envelope power spectra of resampled signals from the training and validation subsets. Since one of the aims of the proposed framework is to perform fast fault diagnosis using a multivariate signal, it has been chosen to assign the maximum number of training epochs to be equal to 200. Such a small number of training iterations allows for the rapid deployment of the fault diagnosis technique in an industrial environment. The convergence curves of loss functions and classification accuracies obtained during 10 experimental trials are demonstrated in Figure 5.6.

As can be seen from the figure, the introduced ModCNN with the proposed signal processing roadmap which includes signal resampling into small overlapping windows and pattern preparation (envelope power spectrum computation) demonstrates similar convergence patterns on both training and validation subsets. The training and validation loss curves sharply decrease during the first 30 training epochs until the loss value of around 0.1 and continue decreasing more steadily during the remaining epochs to a loss value level close to zero. Symmetrically, the training and validation accuracy curves converge to the accuracy level near 100% within the same number of training epochs. The results demonstrated in Figure 5.6 show that the proposed DL model is capable of demonstrating stable convergence results under different training-testing subset permutations. However, from this figure, it can be seen that it is possible to reduce the training

time approximately in two times (around 100 training epochs) in cases when even more rapid deployment is required.

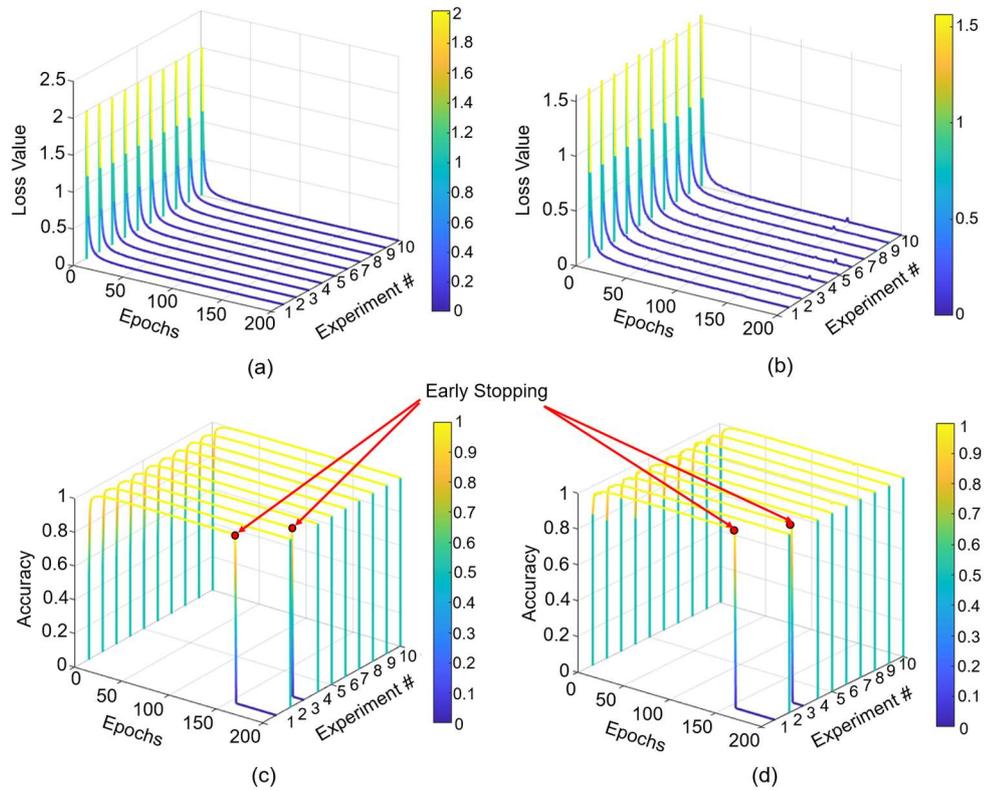


Figure 5.6. The (a) training loss, (b) validation loss, (c) training accuracy, and (d) validation accuracy convergence curves obtained during the 10 experimental trials.

5.3.3. Fault Classification Performance Analysis

To evaluate the fault diagnosis performance, the proposed methodology was compared against three state-of-the-art methods where the combination of signal processing and different types of CNN (i.e., two-dimensional and one-dimensional) were used for diagnosing mechanical faults in rotating machines. The first referenced method is a combination of a fast wavelet kurtogram used as an image-like pattern for the input of a modified LeNet-5 CNN that was utilized for diagnosing rolling element bearing faults [83] (further referred to as Kurt+LeNet5). The second technique is represented by the approach used in the field of bearing fault diagnosis that comprises the envelope power spectra extraction with bandpass filtering and the one-dimensional CNN [111] (further referred to as EPS+CNN). The last approach used for the comparison is represented by a recently published work where a one-dimensional CNN was applied for detecting rubbing faults (i.e., binary differentiation between normal and faulty conditions of the system) using the raw resampled signals collected from the multiple sensors as the inputs to the deep learning technique [141] (further referred to as MSCNN). The number of training epochs was assigned to be the same

as mentioned in Subsection 5.3.2 of this chapter for all the techniques used, including the proposed one.

The fault diagnosis performance for all the methods used in this comparison is expressed in terms of micro-averaged recall (Rec_{μ}), micro-averaged precision ($Prec_{\mu}$), micro-averaged F1-score ($F1_{\mu}$), and fault classification accuracy (FCA). It has been decided to use the averaged versions of these widely used metrics [77] (i.e., sensitivity, precision, and F1-score) due to the possible deviations in the numbers of data instances for different classes in the dataset due to the random sampling procedure. Furthermore, the application of the micro-averaged metrics can provide more fair performance comparisons in the experimental conditions when different signal processing techniques resulting in the different total number of samples are applied to the dataset. These metrics can be calculated using the equations (2.12-2.15) presented in Chapter 2 of this dissertation. The experimental results expressed in terms of performance metrics described above averaged over 10 experiment trials and their standard deviations (Std) are presented in Table 5.1

Table 5.1. The experimental results obtained for the proposed and referenced techniques.

Methods	Metrics (Std) (%)			
	Rec_{μ}	$Prec_{\mu}$	$F1_{\mu}$	FCA
Proposed	99.75 (0.09)	99.75 (0.09)	99.75 (0.09)	99.75 (0.09)
Kurt+ LeNet5	98.54 (0.74)	98.54 (0.74)	98.54 (0.74)	98.54 (0.74)
EPS+CNN	95.76 (0.34)	95.76 (0.34)	95.76 (0.34)	95.76 (0.34)
MSCNN	58.61 (9.2)	58.61 (9.2)	58.61 (9.2)	58.61 (9.2)

As can be seen from Table 5.1, the proposed methodology outperformed all the counterparts from the existing methods used in this comparison with the FCAs equal to 99.75% and standard deviation not exceeding 0.09%. Moreover, the values for other metrics demonstrated by the introduced method are higher than those achieved by the counterpart techniques. Also, the results provided by the referenced methods used in the comparison appear to be less robust, i.e., their fault diagnosis performance differs significantly when applied to different rubbing fault datasets used in this study.

More details regarding fault classification accuracies can be observed in Figures 5.7 and 5.8 where the boxplots with distributions of accuracy values and the confusion matrices obtained during 10 experiments are presented. The black cross in the boxes belonging to different methods in Figure 5.7 corresponds to the average classification accuracy values presented in Table 5.1.

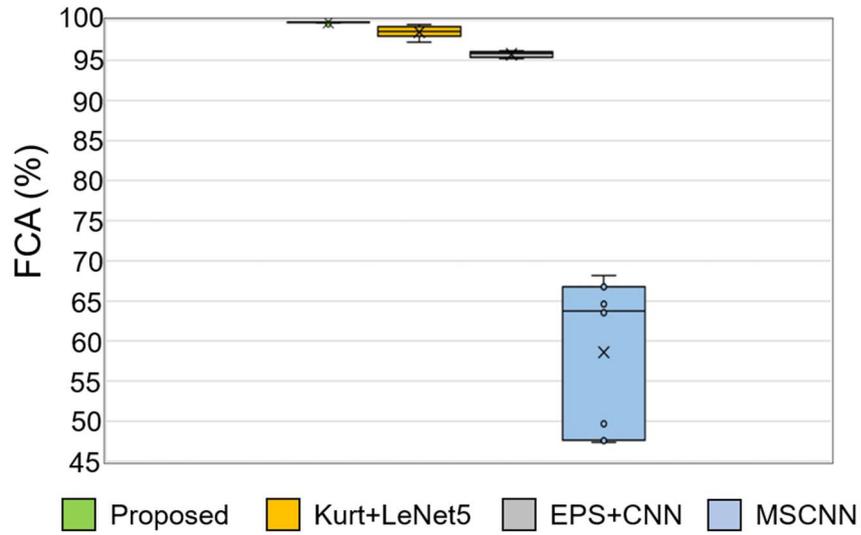


Figure 5.7. The boxplots demonstrating the statistics of the FCA metric over 10 experiments.

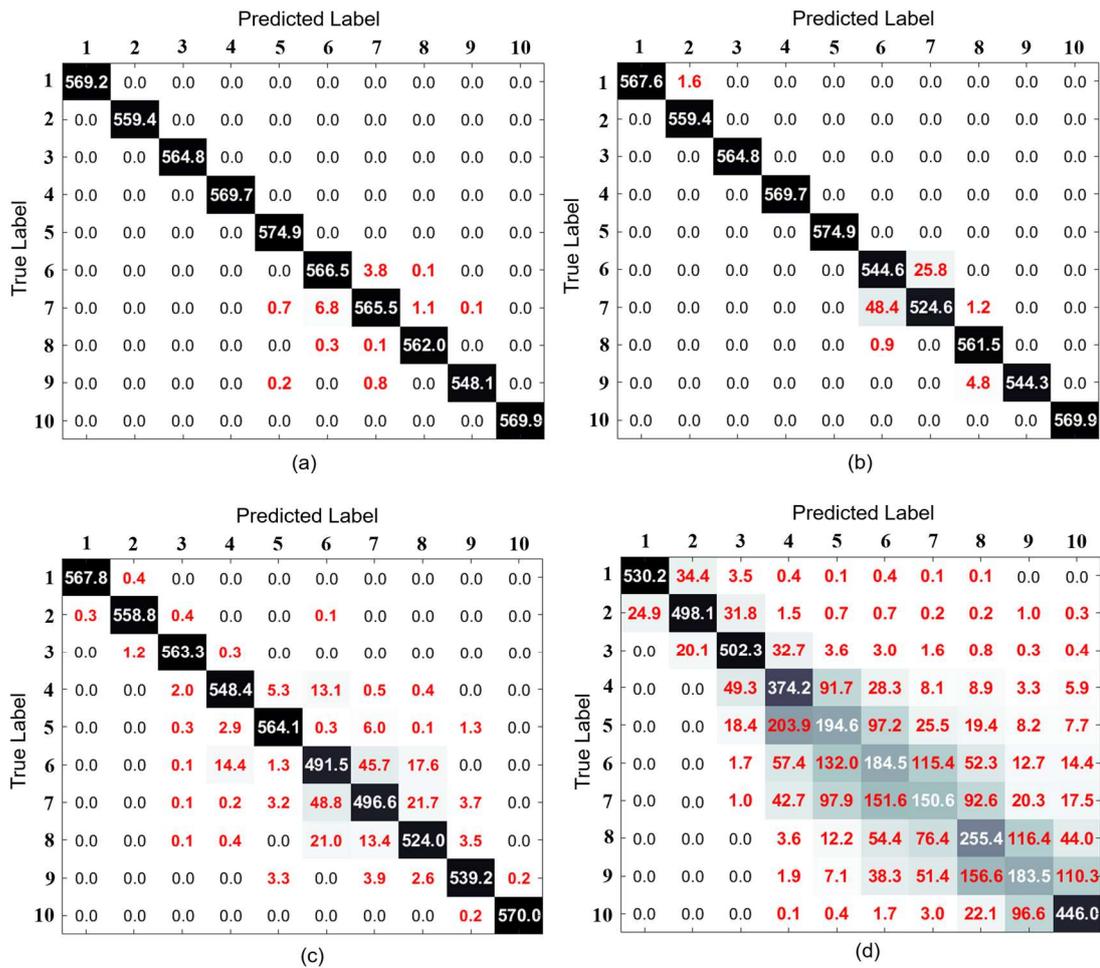


Figure 5.8. The confusion matrices obtained for (a) proposed, (b) Kurt+ LeNet5, (c) EPS+CNN, and (d) MSCNN averaged over 10 experiments.

Among the referenced methods used for the comparison, Kurt+LeNet5 demonstrated the highest fault diagnosis results and reached an average FCA of 98.54% with Std over 10 experiments equal to 0.74%, respectively. The advantage of this technique is that it utilizes a fast kurtogram [142] to create an image-like pattern from a one-dimensional signal. Kurtogram is a type of TFA technique that is frequently utilized for analyzing non-stationary signals and extracting features of non-linear faults. Hence, this signal analysis approach could extract valuable information from the vibration signal and provide recognizable patterns for different intensity levels of blade rubbing faults. However, the drawback of this approach is the time needed for computing the kurtogram representation of the vibration signal along with defining the optimal level of wavelet transformation leading to the recognizable fault patterns. Furthermore, to utilize this particular known CNN architecture, the kurtogram patterns should be resized to meet the dimensionality of the LeNet5 input layer [83] which also might lead to the loss of information and a reduction of the image quality.

Speaking about EPS+CNN, this approach achieved an average FCA value of 95.76% with a standard deviation of 0.34% during 10 experimental trials. This method utilizes the low-frequency range of the envelope power spectra computed from the signals as the inputs to one dimensional CNN. Despite the patterns used in this approach include valuable information about the rub-impact faults, such as the harmonics of the fundamental frequency and its sub-synchronous frequencies (usually considered as valuable evidence of rubbing process [92]) it seems that EPS+CNN failed to achieve a good generalization within the given number of training epochs.

Regarding MSCNN, this technique is the closest one to the proposed method among those used for the comparison as it also utilizes multiple sources of signals. From the results demonstrated in Table 5.1, it can be seen that this method reached an average FCAs of 58.61% and Std of 9.2% which are the worst results among the techniques used for the comparison in this chapter. Despite the main advantages of this approach, such as the simultaneous utilization of multiple sources of signals, it appeared that this technique has some serious limitations. First of all, it seems that this fault diagnosis approach [141] might not be suitable for extracting local features of good quality from signals collected under high rotating speeds and high sampling frequencies. The reason for these results is that the architecture of the referenced one-dimensional CNN that operates on multivariate signals consists of only two layers: the convolutional and the global max-pooling layers. The features extracted using this architecture might not be discriminative enough to characterize a set of input vibration signals of huge dimensionality, even after the resampling process. Furthermore, this approach was initially proposed for fault detection, i.e., differentiating only two states of the rotor system, such as normal and abnormal ones, whereas no variations in fault intensity levels were considered during the design procedure. Also, it is worth noticing that

in the original study, the signals from sensors of different types were utilized as the inputs to the one-dimensional CNN; whereas in this study, all the available signals were collected by sensors with multiple channels that record the displacements of the rotor in different directions.

Overall, it can be concluded that the proposed methodology is capable of accurate diagnosing coupling shaft imbalance and blade rubbing faults of various intensity levels. Furthermore, the simplicity of the vibration signal processing (i.e., pattern preparation) and the shallow architecture of the proposed ModCNN allows for rapid deployment in a real industrial environment as well as its scalable architecture allows for expanding the number of sensors for collecting the multivariable signals.

5.4 Conclusions

In this chapter, a fault identification framework based on a multivariate one-dimensional convolutional neural network (ModCNN) that utilizes data collected by multiple sensors for diagnosing coupling rotor imbalance and blade rub-impact faults of various intensities is introduced. In the proposed methodology, first, windows of a size equal to one period of the vibration time-series are simultaneously picked from the multivariate vibration signals collected using multiple sensors. Then, to perform the normalization and reduce the dimensionality of the cut windows, their envelope power spectra are calculated. Next, the obtained patterns extracted from the windows of the multivariate signal are pushed into the convolution-subsampling layer (CSL) pairs of the ModCNN to extract discriminative local features. Finally, these local features extracted from the windows of the multivariate vibration signal are merged in a fully connected layer and the fault identification is accomplished in the output layer using the SoftMax activation function. Experimental results demonstrated that the proposed ModCNN achieves fast convergence when processing multivariate vibration signals and outperforms the referenced deep learning-based state-of-the-art approaches. The small value of classification accuracy standard deviations emphasizes that the proposed approach is not sensitive to training-testing data permutations. The shallow structure of the ModCNN and its fast convergence allow the employment of the proposed method in a real industrial environment when data of multiple sensors are considered for assessing the health state of the rotor system.

Chapter 6

Hybrid Rubbing Fault Identification Using an Artificial Intelligence-based Observation Technique

6.1 Introduction

Turbines are among the most important rotating machines utilized in power plants. Their operating conditions are closely associated with high temperatures and high rotating speeds. The crucial point in the design of turbines is a small clearance between the stator and rotatory parts (i.e., the turbine blades), which is made to increase the torque and reduce air reluctance. The rubbing phenomenon appears in turbines when the rotor blades start interacting with the stator. The main causes of the rubbing phenomenon are the presence of various types of mechanical faults in the rotating machine, such as misalignment, blade extension due to the high operating temperatures, and self-excited vibrations [40]. Rubbing faults can cause excessive damage to industrial equipment, as well as increase maintenance costs. Thus, the detection and diagnosis of rubbing faults in the early stages are essential to maintain the health of rotating machines.

Rubbing faults are recognized as complex, non-stationary, and non-linear mechanical faults [41]. Due to these properties, most conventional signal analysis approaches that have been created for stationary and linear signals, such as time-domain-based or fast Fourier transform-based techniques, are not useful in extracting the discriminative features of this type of fault. Time-frequency signal analysis (TFA) techniques are preferable over conventional ones due to their ability to simultaneously analyze the characteristics of the collected signals in both domains (i.e., time and frequency).

Some examples of widely utilized TFA approaches for rubbing fault signal processing and feature extraction are the wavelet transform and its variations [44, 85, 86]. The main advantage of these techniques is a good time-frequency localization, which allows the detection of transients appearing in signals. However, the crucial drawback of wavelet-based methods is that their performance is sensitive to the selection of a mother wavelet function [87]. The selection of a mother wavelet function itself requires a series of experiments and subsequent analysis of the obtained results [47, 48].

Another family of TFA approaches that are frequently applied for rubbing signal processing is based on empirical mode decomposition (EMD) [22]. EMD is a method capable of decomposing complex multi-component time sequences into a number of well-behaving intrinsic mode functions (IMFs), where each of the IMFs corresponds to the different frequency bands of

the original signal. EMD is a completely data-driven decomposition technique that does not require any prior knowledge about the signal. Due to its advantages, EMD was successfully applied [56, 57, 59, 88, 128] to extract and analyze valuable information from rubbing fault signals. However, this technique suffers from a problem, referred to as the ‘mode-mixing’ problem, which makes it difficult to find the exact physical meaning of each of the extracted components. Some other approaches that are directly based on EMD have also been introduced [129, 143]; however, they also inherit the same problems as conventional EMD.

To cope with mode-mixing, improvements such as ensemble empirical mode decomposition (EEMD) [23] and complete ensemble empirical mode decomposition [144] were introduced. Despite some differences in these approaches, the main concept behind them is to extract ‘real’ IMF components as the mean of the EMD trials run on the signal with various types of white Gaussian noise added in each trial. These techniques appeared to successfully overcome the mode-mixing problem. Thus, these techniques have been successfully used for diagnosing faults in rolling-element bearings [62], as well as rubbing faults [63]. Despite their capabilities for extracting valuable fault features of various mechanical faults, including rub-impact faults, the application of these methods in real industrial environments is limited due to the significant computational time required by these algorithms to process the incoming mechanical signals samples. This issue is caused by the complexity of these iterative algorithms, as well as the difficulty of finding a reasonable tradeoff between the number of decomposition trials (i.e., the complexity of the algorithm) and the quality of decomposition.

Control theory has suggested different observation techniques that have been used in industry for accurate signal (fault) estimation using model-based algorithms. The stability and robustness are the main advantages of observation techniques. These observation techniques can be categorized into two main groups: a) linear and b) non-linear [103]. Linear observers, such as the autoregressive with exogenous input Laguerre (ARX-Laguerre) proportional-integral observer (PIO), have been used to estimate signals in various applications, such as bearings and robot manipulators [145, 146]. Apart from the various advantages, such as easy implementation in industrial applications and reliability, the accuracy of fault estimation and robustness are the main challenges of this technique. To address robustness issues in the ARX-Laguerre PIO (ARXLPIO), an extended-state ARXLPIO was described in [145]. To address accuracy-related problems in the ARXLPIO and extended ARXLPIO, a hybrid technique was presented in [146]. However, despite the improved accuracy and robustness in the hybrid extended ARXLPIO, modeling highly non-linear and non-stationary signals, such as those associated with rubbing faults, to design a non-linear observer is challenging. In this research, a deep learning technique is represented to construct a hybrid approach.

Once the observation technique, such as ARXLPIO, is designed and the estimated rubbing signal is obtained, the decision about the system's state should be performed. Recently, in the field of fault diagnosis, decision-making about the health condition of a rotating machine is usually accomplished through machine learning (ML)- or deep learning (DL)-based approaches. Among ML techniques utilized for fault assessment, the k-nearest neighbors (k-NNs) technique [29], support vector machines (SVMs) [89], and artificial neural networks (ANNs) [104] are the most popular solutions. However, one of the main drawbacks of these approaches is that they rely on very precise feature engineering, which is difficult when considering complex non-linear faults, such as a rub-impact fault. Without appropriate signal processing and high-quality feature extraction with feature selection, these approaches struggle to deliver good fault diagnosis results when applied to time sequences (i.e., high-dimensional feature spaces) of the acquired signals. It can be difficult for these techniques to identify hidden patterns and dependencies in data sequences. Thus, approaches capable of extracting features or patterns directly from the signals are favorable for industrial applications. Due to this problem, DL-based methods have drawn significant attention; these methods are focused on learning data representations themselves to achieve a target goal, unlike ML-based approaches. The most popular DL-based solutions used for solving various fault diagnosis problems use deep neural networks (DNNs) [113], convolutional neural networks [83] (CNNs) (for fault diagnosis), and various types of autoencoders (AEs) [147] (for feature extraction and data augmentation). The main ideas behind DL-based approaches are similar to those of ANNs; during training, they adjust the weights of neurons and bias terms while tuning hyperparameters to meet the target requirements. However, the main difference between DNNs and ANNs is that DNNs utilize large numbers of hidden layers and nodes to learn specific patterns from the data representation. DNNs are capable of resolving very complex problems and achieving high generalization performance; however, to learn a large number of deep network parameters, it is necessary to use huge datasets. Furthermore, when designing the DNN for learning to extract valuable discriminative patterns when directly working with signals, there exists a challenge of adjusting the architecture of the network in general and the dimensionality of its hidden layers specifically due to the variation of rotating speeds and differences in signal sampling rates during the data acquisition procedure (uncertainty). Therefore, this chapter introduces the robust scalable DNN (S-DNN) which is suitable for working with 1D signals, i.e., the network that includes layers the number and dimensionality of which adjusts according to the dimensionality of the input signal and the complexity of solving fault diagnosis problem. This network is used in conjunction with an estimated signal resampling approach to improve the fault diagnosis capabilities of the conventional linear ARXLPIO and identify rubbing faults of different intensity levels. Specifically, the signal resampling technique is first applied to the rubbing signal estimated by the linear observation technique to extend the dataset and obtain more data instances for training, validation, and testing tasks. Next, these resampled estimated signals are used as inputs for the S-DNN to

make a decision about the system’s state. The main contributions of this chapter are introducing a flexible hybrid method based on simple linear observation technique (ARXLPIO) and DL (S-DNN) that can be applied for diagnosing nonlinear rub-impact faults in industrial applications along with a specific algorithm for designing the adaptive architecture of S-DNN.

The remainder of this chapter is organized as follows. Section 6.2 provides insights into the proposed hybrid fault diagnosis methodology. Section 6.3 presents the experimental validation of the introduced technique and discusses the obtained results. Finally, Section 6.4 contains concluding remarks.

6.2 Proposed Methodology

The proposed rub-impact fault identification framework is demonstrated in Figure 6.1. It consists of three important steps: data collection, system modeling and signal estimation, and decision-making. First, the rub-impact fault signals of different intensity levels are collected from the experimental testbed. Next, these vibration signals are used to build a linear ARXLPIO to estimate the rubbing signals in normal and abnormal conditions. To perform decision-making about the condition of the system and improve the performance of the ARXLPIO, the S-DNN is utilized for decision-making during the final step of the algorithm. To use the S-DNN effectively, the signals estimated by observation techniques are first resampled concerning the revolutions of the rotor, and their frequency power spectra are obtained using a fast Fourier transform (FFT).

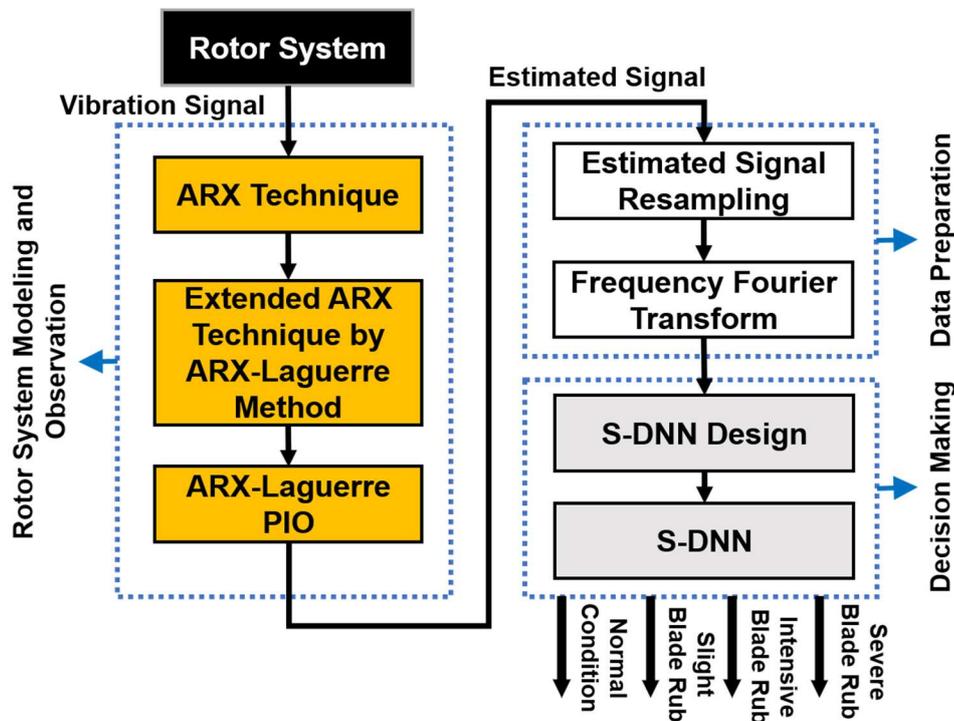


Figure 6.1. Proposed hybrid rub-impact fault diagnosis approach.

Finally, the patterns that appeared in the power spectra of the resampled estimated signals are used as inputs for the S-DNN to identify rub-impact faults of various intensity levels.

6.2.1. ARX-Laguerre PI-observer (ARXLPIO)

The mathematical formulation of the rubbing phenomenon is nonlinear and complicated. Also, the dynamic behavior in theoretical and practical applications may be different, especially in uncertain conditions; this causes challenges in system modeling. To decrease the challenge of system modeling and estimation for rubbing, the ARXLPIO technique is depicted. This technique can be designed based on the following steps. In the first step, the ARX technique for the vibration signal of rubbing is defined. Next, to improve the robustness and reliability, the ARX-Laguerre procedure is represented, as shown in the second step. To improve the accuracy of the signal estimation and increase the differentiation between normal and abnormal conditions, the ARXLPIO can be represented as shown in the third step. The ARX technique for rubbing signal estimation is formalized by the following equation [103, 145, 148].

$$\bar{Y}(k+i) = \sum_{i=1}^{n_{\bar{y}}} \varepsilon_{\bar{y}}(i) \bar{Y}(k) + \sum_{i=1}^{n_u} \varepsilon_u(i) U(k). \quad (6.1)$$

Here, $U(k)$, $\bar{Y}(k)$, $(n_{\bar{y}}, n_u)$, and $(\varepsilon_{\bar{y}}(i), \varepsilon_u(i))$ are the input, output, order, and model parameters, respectively. Regarding (6.1), the model parameters $(\varepsilon_{\bar{y}}(i), \varepsilon_u(i))$ are represented by the following equation [148].

$$\begin{cases} \varepsilon_u(i) = \sum_{n=0}^{\infty} q_{n,u} W_n^u(i, C_u) \\ \varepsilon_{\bar{y}}(i) = \sum_{n=0}^{\infty} q_{n,\bar{y}} W_n^{\bar{y}}(i, C_{\bar{y}}) \end{cases}, \quad (6.2)$$

where $(W_n^u(i, C_u), W_n^{\bar{y}}(i, C_{\bar{y}}))$ are n -th orthonormal functions, $(q_{n,u}, q_{n,\bar{y}})$ are the Fourier decomposition coefficients, and $(C_u, C_{\bar{y}})$ are orthonormal bases. To improve the robustness, ARX-Laguerre signal estimation is represented. Therefore, the ARX orthonormal can be defined by the following equation.

$$\bar{Y}(k+i) = \sum_{n=0}^{(n_{\bar{y}})-1} q_{n,\bar{y}} \partial_{n,\bar{y}_o}(k, C_{\bar{y}}) + \sum_{n=0}^{(n_u)-1} q_{n,u} \partial_{n,u}(k, C_u). \quad (6.3)$$

Here, $\partial_{n,u}$ and ∂_{n,\bar{y}_o} are the orthonormal functions for the input and output, respectively, which can be represented based on the following equation.

$$\begin{cases} \partial_{n,\bar{y}_o}(k, C_{\bar{y}}) = \sum_{i=1}^{\infty} W_n^{\bar{y}}(i, C_{\bar{y}}) \bar{Y}(k-i) \\ \partial_{n,u}(k, C_u) = \sum_{i=1}^{\infty} W_n^u(i, C_u) U(k-i) \end{cases} \quad (6.4)$$

In addition, the Laguerre functions for the input (Y_n^u) and output ($Y_n^{\bar{y}}$) are represented by the following equation.

$$\begin{cases} Y_n^u = \frac{\sqrt{1-C_u^2}}{Z-C_u} \left(\frac{1-C_u z}{Z-C_u}\right)^n \\ Y_n^{\bar{y}} = \frac{\sqrt{1-C_{\bar{y}}^2}}{Z-C_{\bar{y}}} \left(\frac{1-C_{\bar{y}} z}{Z-C_{\bar{y}}}\right)^n \end{cases} \quad (6.5)$$

Therefore, based on (6.4) and (6.5), the modified orthonormal functions for the input ($\bar{\partial}_{n,u}(k, C_u)$) and output ($\bar{\partial}_{n,\bar{y}_o}(k, C_{\bar{y}})$) can be represented by the following formulation.

$$\begin{cases} \bar{\partial}_{n,\bar{y}_o}(k, C_{\bar{y}}) = \sum_{j=1}^{\infty} Y_n^{\bar{y}}(k) \bar{Y}(k) \\ \bar{\partial}_{n,u}(k, C_u) = \sum_{j=1}^{\infty} Y_n^u(k) U(k) \end{cases} \quad (6.6)$$

Thus, if the filtering for the output and input signals is defined by $(W(k), I(k))$, the ARX-Laguerre rubbing modeling and estimation based on the vibration signals are represented by the following [103, 145, 148].

$$\bar{Y}(k+i) = \sum_{n=0}^{(n_{\bar{y}})-1} q_{n,\bar{y}} \bar{\partial}_{n,\bar{y}_o}(k, C_{\bar{y}}) W(k) + \sum_{n=0}^{(n_u)-1} q_{n,u} \bar{\partial}_{n,u}(k, C_u) I(k). \quad (6.7)$$

Therefore, the state-space ARX-Laguerre method is denoted by the following equation.

$$\begin{cases} \bar{X}(k+1) = [\alpha_{\bar{X}} \bar{X}(k) + \alpha_{\bar{Y}} \bar{Y}(k) + \alpha_U U(k)] + e(k) \\ \bar{Y}(k+1) = (\alpha_O)^T \bar{X}(k) \end{cases} \quad (6.8)$$

Here, $\bar{X}(k), \bar{Y}(k), U(k), e(k), (\alpha_{\bar{X}}, \alpha_{\bar{Y}}, \alpha_U)$, and $(\alpha_O)^T$ are the rubbing state; measured rubbing output; measured rubbing input; error signal; state, output, and input coefficients; and Fourier coefficients of the rubbing output, respectively. The error signal is represented by the following equation.

$$e(k) = Y(k) - \bar{Y}(k). \quad (6.9)$$

Here, $Y(k)$ is an original rubbing signal. Moreover, the state, output, and input coefficients are represented by the following.

$$\alpha_{\bar{X}} = \begin{bmatrix} \alpha_{\bar{X}o} & O_{n_u, n_{\bar{y}}} \\ O_{n_{\bar{y}}, n_u} & \alpha_{\bar{X}i} \end{bmatrix}. \quad (6.10)$$

Here, $O_{n_u, n_{\bar{y}}}$ and $O_{n_{\bar{y}}, n_u}$ are null matrices. The values of $\alpha_{\bar{X}o}$ and $\alpha_{\bar{X}i}$ are characterized by (6.11) and (6.12).

$$\alpha_{\bar{X}o} = \begin{bmatrix} C_{\bar{y}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ (-C_{\bar{y}})^{(n_{\bar{y}})-1}(1 - C_{\bar{y}}^2) & \cdots & C_{\bar{y}} \end{bmatrix} \quad (6.11)$$

$$\alpha_{\bar{X}i} = \begin{bmatrix} C_u & \cdots & 0 \\ \vdots & \ddots & \vdots \\ (-C_u)^{(n_u)-1}(1 - C_u^2) & \cdots & C_u \end{bmatrix} \quad (6.12)$$

$$\alpha_{\bar{Y}} = \sqrt{1 - C_{\bar{y}}^2} \begin{bmatrix} 1 \\ -C_{\bar{y}} \\ -C_{\bar{y}}^2 \\ \vdots \\ (-C_{\bar{y}})^{(n_{\bar{y}})-1} \end{bmatrix} \quad (6.13)$$

$$\alpha_U = \sqrt{1 - C_u^2} \begin{bmatrix} 1 \\ -C_u \\ -C_u^2 \\ \vdots \\ (-C_u)^{(n_u)-1} \end{bmatrix} \quad (6.14)$$

To improve the performance (accuracy) of signal estimation in the ARX-Laguerre technique, the PIO is introduced. The state-space ARXLPIO is represented by the following equation.

$$\begin{cases} \widehat{\bar{X}}(k+1) = [\alpha_{\bar{X}}\widehat{\bar{X}}(k) + \alpha_{\bar{Y}}\widehat{\bar{Y}}(k) + \alpha_U U(k)] + \widehat{e}(k) + \alpha_p(Y(k) - \widehat{\bar{Y}}(k)) \\ \widehat{\bar{Y}}(k+1) = (\alpha_o)^T \widehat{\bar{X}}(k) \end{cases} \quad (6.15)$$

To increase the accuracy of $\widehat{\bar{Y}}(k+1)$, the accuracy of $\widehat{\bar{X}}(k)$ must be improved. Regarding (6.15), to increase the accuracy of $\widehat{\bar{X}}(k)$, various parameters have important roles. One of the main

parameters is $\widehat{e(k)}$. Equation (16) represents the estimation of $e(k)$ based on the ARX-Laguerre PIO technique.

$$e(\widehat{k+1}) = \widehat{e(k)} + \alpha_i(Y(k) - \widehat{Y(k)}). \quad (6.16)$$

Here, $\widehat{X(k)}$, $\widehat{Y(k)}$, $\widehat{e(k)}$, and (α_p, α_i) are the state estimation based on the ARXLPIO, signal output estimation based on the ARXLPIO, signal error (e.g., uncertainties and fault) estimation based on the ARXLPIO, and ARXLPIO coefficients used to improve the signal's estimation accuracy, respectively. After estimating the signals based on the ARXLPIO, the residual signal can be calculated as follows.

$$r_{\widehat{Y(k)}} = Y(k) - \widehat{Y(k)}. \quad (6.17)$$

Various techniques have been used to increase the accuracy of the ARXLPIO, such as the extended ARXLPIO [145] and fuzzy extended ARXLPIO [146]. In this research, the S-DNN is used to improve the fault identification capabilities of the designed observation method to increase the accuracy of fault detection and diagnosis of the rubbing faults.

6.2.2. Artificial Intelligence-based Approach for Blade Rub Fault Identification

6.2.2.1. Signal Resampling Process

Once the estimated signal of the rubbing fault is obtained by the ARXLPIO technique, this estimated signal is used to detect rubbing faults in a system and identify the intensity. However, since the deep learning networks learn from data representations, it is essential to convert the estimated time-domain vibration signal into a meaningful and discriminative representation. To convert the estimated signal into a meaningful form and help the designed DNN achieve good generalization on the target task, first, the estimated rub-impact fault signals are resampled into a series of windows, where each window has the length of the time-domain samples corresponding to one revolution of the shaft. It is important to notice that in this chapter the linear signal approximation technique (i.e., ARX-Laguerre) and signal estimation algorithm (i.e., PIO) are used. Therefore, due to the linearity of those techniques, their sampling rates are optimized by the observer: by increasing or decreasing the sampling rates, the accuracy of signal estimation changes. Due to this behavior of observer, the numbers of data points in the raw vibration signals collected from the testbed and ones estimated by ARXLPIO are different such as 65,536 points for the raw and 16,949 points for the estimated signals, respectively.

To resample the initially estimated signals, the same procedure as was described in previous chapters is applied. First, the number of revolutions performed in one second (RPS) should be calculated using the following formulation:

$$RPS = \frac{RPM}{60} = \frac{2500}{60} = 41.6, \quad (6.18)$$

where RPM is the rotational speed in RPM that is used during data collection. After computing the number of revolutions per second, the time (6.19) and the number of signal points (6.20) needed for one revolution can be obtained as shown below:

$$TFOR = \frac{1}{RPS} = \frac{1}{41.6} \approx 0.024, \quad (6.19)$$

$$w_{length} = f_{sampling} \times TFOR = 65536 \times 0.024 \approx 1572. \quad (6.20)$$

Here, $TFOR$ corresponds to the time of one revolution expressed in seconds, w_{length} is the length of each new instance of the resampled estimated signal expressed in a number of data points, and $f_{sampling}$ is the sampling frequency used during data acquisition.

To ensure that each of the new resampled instances has a relation with the previous one, the estimated signals are cut into windows of 1,572 data points with a 15% overlap, as shown in Figure 6.2. After resampling the estimated signals with the overlap procedure described above the resampled version of the dataset containing estimated signals was of a size of 7,470 instances in total with 747 samples per class, respectively.

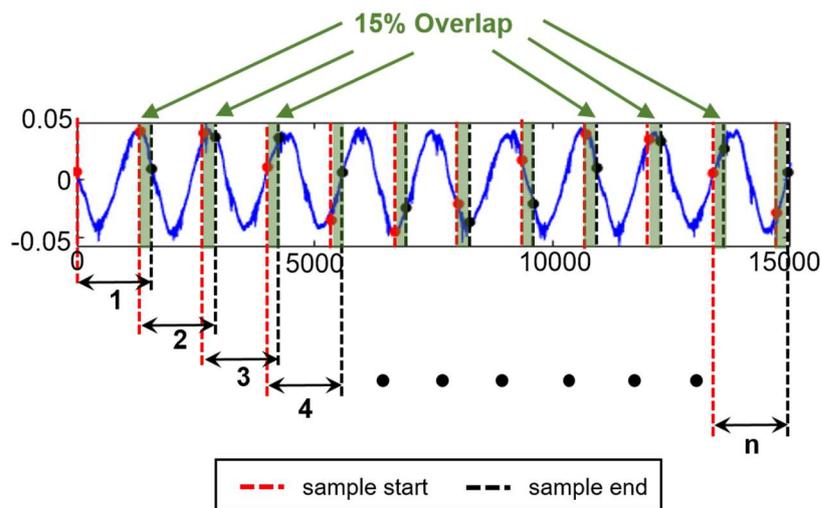


Figure 6.2. The estimated signal resampling process with 15% overlap.

Once resampling is completed, the obtained time-domain windowed instances are converted into the frequency domain using a fast Fourier transform (FFT) to help the S-DNN learn patterns from the estimated signal and achieve high accuracy for the task of rub-impact fault identification. Example frequency spectra computed for the resampled estimated signals in the normal condition (i.e., when neither rubbing nor imbalances are present in the system) and abnormal conditions are demonstrated in Figure 6.3. As can be seen from this figure, the frequency spectra of signals in normal and faulty conditions show the presence of frequency harmonics in the same frequency ranges but differ in terms of the amplitude, which means that this type of signal representation is discriminative and can be used as an input for the deep learning-based technique to learn the patterns corresponding to various signal classes and perform decision-making. Note that since the FFT frequency spectrum is symmetric, it is recommended to cut off the mirrored part of the frequency spectrum to reduce the input size of the data vectors before fault identification.

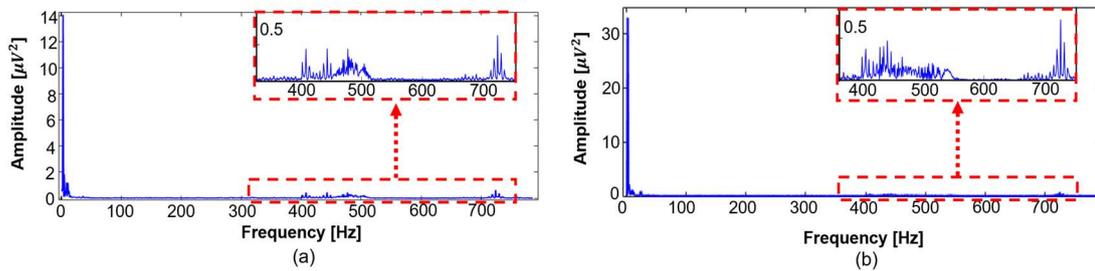


Figure 6.3. The frequency power spectra of the resampled estimated signals corresponding to (a) the normal system condition and (b) the severe rubbing fault system condition.

6.2.2.2. Scalable Deep Neural Network (S-DNN) Design

To improve the performance of the linear ARXLPIO and perform rub-impact fault identification, the S-DNN of the triangular architecture type (i.e., the number of neurons in each layer is smaller than in the previous one when moving through the network from its input towards the output layer) is utilized in this chapter. As the input data, the frequency power spectra of the resampled estimated signals are used. The main feature of S-DNN is the presence of scalable layers. It means, that the dimensionality of these layers depends on the number of neurons in the previous layer and reduces from layer to layer according to a reduction rate which is a tunable selectable parameter of the model. Additionally, the number of scalable layers in the proposed network represents the second adjustable parameter that can be chosen during the design process.

The process of designing the proposed S-DNN starts from defining the activation functions of hidden and output layers along with the loss function based on the knowledge about input data type and type of problem that should be resolved (in this chapter, this is the problem of fault classification). Next, the optimization algorithm should be chosen for network training. Finally, the parameters of S-DNN such as the number of scalable hidden layers and their reduction rate of

nodes that allow resolving the defined problem with the highest value of evaluation metric should be selected according to the proposed algorithm that will be introduced in this section.

As mentioned above, first, the activation functions for the layers of the neural network should be defined. In this chapter, for all the layers of the future S-DNN, except for the output layer, the activation function is set to be the ‘rectified linear unit’ (ReLU) activation function. There are two main reasons for choosing this activation function for the input and hidden layers. 1) The representation form of the input data (i.e., the power spectra of the estimated signals) does not allow it to have any negative values; thus, the lower limit of the activation can be set to zero without concerns related to losing valuable information. 2) The ReLU activation function is non-saturating, which eases the computation of the loss function gradients during the training process. As the activation function of the output layer, the SoftMax activation function is chosen due to its ability to estimate the probability that a particular data instance belongs to a specific class by computing the scores of each class for this data sample. The ReLU and SoftMax activation functions can be described using the following equations:

$$f_{ReLU}(x) = \max(0, x), \quad (6.22)$$

$$\hat{P}_m = \exp(s_m(x)) / \sum_{i=1}^M \exp(s_i(x)), \quad (6.23)$$

where M is the total number of classes and $s(x)$ is a vector with the scores of every available class for the specific data sample x . The input data sample is assigned to the class with the highest estimated probability \hat{P}_m (i.e., the class that has the highest computed score for this instance). The connection weights of the layers of the S-DNN are initialized using Glorot initialization [126] with a normal distribution.

To perform the training process, the outputs of the SoftMax activation are used with the categorical cross-entropy loss function to resolve problems related to multi-class classification. The categorical cross-entropy loss function can be formulated as:

$$Loss(\theta) = -\frac{1}{n} \sum_{i=1}^n \sum_{m=1}^M y_m^i \log(\hat{P}_m). \quad (6.24)$$

Here, θ is the set of model parameters and y_m^i and \hat{P}_m are the target and estimated probabilities that the $i - th$ data sample belongs to the class m , respectively.

As the optimizer, this chapter employs the Adam [127] optimization algorithm, which states for ‘adaptive moment estimation.’ This optimization method is one of the most popular

optimizers due to its specific properties. For example, Adam optimizer keeps track of the exponentially decaying average value of the past gradients of the loss function. It also keeps track of the exponentially decaying average of the past squared gradients of the loss function. The learning rate in this work is equal to 0.001.

Once the activation functions and the parameters of the optimization algorithm are fixed, the process of building the S-DNN. Specifically, at this step, the numbers of hidden scalable layers and the node reduction rate that allow for achieving the best fault classification performance should be selected. In this chapter, the algorithm for selecting the parameters mentioned above is proposed and represented as below:

1) Define the numbers of input and output layer nodes: n_{inp}, n_{out}

2) Define the ranges for the parameters of node reduction rate m and number of hidden layers h : $m \in \{2,3,4, \dots, M\}$ and $h \in \{1,2,3,4,5, \dots, H\}$.

3) Create the lookup grid L of size $(M - 1) \times H$ where each cell contains the unique $\langle m, h \rangle$ combination, as shown in Table 6.1, and the empty search grid G of the same size. Randomly permute the elements of L .

Table 6.1. The example of lookup grid L .

	1	2	3	4	...	H
1	$\langle m, h \rangle$		$\langle m, h \rangle$			
2	$\langle m, h \rangle$		$\langle m, h \rangle$			
3	$\langle m, h \rangle$. . .	$\langle m, h \rangle$			
4	$\langle m, h \rangle$		$\langle m, h \rangle$			
5	$\langle m, h \rangle$		$\langle m, h \rangle$			
.			.		.	.
.			.		.	.
.			.		.	.
M-1	$\langle m, h \rangle$		$\langle m, h \rangle$			

Note, in the heads of the rows and columns of the grid the indices are located, not the actual values of m and h .

4) Check the satisfaction of the network triangularity condition by the network architectures $A(m, h)$ formed using the $\langle m, h \rangle$ pairs from L . Triangularity is held when each hidden layer $i \in \{1, \dots, h\}$ (excluding the input and output ones) of the architecture $A(m, h)$ satisfies the following criteria:

$$\begin{aligned}
 & n_i > n_{out} \wedge n_i \neq 0 \\
 & s. t. n_i = n_{i-1} - n_{i-1} * (1/m),
 \end{aligned}
 \tag{6.25}$$

where n_i is the number of nodes in the hidden layer i .

5) Fill the cells of the search grid G with zeros for $\langle m, h \rangle$ pairs that hold the triangularity condition and with ones for those pairs that do not. The example of the search grid G view after this step is presented in Table 6.2.

Table 6.2. The example of search grid G .

	1	2	3	4	...	H
1	0	0	0	0		0
2	0	0	0	0		0
3	0	0	0	0	. . .	0
4	0	0	0	0		0
5	1	1	0	1		0
.			.		.	.
.			.		.	.
.			.		.	.
M-1	0	0	0	0		0

6) Run Astar [149] algorithm on the search grid G with the starting point (1,1) and the ending point $(M - 1, H)$. In an extremal case when starting and/or ending points in G are not available (the value of the cell is equal to 1), randomly permute the elements of L and their respective cells in G holding the relation between the contents of the cells in both grids. The Astar algorithm will provide the indices of the cells in G corresponding to the found optimal path P between starting and ending points.

7) Using L match the coordinates of cells from the P in G with actual $\langle m, h \rangle$ values. Build the network architectures $A\langle m, h \rangle$, train them for a small number of epochs (equal to 25 in this chapter) to observe the tendency and validate them using the validation subset.

8) Pick the architecture $A_*(m, h)$ with the smallest validation loss value. Use this architecture to build and train the actual network and to adjust the model weights with biases. If several architectures deliver similar validation losses, it is recommended to choose the architecture with the smaller value of h to reduce computational complexity.

Using the algorithm of selecting numbers of hidden layers and nodes proposed in this Subsection, it is possible to build an architecture of the DNN to solve the specific problem given the information about the dimensionality of the input data, a number of target classes, and ranges of the m and h parameters in an autonomous manner. Here, there are some suggestions for selecting these ranges since both m and h values affect the number of parameters that should be tuned during the training of the actual S-DNN. It is recommended to keep the value of M equal to 10 to avoid a small reduction of node numbers in hidden layers. The maximum number of hidden layers H is limited by the computational capabilities of the target system. The practical example

and the architecture of actual S-DNN created for solving the problem of blade rub-impact fault identification in this chapter using the proposed methodology is presented in Section 6.3.

6.3. Experimental Results and Discussion

6.3.1. Training, Validation, and Testing Subsets Configurations

To evaluate the fault diagnosis capabilities of the proposed hybrid approach, the resampled dataset corresponding to the estimated rubbing fault signals of different intensity levels (which contained 7,470 data samples) was randomly split into training, validation, and testing subsets. The complete process of subset creation is as follows. First, the whole dataset was separated into training and testing subsets at a ratio of 8:2. Then, the obtained training subset was again divided into training and validation subsets using the same proportion as in the previous step. Thus, the overall training, validation, and testing subset configurations can be summarized as follows. The training subset consisted of 4,780 samples, the validation subset comprised 1,196 instances, and the remaining 1,494 unseen data instances were used for testing the proposed solution. To avoid the effect of randomness, the experimental validation in this section has been performed 10 times and at each experimental trial, the training, validation, and testing subsets were sampled randomly according to the procedure described above.

6.3.2. Case Study of Building the S-DNN

In this subsection, the details on building the S-DNN are presented. Before starting the process of creating S-DNN, let us summarize the initial conditions that are utilized to select the architecture best suitable for solving the classification problem in this work. The number of the nodes in the input layer n_{inp} is equal to 787 to match the dimensionality of the input data vectors that contain the frequency power spectra of resampled signals without a mirrored part. The number of nodes in the output layer n_{out} is equal to 10 which is the same as the number of signal classes used in this study. To start the selection of m and h parameters, the values of M and H have been assigned to be equal to 10, and 15, respectively. Thus, the first two steps of the S-DNN building algorithm are completed. At step 3 of the algorithm, the lookup grid L and search grid G of size 9×15 are created. It can be calculated, that at this step of the algorithm 135 unique candidate architectures $A(m, h)$ exist. Next, after checking whether these candidates satisfy the condition of the network triangularity, we are left with 121 possible architectures that are marked as zeros are left and 14 unsuccessful architectures marked as ones in the search grid G (steps 4-5). However, training and validating all the 121 architectures of DNN to find the best one is a time and computationally expensive process. Hence, the Astar algorithm is applied to the search grid G to find the shortest path from the start point to the endpoint of the grid while avoiding the architectures

that do not satisfy the triangularity condition. The idea of employing the Astar method at step 6 of the S-DNN building algorithm is to reduce the time for finding the optimal architecture while examining the search space. The path P delivered by the Astar method contained indices of 15 cells of the search grid G that satisfied the network triangularity condition. At step 7, using the indices from path P , the actual candidate architectures $A(m, h)$ that should be evaluated using the validation subset to find the best one are recovered. For this, each candidate architecture was trained during the 25 epochs and their validation losses were recorded and presented in Figure 6.4. From this figure, it can be seen, that the candidate architecture $A(10,4)$ demonstrated the smallest value of the validation loss among the other candidates. Hence, the parameters $m = 10$ and $h = 4$ are chosen to build the actual S-DNN.

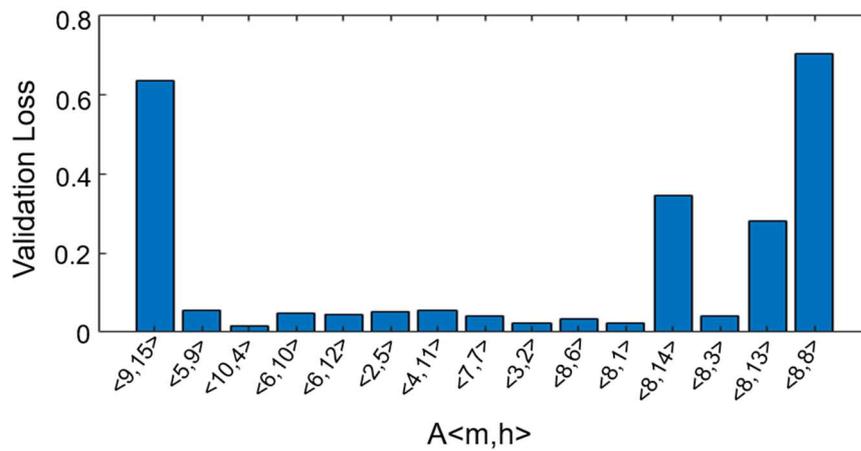


Figure 6.4. Validation losses of candidate network architectures.

The architecture of the S-DNN selected and built using the proposed algorithm is tabulated in Table 6.3. As can be seen from the table, the algorithm for selecting numbers of hidden scalable layers and node number introduced in this Subsection delivered a neural network model with a triangular architecture that is described as $A_*(10,4)$. Thus, it consists of one input layer, one output layer, and four hidden fully connected scalable layers. The number of nodes in hidden scalable layers decreases by 10 % when going deeper from the input layer to the output one.

To prevent overfitting during the training of the actual selected configuration of S-DNN, dropout regularization with a rate of 0.1 to the input layer and the next four scalable hidden layers is applied in this chapter. Specifically, at each step of the training, 10% of the nodes in the aforementioned layers are randomly ‘dropped out’ from the gradient computation process. This method allows the network to avoid overfitting and better generalize previously unseen data since each neuron of the network learns to be as useful as possible on its own because it cannot always rely on cooperation with neighboring nodes. Additionally, the early stopping approach is employed in this study. The main idea of early stopping is to stop the training process once the validation loss stops decreasing or starts increasing after a certain standby time. Early stopping, in combination with dropout regularization, significantly decreases the chance that the proposed S-DNN model will overfit the training data.

Table 6.3. The Architecture of S-DNN selected using the proposed algorithm.

Layer Type, #	Node Number	Activation Function
Input, #1	n_{inp}	ReLU
Hidden (scalable), #2	$n_2 = n_{inp} - (1/m) \cdot n_{inp}$	ReLU
Hidden (scalable), #3	$n_3 = n_2 - (1/m) \cdot n_2$	ReLU
Hidden (scalable), #4	$n_4 = n_3 - (1/m) \cdot n_3$	ReLU
Hidden (scalable), #5	$n_5 = n_4 - (1/m) \cdot n_4$	ReLU
Output, #6	n_{out}	SoftMax

n_{inp} is the number of nodes in the input layer (787 nodes); n_{out} is the number of nodes in the output layer (10 nodes); m is the reduction rate used for assigning the number of nodes in scalable hidden layers (equal to 10).

6.3.3. Training and Validation of S-DNN Model

Before evaluating the fault diagnosis capabilities of the proposed methodology, the actual S-DNN with the chosen architecture should be trained for assessing the condition of the rotor system. In this subsection, the S-DNN has trained with the frequency spectra of resampled estimated signals from the training and validation subsets. The convergence curves of loss functions and classification accuracies obtained during 10 experimental trials on training and validation subsets are demonstrated in Figure 6.5.

As can be seen from Figure 6.5 (c-d), from the beginning of the training process, training and validation accuracy curves sharply increase within the first 10 epochs until they reach an accuracy level near 99%. Then, for the next 40 epochs, some fluctuations can be observed in both training and validation accuracy curves after which they converge to an accuracy level of 99.7%. The training and validation loss curves depicted in Figure 6.5 (a-b) demonstrate similar behavior as accuracy curves but in a decreasing direction. The ‘steps’ in curves in this figure correspond to the activation of the early stopping algorithm.

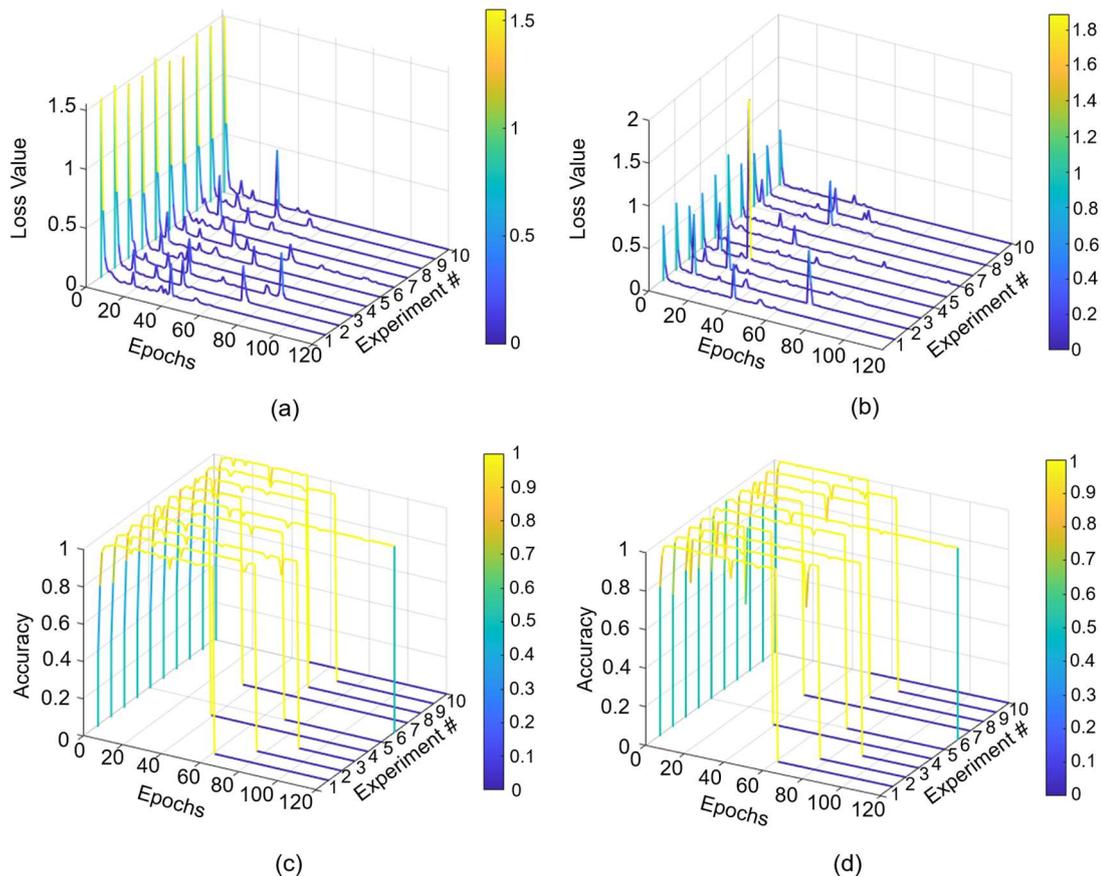


Figure 6.5. The (a) training loss, (b) validation loss, (c) training accuracy, and (d) validation accuracy convergence curves obtained during the 10 experimental trials.

Overall, it can be concluded, that the proposed fault diagnosis framework demonstrates stable convergence patterns in each experimental trial which means that it is capable of delivering stable classification results.

6.3.4 Fault Diagnosis Performance Analysis

To demonstrate the effectiveness of the proposed hybrid rub-impact fault diagnosis methodology, its classification accuracy is compared against four counterpart methods representing two different experimental scenarios. In the first scenario, it is reasonable to evaluate

how ARXLPIO and S-DNN operate separately. First, the S-DNN is applied directly to the frequency spectra of the resampled raw vibration signals collected from the testbed without signal estimation performed by ARXLPIO; this method is subsequently referred to as RAW+S-DNN. In the second approach from this experimental scenario, the only ARXLPIO is applied to estimate the blade rub-impact fault signal, and the control theory-based fault diagnosis method using the residual signal characterization and machine learning algorithm is applied to classify the fault types, as was proposed in [103]; this technique is subsequently referred to as the ARXLPIO. In the second experimental scenario, the applicability of S-DNN for the decision-making process is evaluated and its performance is compared against other deep learning-based techniques. For this, in the first method of this experimental scenario, the S-DNN in the proposed fault diagnosis pipeline is replaced with the one-dimensional CNN that has been successfully applied for bearing fault classification in [111]. This combination is further referred to as ARXLPIO + CNN. And in the second counterpart method from this scenario, S-DNN is replaced with sparse stacked autoencoders (SSAE) [150] and this approach will be further referred to as ARXLPIO+SSAE.

The fault diagnosis performance for all the approaches used in this Chapter for the comparison is expressed in terms of micro-averaged recall (Rec_{μ}), micro-averaged precision ($Prec_{\mu}$), micro-averaged F1-score ($F1_{\mu}$), and fault classification accuracy (FCA). It has been decided to use the micro-averaged versions of these widely used metrics [77] (i.e., sensitivity, precision, and F1-score) due to the possible deviations in the numbers of data instances for different classes in the dataset due to the random sampling procedure. Furthermore, the application of the micro-averaged metrics can provide more fair performance comparisons in the experimental conditions when different signal processing techniques resulting in the different total number of samples are applied to the dataset. These metrics can be calculated using the equations (2.12-2.15) presented in Chapter 2 of this dissertation. The experimental results expressed in terms of performance metrics described above averaged over 10 experiment trials and their standard deviations (Std) are presented in Table 6.4.

Table 6.4. The experimental results obtained for the proposed and referenced techniques.

Methods	Metrics (Std) (%)			
	Rec_{μ}	$Prec_{\mu}$	$F1_{\mu}$	FCA
Proposed	99.79 (0.11)	99.79 (0.11)	99.79 (0.11)	99.79 (0.11)
ARXLPIO+SSAE	99.78 (0.11)	99.78 (0.11)	99.78 (0.11)	99.78 (0.11)
ARXLPIO+CNN	99.3 (0.15)	99.3 (0.15)	99.3 (0.15)	99.3 (0.15)
RAW+S-DNN	97.7 (0.13)	97.7 (0.13)	97.7 (0.13)	97.7 (0.13)
ARXLPIO	89.6 (3.7)	89.6 (3.7)	89.6 (3.7)	89.6 (3.7)

The results provided in Table 6.4 demonstrate that the proposed hybrid approach achieved high rubbing fault classification accuracy with an average result of 99.79% over 10 experiments. One of the counterpart methods, ARXLPIO+SSAE, demonstrated a similar averaged FCA value – 99.78%, while the remaining approaches used for the comparison, i.e., ARXLPIO+CNN, RAW+S-DNN, and ARXLPIO methods, reached averaged FCA values 99.3%, 97.7%, and 89.6%, respectively. Furthermore, the statistics of FCA metric distributions over 10 experiments for each method used in the comparison are presented in form of boxplots in Figure 6.6. The black cross in each boxplot corresponds to the mean value of the FCA metric which is presented in Table 6.4.

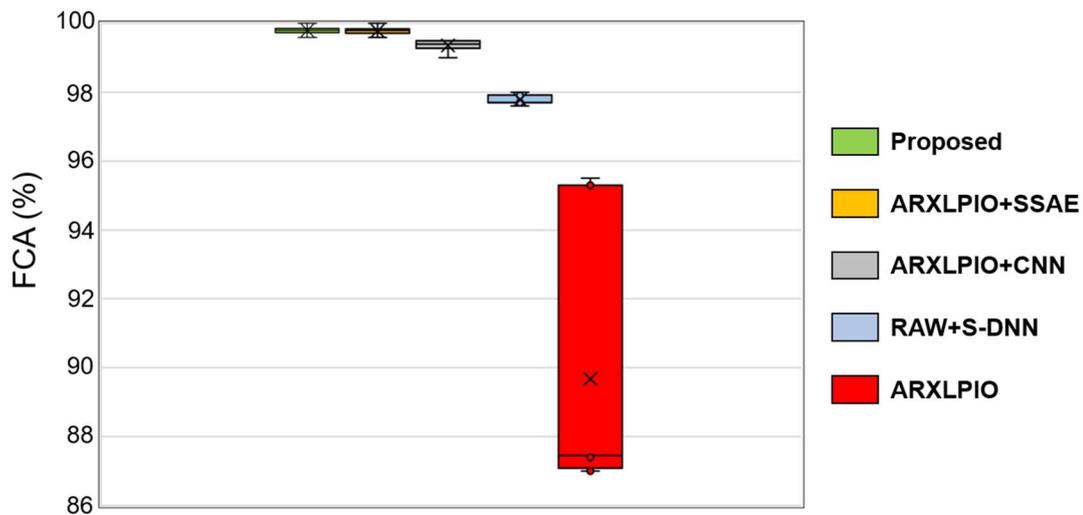


Figure 6.6. The boxplots demonstrating the statistics of the FCA metric over 10 experiments.

As can be seen from the boxplots in Figure 6.6, the FCA of the proposed method did not deviate significantly from the mean and its median values during the experiment and demonstrated the smallest Std value equal to 0.11% which shows that the proposed technique is not sensitive to training-testing data permutations and ensures the repeatability of the results. For ARXLPIO+SSAE it can be seen that the deviation of FCA also not significant and equal to one of the proposed approach. Regarding ARXLPIO+CNN, despite the relatively low Std of FCA, its average and median FCA values are located lower than those of the proposed and ARXLPIO+SSAE techniques. RAW+S-DNN approach demonstrates lower FCA values in all 10 experiments comparing to the first three approaches, whereas the Std of its FCAs is not high and just 0.13%. However, it is seen that the use of ARXLPIO instead of raw vibration signals in conjunction with S-DNN is capable of achieving higher results. Regarding the single application

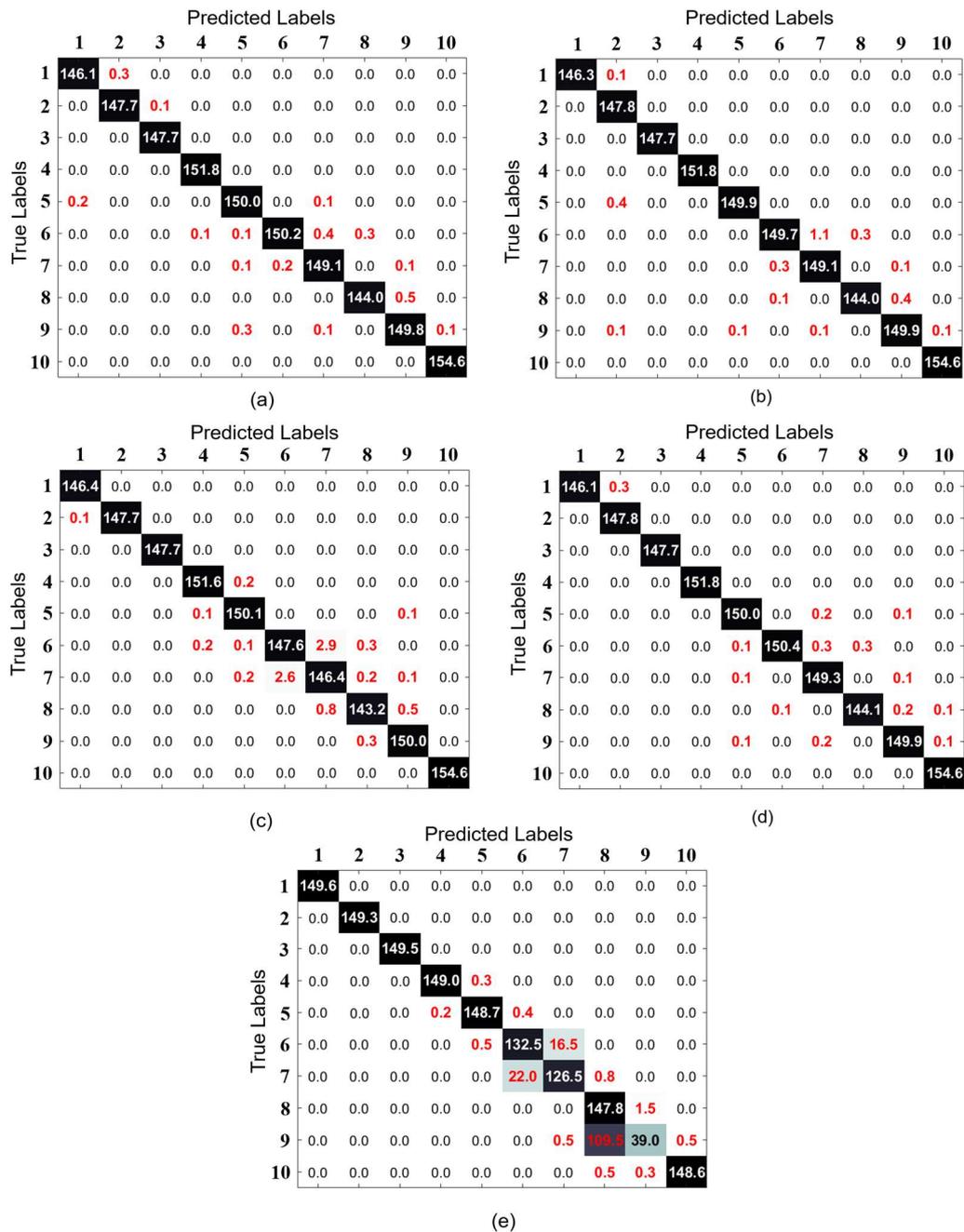


Figure 6.7. The confusion matrices obtained for (a) proposed, (b) ARXLPIO+SSAE, (c) ARXLPIO+CNN, (d) RAW+S-DNN, and (e) ARXLPIO averaged over 10 experiments.

of ARXLPIO, it is observed that its average FCA is the smallest one among the techniques used for the comparison and the Std value is equal to 3.7% which is the worst result.

The averaged confusion matrices depicted in Figure 6.7 show more detailed information about how different blade rubbing fault severity levels have been differentiated by the methods used in this comparison. From this figure, it can be observed that all of the approaches successfully differentiated signal conditions when no rub-impact faults were present in the system and the group

corresponding to signals collected under severe rubbing conditions. However, when the nonlinearity of the signal increases (i.e., signal classes from #5 to #9), some misclassifications between the data instance groups appear, especially it can be seen when utilizing the ARXLPIO and RAW+S-DNN methods.

When the rub-impact fault appears in the system and its intensity increases, the signal becomes highly nonlinear. Despite the simplicity of implementing a PIO in industrial applications, this type of observer is linear, which means that it cannot accurately estimate the uncertainty term of nonlinear signals collected from nonstationary rotor systems. Thus, according to Table 6.4 and the confusion matrix demonstrated in Figure 6.7 (e), a drop in the classification accuracy, as well as high rates of misclassification, can be observed when the intensity of rubbing in the system increases. According to the residual signals computed using (6.17) and demonstrated in Figure 6.8, it can be seen that the linear ARXLPIO has problems estimating the signals under slight and intensive rubbing conditions. From this figure, it is observed that the residual signals used for differentiating different signal groups (as in [21]) are well differentiable when no rubbing is present in the system, which means that they can be used for fault detection (i.e., differentiation between faulty and not faulty conditions); however, the estimation error increases as the rubbing intensity increases and the obtained residual signals for slight and intensive rubbing conditions are just slightly different from each other. This causes difficulties for the direct application of the linear ARXLPIO for blade rub-impact fault identification. To increase the capabilities of the linear observer when estimating nonlinear systems, robust nonlinear observation techniques, or hybrid approaches, such as the one presented in this chapter, should be considered.

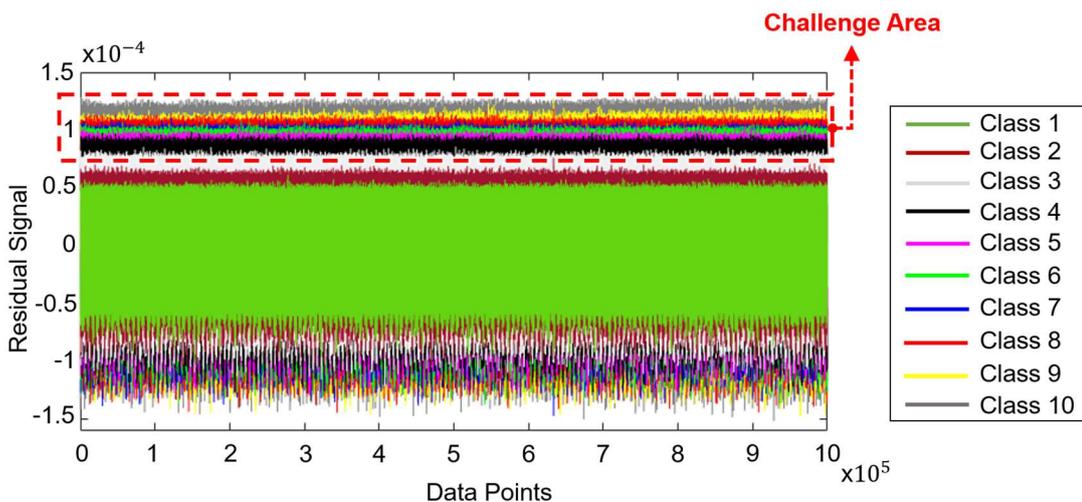


Figure 6.8. The error signals obtained by the ARXLPIO after signal estimation process.

Regarding the RAW+S-DNN, in this method, the frequency spectrum representation of the original vibration signal is used as the input for the same S-DNN that was utilized in the proposed hybrid technique. With the help of the S-DNN, the fault classification accuracy has been increased significantly in comparison with the ARXLPIO; however, because the raw signal is used in this method, some misclassifications appear when the complexity of the vibration signal increases with increasing rubbing fault intensity. This behavior is caused by some similarities and small variations in the frequency power spectra computed for those signal groups; therefore, the S-DNN might have issues when identifying patterns in specific areas of the frequency spectra of these signals.

Talking about ARXLPIO+SSAE and ARXLPIO+CNN methods from the second experimental scenario, it can be seen that these methods accomplished the fault diagnosis task with high values of performance metrics that are similar or slightly worse than ones demonstrated by the proposed approach. On the one hand, this observation tells us about the high capabilities of improving the accuracy of fault classification by deep learning-based techniques when they are combined with simple control theory-based linear observation techniques, such as ARXLPIO. Though the linear observer cannot accurately estimate the nonlinear signal that inevitably leads to degradation of fault diagnosis performance in general cases when the conventional control theory approaches (for instance, threshold-based methods) are used for classification, the deep learning methods can improve the fault diagnosis performance significantly when utilizing the output of the observer. On another hand, these approaches used for the comparison have some challenges and drawbacks as well. First, there is a problem of result reasoning in the ARXLPIO+CNN method: unlike the application of CNNs to two-dimensional image patterns where it is possible to evaluate what the exact local features have been learned by the network, it is difficult to interpret the local features extracted by CNN from one-dimensional signal patterns. Second, for the ARXLPIO+SSAE framework there exist a challenge of designing the appropriate architecture. Unlike the proposed S-DNN where the numbers of layers, nodes, and the best suitable architecture have been determined using the algorithm proposed in Section 6.2, for SSAE the numbers of stacked autoencoders (i.e., numbers of hidden pretrained layers), numbers of nodes in them, sparsity parameter should be found experimentally. Moreover, the learning process of stacked autoencoders (pretraining of hidden layers and tuning the final architecture) might be time-consuming that can be a sensitive issue for some applications.

Finally, the proposed method demonstrates high values of FCA and of performance metrics that are better than ones achieved by most of the counterpart methods used in the comparison, except ARXLPIO+SSAE. In our introduced solution, the linear ARXLPIO technique is improved by the automatically designed proposed S-DNN (i.e., the deep learning technique);

thus, the hybrid method for diagnosing blade rub-impact faults is introduced. The advantage of applying S-DNN to the output of ARXLPIO is the scalability of the network that using the steps proposed in Section 6.2 adapts to the dimensions of the input signals and thus finding the reasonably good architecture that solves the current fault classification problem well. Also, the increase of the FCA in comparison with the RAW+S-DNN method where no observer was used can be explained by the fact that linear model-reference techniques, such as the ARXLPIO method, in general, tend to be more robust to various disturbances than signal-based solutions because they are based on the rotor system itself and the behavior of this system. This is the reason why the proposed approach where the ARXLPIO first derived the nonlinear signal estimation and then the S-DNN was applied to improve the fault diagnosis capabilities of ARXLPIO demonstrated better accuracy than the method where S-DNN was directly applied to the spectra of raw resampled signals. As can be seen from Figure 6.7 (a) and (e), this combination also leads to the reduction of a number of misclassified samples corresponding to slight and intensive rubbing conditions in comparison with the case when no deep learning approach is used to improve the performance of ARXLPIO.

Overall, based on the experimental results, it can be concluded that the proposed hybrid deep learning-based observation technique can be successfully applied to identify rubbing faults of various intensity levels. Furthermore, the simplicity of creating the ARXLPIO in real scenarios, the autonomous design of S-DNN architecture according to the proposed algorithm and speed of its convergence along with the absence of TFA methods that are slow in terms of computation times (i.e., it might be difficult to use them in real-time or near-real-time fault diagnosis) makes the proposed method favorable for industrial applications.

6.4. Conclusions

In this chapter, a novel, hybrid deep learning-based observation technique is presented for diagnosing rub-impact faults of various intensity levels. In the proposed model, system modeling and observation is first accomplished by the linear ARXLPIO. Next, the blade rubbing signal estimated by the ARXLPIO is resampled with a 15% overlap based on the rotating speed and the time needed to perform one shaft revolution. Finally, the frequency power spectra of these resampled estimated signals are computed and used as inputs for the S-DNN designed according to the proposed algorithm. The advantage of the introduced scalable network is the presence of scalable layers that can be adjusted based on the input signal dimensionality and simplify the process of developing the DNN architecture for a specific task. In this work, the S-DNN is used to increase the performance of the linear ARXLPIO (when it is used for estimating nonlinear signals) and to perform a rubbing fault identification procedure. The proposed hybrid approach

outperformed most of the counterpart techniques used for the comparison. Moreover, the simplicity and flexibility of designing ARXLPIO along with the autonomous design of S-DNN allow the proposed framework to be applied in industrial environments.

Part III

Summary and Future Work

Chapter 7

Summary of Contributions and Future Work

7.1 Introduction

The main contributions of this dissertation and future aspects of the current work are given in this chapter. Section 7.2 highlights the main contributions of this thesis whereas future research direction is given in Section 7.3.

7.2 Summary of Contributions

This dissertation has focused on the condition-based maintenance of rotating machinery. It investigates the two different types of fault diagnosis pipelines that can be used for diagnosing blade rub-impact faults in turbomachinery, i.e., turbines. Specifically, in this thesis, the fault diagnosis of rotating machinery is performed by utilizing the general fault diagnosis scheme that consists of advanced signal processing and hybrid feature extraction with machine learning-based decision making and by creating end-to-end data-driven artificial intelligence-based and hybrid solutions. Thus, Chapters 2 to 3 are about the schemes and investigations into advanced iterative time-frequency signal processing methods and hybrid feature models for fault diagnosis of blade rubbing faults of various intensity levels, while Chapters 4, 5, and 6 focus on data-driven blade rub-impact fault identification by using different types of deep neural architectures and hybrid approaches. The contributions of this dissertation to condition monitoring and diagnosis of rotating machinery are given below:

- Advanced iterative time-frequency domain signal processing techniques such as empirical mode decomposition (EMD) are frequently utilized for analyzing complex vibration signals. The crucial drawback of the conventional EMD algorithm when applied to highly nonlinear and nonstationary signals is the mode-mixing problem. Furthermore, the iterative EMD-like time-frequency analysis approaches suffer the problem of fusing the informative intrinsic modes related to the mechanical faults being investigated. In Chapter 2 a novel algorithm for blade rub-impact fault identification is presented to tackle the problems discussed above. Firstly, the ensemble empirical mode decomposition (EEMD) algorithm replaced the original EMD to overcome the nonstationary and nonlinear behavior of the vibration signals containing the rubbing faults and deliver clearer intrinsic mode functions (IMFs). Next, the novel criterion for selecting optimal IMFs containing the frequency features of blade rub-impact faults is proposed. This criterion is comprised of a ratio between the degree-of-frequency-presence (DFP) of the specific frequency

harmonics related to rubbing faults and the Kullback-Leibler divergence (KLD). Hence, the modes for which the objective function value exceeded the arbitrarily assigned threshold were selected as the optimal components containing rub-impact fault information out of the complete finite set of the extracted intrinsic modes. The signals reconstructed using the selected optimal IMF components contained valuable information related to the rubbing process allow to distinguish the rub-impact faults of various intensity levels and contain less high-frequency noise making the feature models more accurate in describing these faults. These reconstructed signals were used for the hybrid feature model extraction and this hybrid feature model was used as the input to the support vector machines (SVM) algorithm to identify the faults of various intensity levels. The effectiveness of the proposed methodology was validated using the vibration acceleration dataset provided by UIAI Lab., where the rub-impact fault was simulated by creating the rotor imbalance that caused the interactions between rotor blades and rotor cage. Experimental results demonstrated that the proposed methodology allows for the identification of rubbing faults of various intensity levels with a high average classification accuracy equal to 99.2% over ten experimental trials even when the blade rub-fault is presented in the system along with another mechanical fault (i.e., rotor imbalance). It is worth noticing that the proposed approach outperformed the known feature models used for rubbing fault diagnosis.

- The increase of the rubbing fault intensity inevitably leads to the drastic increase of the energy of the IMF components extracted from the vibration signals and the amplitudes of blade rub- related frequency harmonics. In some circumstances, this phenomenon causes the selection of non-informative intrinsic modes or modes containing high-frequency noise along with the informative components when the objective function proposed in Chapter 2 is applied for fusing rub-related IMFs. This problem is mainly caused by the arbitrarily selected threshold introduced in Chapter 2 and by the increasing difference between the scales of objective values of IMFs corresponding to different rubbing intensity levels. To address this issue, in Chapter 3, the improvements to the IMF selection algorithm are introduced. First, the scaling procedure is proposed to normalize the objective values of the objective function computed for the IMFs extracted from the vibration signals. Next, the statistical properties of the set of normalized objective values are used to compute the adaptive threshold value. The improved selection algorithm of rub-impact fault-related IMFs selects the components for which the normalized objective value exceeds the adaptive threshold. The effectiveness of the proposed methodology was tested using the vibration acceleration dataset provided by UIAI Lab. where the rub-impact fault was simulated by creating the rotor imbalance that caused the interactions between rotor blades

and rotor cage. The experimental results demonstrate that the signals reconstructed using the IMF components selected by the improved procedure preserve the valuable blade rub-impact fault-related frequency harmonics but contain a significantly smaller amount of high-frequency noise in comparison with the signals reconstructed using the original IMF selection algorithm introduced in Chapter 2. Furthermore, the proposed improved method in combination with the hybrid feature model presented in Chapter 2 demonstrates slightly better rub-impact fault diagnosis capabilities in comparison with the original technique – 99.8% against 99.2% of average classification accuracy over ten experimental trials with different training and testing data configurations.

The second part of the dissertation contributes to the literature on deep learning-based and hybrid fault diagnosis of rotating machines as below:

- Despite the effectiveness of fault diagnosis techniques that are based on “advanced signal processing-feature extraction-decision making” pipelines, those techniques, in general, require a deep knowledge about the mechanical faults being investigated as well as the quality of extracted features. Usually, as features that are used to characterize mechanical faults based on signals, statistical or handcrafted parameters are utilized. However, since the statistical parameters are extracted from the signals collected from the rotor system, their values are affected by various disturbances such as environmental noise. Furthermore, without a hard signal preprocessing routine, the statistical feature parameters are likely to fail to characterize the mechanical faults hidden in nonstationary signals well. Regarding handcrafted feature parameters, it is not easy to verify whether these features are optimal ones for diagnosing the specific types of faults. The system-based fault diagnosis can be used to address the issues above when dealing with nonlinear rotor systems that produce nonstationary vibration signals. Chapter 4 introduces the data-driven system-based fault diagnosis framework that consists of a deep undercomplete denoising autoencoder (DUDAE) to perform artificial intelligence-based nonlinear system identification and deep neural network (DNN) for decision making about the current state of the system. In this approach, first, the resampled vibration signal collected under the normal operating condition of the system is used as input to DUDAE to learn how to perform nonlinear function approximation of the system under normal operating conditions. Then, this trained algorithm is used to estimate the current state of the rotor system by using the unknown vibration signal as the input. Then, the residual signals between the unknown vibration signal and its estimate by DUDAE are computed. Those residual signals appeared to be sensitive to the system degradation and were considered as discriminative features to be used by DNN for the decision making. The series of experiments show that

the proposed fault diagnosis model demonstrated stable convergence behavior under different training-testing data permutations and achieved average accuracy values of 95.64% and 95.11% over ten experimental trials with different techniques used for noise corruption at the input of DUDAЕ. The small value of fault classification accuracy standard deviation indicates the capabilities of the proposed approach to deliver stable results under different training and testing subsets configurations.

- To overcome the problem of multivariate signal analysis and to increase generalization capabilities by utilizing signals collected from several sensors, the multivariate one-dimensional convolutional neural network (ModCNN) is introduced in Chapter 5. Here, the multivariate vibration signal collected by multiple sensors is first cut into the windows with overlap considering the operating speed of rotary machinery. Since ModCNN belongs to the family of representation learning algorithms, the envelope power spectra of the obtained windows are computed to form discriminative patterns. The extraction of envelope spectra allowed for demodulating the multivariate vibration signal containing blade rub fault and for a discriminative representation of the changes in rub-impact fault intensity levels. Therefore, when these spectra were used as the inputs to ModCNN, it yielded satisfactory fault identification results. The experimental results demonstrated that the proposed deep learning-based fault diagnosis framework achieved a high fault classification accuracy of 99.75% averaged over ten experimental trials. The small value of fault classification accuracy standard deviation indicates the capabilities of the proposed approach to deliver stable results under different training and testing subsets configurations.
- For diagnosing complex engineering systems, many techniques from various fields have been proposed by researchers such as signal processing-, artificial intelligence-, and classical control theory-based methods. Each of those fields has its advantages that allow for demonstrating reasonably good results. However, due to the limitations in each of the fields, the fault diagnosis techniques that rely only on one of them suffer the problems of robustness when the uncertain conditions are presenting in the engineering systems as well as their stability degrades. To address these issues, a hybrid (i.e., the approach combining the techniques belonging to different fields) fault diagnosis approach using the deep learning-based observation technique is proposed in Chapter 6 of this dissertation. Here, the modeling of the system and its observation are first accomplished by the linear control theory-based approach called Autoregressive with eXogenous input Laguerre proportional-integral observer (ARXLPIO). Then, the signals estimated by ARXLPIO were cut with overlap into a set of windows and their frequency power spectra were

computed to derive informative data representation. These power spectra are then used as the inputs to a scalable deep neural network (S-DNN) with the architecture, autonomously designed by the novel proposed algorithm, for fault identification. In this chapter, S-DNN (deep learning-based approach) is used to improve the fault diagnosis performance of linear ARXLPIO (control theory-based approach) when it is applied to the nonlinear signals (i.e., blade rub impact fault signals) and to perform fault identification. The series of experiments demonstrated that the proposed hybrid fault diagnosis technique achieved a high fault classification accuracy of 99.79% averaged over ten experimental trials. The small value of fault classification accuracy standard deviation (0.11%) indicates that the proposed approach can deliver stable fault identification results under different training and testing subsets configurations. Moreover, the simplicity and flexibility of designing ARXLPIO along with the autonomous design of S-DNN architecture allow the proposed framework to be employed in industrial environments.

7.3 Future Work

As explained in the previous section, the primary focus of this dissertation is the condition-based maintenance of rotating machinery. Chapters 2 to 6 elaborated on the contributions and the investigations carried out on the topics. Though the research on these areas was a detailed one there are other topics and techniques which are needed to be investigated for exhaustive details. Some of these future topics are presented below:

- Hybrid fault detection, diagnosis, and prognosis: When the mechanical faults appear in different components of real industrial systems, their dynamic behavior changes drastically. The vibration signals collected from these types of systems appeared to be nonstationary ones due to the induced nonlinearity of systems. To resolve the problems of fault detection (i.e., discriminating the states of a system between healthy and unhealthy ones), fault diagnosis (i.e., classifying fault types within), and fault prognosis (i.e., estimating remaining useful lifetime, deriving health index sensitive to fault progression, and defining the end-of-life thresholds) in nonlinear industrial systems, various methods have been proposed by researchers. Those methods can be roughly divided into signal processing-, artificial intelligence-, and control theory-based approaches. Each type of them has its advantages as well as limitations which may cause problems of the robustness of each technique on its own in uncertain conditions. In my future work, I would like to continue researching and developing hybrid techniques by integrating methods of different fields into each other or combining them in various forms to simultaneously utilize their strong points for resolving complex problems and thus, enhancing their robustness and stability.

-
- **Multivariate signal analysis and data fusion:** The multivariate signal analysis and data fusion is a useful topic for industrial applications and a possible extension of the works presented in this dissertation. The usage of multivariate signals of the same or various types can provide in-depth details about the mechanical defects and faults ongoing in the components of industrial systems which can potentially benefit the processes of fault detection and identification. Furthermore, the appropriate strategies of data fusion from multiple sensors can be useful for accelerating the operation of condition monitoring frameworks by both picking the most valuable information related to the faults and reducing the total dimensionality of the data arrays needed to be processed. In my future research, I would like to investigate the ways of signal fusion and creating scalable and reliable end-to-end fault diagnosis systems that can be applied for analyzing health conditions of various systems.

Publications

International Journal Papers

1. Nguyen, C.D.; **Prosvirin, A.E.**; Kim, J.-M. Construction of a Sensitive and Speed Invariant Gearbox Fault Diagnosis Model Using an Incorporated Utilizing Adaptive Noise Control and a Stacked Sparse Autoencoder-based Deep Neural Network. *Sensors* **2021**, *21*, 18.
2. Ahmad, Z.; **Prosvirin, A.E.**; Kim, J.; Kim, J.-M. Multistage Centrifugal Pump Fault Diagnosis by Selecting Fault Characteristic Modes of Vibration and Pearson Linear Discriminant Analysis. *IEEE Access*, **2020**.
3. **Prosvirin, A.E.**; Piltan, F.; Kim, J.-M. Blade Rub-Impact Fault Identification Using Autoencoder-based Nonlinear Function Approximation and a Deep Neural Network. *Sensors*, **2020**, *20*, 6265.
4. **Prosvirin, A.E.**; Piltan, F.; Kim, J.-M. Hybrid Rub-impact Fault Diagnosis Using a Deep Learning-based Observation Technique. *IEEE Transactions on Neural Networks and Learning Systems*, **2020**.
5. Nguyen, C.D.; **Prosvirin, A.E.**; Kim, J.-M. A Reliable Fault Diagnosis Method for a Gearbox System with Varying Rotational Speeds. *Sensors* **2020**, *20*, 3105.
6. Toma, R. N.; **Prosvirin, A.E.**; Kim, J.-M. Bearing Fault Diagnosis of Induction Motors Using a Genetic Algorithm and Machine Learning Classifiers. *Sensors* **2020**, *20*, 1877.
7. Piltan, F.; **Prosvirin, A.E.**; Kim, J.-M. Robot Manipulator Active Fault-Tolerant Control Using a Machine Learning-based Automated Robust Hybrid Observer. *Journal of Intelligent & Fuzzy Systems*, **2020**.
8. Piltan, F.; **Prosvirin, A.E.**; Sohaib, M.; Saldivar, B.; Kim, J.-M. An SVM-Based Neural Adaptive Variable Structure Observer for Fault Diagnosis and Fault-Tolerant Control of a Robot Manipulator. *Applied Sciences*, **2020**, *10*, 1344.
9. Piltan, F.; **Prosvirin, A.E.**; Jeong, I.; Im, K.; Kim, J.-M. Rolling-Element Bearing Fault Diagnosis Using Advanced Machine Learning-Based Observer. *Applied Sciences*, **2019**, *9*, 5404.
10. **Prosvirin, A.E.**; Islam, M.M.M.; Kim, J.-M. An Improved Algorithm for Selecting IMF Components in Ensemble Empirical Mode Decomposition for Domain of Rub-Impact Fault Diagnosis. *IEEE Access*, **2019**.
11. **Prosvirin, A.E.**; Islam, M.; Kim, J.; Kim, J.-M. Rub-Impact Fault Diagnosis Using an Effective IMF Selection Technique in Ensemble Empirical Mode Decomposition and Hybrid Feature Models. *Sensors*, **2018**, *18*, 2040.

12. Appana, D.K.; **Prosvirin, A.E.**; Kim, J.-M. Reliable fault diagnosis of bearings with varying rotational speeds using envelope spectrum and convolution neural networks. *Soft Computing*, **2018**.

Journal Papers Under-review

1. **Prosvirin, A.E.**; Maliuk, A.S.; Kim, J.-M. Intelligent Rubbing Fault Identification Using Multivariate Signals and a Multivariate One-Dimensional Convolutional Neural Network (*Expert Systems with Applications*).
2. Islam, M.M.M.; **Prosvirin, A.E.**; Kim, J.-M. Data-Driven Prognostic Scheme for Rolling-Element Bearings Using a New Health Index and Variants of Least-Square Support Vector Machines. (*Mechanical Systems and Signal Processing*).
3. **Prosvirin, A.E.**; Ahmad, Z.; Kim, J.-M. Global and Local Feature Extraction Using AI-based Solutions for Diagnosing Centrifugal Pump Mechanical Faults. (*IEEE Access*).

Book Chapters

1. **Prosvirin, A.E.**; Duong, B.P.; Kim, J.-M. SVM Hyperparameter Optimization Using a Genetic Algorithm for Rub-Impact Fault Diagnosis. In *Advances in Intelligent Systems and Computing*; Bhatia S., Tiwari S., Mishra K., Trivedi M., Eds.; Springer Singapore: Singapore, **2019**, Vol. 924.
2. **Prosvirin, A.E.**; Kim, J.; Kim, J.-M. Efficient Rub-Impact Fault Diagnosis Scheme Based on Hybrid Feature Extraction and SVM. In *Advances in Computer Communication and Computational Sciences*; Bhatia, S.K., Tiwari, S., Mishra, K.K., Trivedi, M.C., Eds.; Springer Singapore: Singapore, **2019**; Vol. 759, pp. 405–415 (**Best Research Award**).
3. Islam, M.M.M.; **Prosvirin A.E.**; Kim, J.-M.; Intelligent Rub-Impact Fault Diagnosis Based on Genetic Algorithm-Based IMF Selection in Ensemble Empirical Mode Decomposition and Diverse Features Models. In *Lecture Notes in Computer Science*; Yin H., Camacho D., Novais P., Tallón-Ballesteros A., Eds.; Springer Cham: Cham, **2018**; Vol. 11314.
4. **Prosvirin, A.E.**; Kim, J.; Kim, J.-M. Bearing Fault Diagnosis Based on Convolutional Neural Networks with Kurtogram Representation of Acoustic Emission Signals. In *Advances in Computer Science and Ubiquitous Computing*; Park, J.J., Loia, V., Yi, G., Sung, Y., Eds.; Springer Singapore: Singapore, **2018**; Vol. 474, pp. 21–26

International Conferences

1. **Prosvirin, A.E.**; Kim, J.; Kim, J.-M.; “Tooth Cut Fault Identification in Fixed-Axis Gearbox Using Acoustic Emission Hit-based Features and Decision Trees.” *World Congress on Condition Monitoring*, Singapore, Singapore, **2019**.

2. Khan, S.A.; **Prosvirin A.E.**; Kim, J.-M.; “Towards Bearing Health Prognosis Using Generative Adversarial Networks: Modeling Bearing Degradation.” *International Conference on Advancements in Computational Sciences*, Lahore, Pakistan, **2018**.
3. **Prosvirin, A.E.**; Duong, B.P.; Kim, J.-M.; “SVM Hyperparameter Optimization Using a Genetic Algorithm for Rub-Impact Fault Diagnosis.” *The International Conference on Computer, Communication, and Computational Sciences*, Bangkok, Thailand, **2018**.
4. Islam, M.M.M.; **Prosvirin A.E.**; Kim, J.-M.; “Intelligent Rub-Impact Fault Diagnosis Based on Genetic Algorithm-Based IMF Selection in Ensemble Empirical Mode Decomposition and Diverse Features Models.” *Intelligent Data Engineering and Automated Learning*, Madrid, Spain, **2018**.
5. **Prosvirin, A.E.**; Kim, J.; Kim, J.-M. “Efficient Rub-Impact Fault Diagnosis Scheme Based on Hybrid Feature Extraction and SVM.” *The International Conference on Computational Sciences, Advanced Database, and Computing*, Phuket, Thailand, **2017**. (**Best Research Award**).
6. **Prosvirin, A.E.**; Kim, J.; Kim, J.-M. “Bearing Fault Diagnosis Based on Convolutional Neural Networks with Kurtogram Representation of Acoustic Emission Signals.” *The 11th International Conference on Multimedia and Ubiquitous Engineering*, Seoul, South Korea, **2017**.

Domestic Conferences

1. **Prosvirin, A.E.**; Kim, C.; Kim, J.-M. “Gearbox Fault Diagnosis Using Adaptive Thresholding-based Acoustic Emission Hit Feature Extraction and Machine Learning.” *The Korean Engineering Safety Health Art Society*, University of Ulsan, South Korea, 2019. (**Best Paper Award**).
2. Maliuk, A.S.; **Prosvirin, A.E.**; Kim, J.-M. “Induction Motor Stator Fault Detection Using Motor Current Signature Analysis Park Vector Approach and Machine Learning.” *The Korean Engineering Safety Health Art Society*, University of Ulsan, South Korea, 2019.
3. Islam, M.M.M.; **Prosvirin, A.E.**; Kim, J.-M. ”Robust Condition Monitoring of Rub-Impact Faults in Rotor System Using Empirical Mode Decomposition with IMF Selection.” *The Engineering and Art Society in Korea*, University of Ulsan, South Korea, **2017**.
4. **Prosvirin, A.E.**; Islam, M.M.M.; Kim, C.; Kim, J.-M. “Fault Prediction of Rolling Element Bearings Using One Class Least Squares SVM.” *The Engineering and Art Society in Korea*, University of Ulsan, South Korea, **2017**. (**Best Paper Award**).

References

- [1] R. B. Randall and J. Antoni, “Rolling element bearing diagnostics—A tutorial,” *Mechanical Systems and Signal Processing*, vol. 25, no. 2, pp. 485–520, Feb. 2011, doi: 10.1016/j.ymssp.2010.07.017.
- [2] Y. Qu, D. He, J. Yoon, B. Van Hecke, E. Bechhoefer, and J. Zhu, “Gearbox Tooth Cut Fault Diagnostics Using Acoustic Emission and Vibration Sensors — A Comparative Study,” *Sensors*, vol. 14, no. 1, pp. 1372–1393, Jan. 2014, doi: 10.3390/s140101372.
- [3] A. M. Umbrajkaar, A. Krishnamoorthy, and R. B. Dhumale, “Vibration Analysis of Shaft Misalignment Using Machine Learning Approach under Variable Load Conditions,” *Shock and Vibration*, vol. 2020, pp. 1–12, Jul. 2020, doi: 10.1155/2020/1650270.
- [4] A. K. Nandi and H. Ahmed, *Condition monitoring with vibration signals: compressive sampling and learning algorithms for rotating machines*. Hoboken, NJ, USA: Wiley-IEEE Press, 2019.
- [5] H. Henao *et al.*, “Trends in Fault Diagnosis for Electrical Machines: A Review of Diagnostic Techniques,” *IEEE Industrial Electronics Magazine*, vol. 8, no. 2, pp. 31–42, Jun. 2014, doi: 10.1109/MIE.2013.2287651.
- [6] H. Vogelesang, “An introduction to energy consumption in pumps,” *World Pumps*, vol. 2008, no. 496, pp. 28–31, Jan. 2008, doi: 10.1016/S0262-1762(07)70434-0.
- [7] X. Q. Fu, W. T. Jia, H. Xu, and S. L. Song, “Imbalance–misalignment–rubbing coupling faults in hydraulic turbine vibration,” *Optik*, vol. 127, no. 8, pp. 3708–3712, Apr. 2016, doi: 10.1016/j.ijleo.2016.01.006.
- [8] A. M. Abdelrhman, L. M. Hee, M. S. Leong, and S. Al-Obaidi, “Condition Monitoring of Blade in Turbomachinery: A Review,” *Advances in Mechanical Engineering*, vol. 6, p. 210717, Jan. 2014, doi: 10.1155/2014/210717.
- [9] W. K. Ngui, M. S. Leong, M. I. Shapiai, and M. H. Lim, “Blade Fault Diagnosis using Artificial Neural Network,” vol. 12, no. 4, p. 8, 2017.
- [10] Mohamed. A. Abuzaid, M. E. Eleshaky, and M. G. Zedan, “Effect of partial rotor-to-stator rub on shaft vibration,” *Journal of Mechanical Science and Technology*, vol. 23, no. 1, pp. 170–182, Jan. 2009, doi: 10.1007/s12206-008-0717-x.
- [11] P. Goldman and A. Muszynska, “Rotor-to-stator, rub-related, thermal/mechanical effects in rotating machinery,” *Chaos, Solitons & Fractals*, vol. 5, no. 9, pp. 1579–1601, Sep. 1995, doi: 10.1016/0960-0779(94)00165-M.
- [12] Z. Feng, M. Liang, and F. Chu, “Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples,” *Mechanical Systems and Signal Processing*, vol. 38, no. 1, pp. 165–205, Jul. 2013, doi: 10.1016/j.ymssp.2013.01.017.

-
- [13] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, Aug. 2018, doi: 10.1016/j.ymssp.2018.02.016.
- [14] A. E. Prosvirin, F. Piltan, and J.-M. Kim, "Hybrid Rubbing Fault Identification Using a Deep Learning-Based Observation Technique," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–12, 2020, doi: 10.1109/TNNLS.2020.3027160.
- [15] G. H. Farrahi, M. Tirehdast, E. Masoumi Khalil Abad, S. Parsa, and M. Motakefpoor, "Failure analysis of a gas turbine compressor," *Engineering Failure Analysis*, vol. 18, no. 1, pp. 474–484, Jan. 2011, doi: 10.1016/j.engfailanal.2010.09.042.
- [16] C. Vilcu, F. Niculescu, A. Mitru, C. Nechifor, C.-I. Borzea, and C. Cornea, "HolderCPS' - A new type data recorder system for proactive maintenance to rotary blade machines," in *2017 10th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*, Bucharest, 2017, pp. 655–660, doi: 10.1109/ATEE.2017.7905190.
- [17] S. M. A. Al-Obaidi, M. S. Leong, R. I. R. Hamzah, and A. M. Abdelrhman, "A Review of Acoustic Emission Technique for Machinery Condition Monitoring: Defects Detection & Diagnostic," *Applied Mechanics and Materials*, vol. 229–231, pp. 1476–1480, Nov. 2012, doi: 10.4028/www.scientific.net/AMM.229-231.1476.
- [18] K. Mathioudakis, A. Papathanasiou, E. Loukis, and K. Papailiou, "Fast Response Wall Pressure Measurement as a Means of Gas Turbine Blade Fault Identification," *Journal of Engineering for Gas Turbines and Power*, vol. 113, no. 2, pp. 269–275, Apr. 1991, doi: 10.1115/1.2906558.
- [19] K. M. Kim, J. S. Park, D. H. Lee, T. W. Lee, and H. H. Cho, "Analysis of conjugated heat transfer, stress and failure in a gas turbine blade with circular cooling passages," *Engineering Failure Analysis*, vol. 18, no. 4, pp. 1212–1222, Jun. 2011, doi: 10.1016/j.engfailanal.2011.03.002.
- [20] E. Rubio and J. C. Jáuregui, "15 Time-Frequency Analysis for Rotor-Rubbing Diagnosis," *Advances in Vibration Analysis Research*, p. 22.
- [21] F. Al-Badour, M. Sunar, and L. Cheded, "Vibration analysis of rotating machinery using time–frequency analysis and wavelet techniques," *Mechanical Systems and Signal Processing*, vol. 25, no. 6, pp. 2083–2101, Aug. 2011, doi: 10.1016/j.ymssp.2011.01.017.
- [22] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 1998, vol. 454, pp. 903–995, Accessed: Jun. 20, 2017. [Online]. Available: <http://rspa.royalsocietypublishing.org/content/royprsa/454/1971/903.full.pdf>.
- [23] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in adaptive data analysis*, vol. 1, no. 01, pp. 1–41, 2009.

-
- [24] Y. Yang, J. Cheng, and K. Zhang, “An ensemble local means decomposition method and its application to local rub-impact fault diagnosis of the rotor systems,” *Measurement*, vol. 45, no. 3, pp. 561–570, Apr. 2012, doi: 10.1016/j.measurement.2011.10.010.
- [25] Y. Wang, R. Markert, J. Xiang, and W. Zheng, “Research on variational mode decomposition and its application in detecting rub-impact fault of the rotor system,” *Mechanical Systems and Signal Processing*, vol. 60–61, pp. 243–251, Aug. 2015, doi: 10.1016/j.ymssp.2015.02.020.
- [26] L. Deng and R. Zhao, “Fault feature extraction of a rotor system based on local mean decomposition and Teager energy kurtosis,” *Journal of Mechanical Science and Technology*, vol. 28, no. 4, pp. 1161–1169, Apr. 2014, doi: 10.1007/s12206-013-1149-9.
- [27] Q. Gao, C. Duan, H. Fan, and Q. Meng, “Rotating machine fault diagnosis using empirical mode decomposition,” *Mechanical Systems and Signal Processing*, vol. 22, no. 5, pp. 1072–1081, Jul. 2008, doi: 10.1016/j.ymssp.2007.10.003.
- [28] Z. K. Peng, P. W. Tse, and F. L. Chu, “A comparison study of improved Hilbert–Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing,” *Mechanical Systems and Signal Processing*, vol. 19, no. 5, pp. 974–988, Sep. 2005, doi: 10.1016/j.ymssp.2004.01.006.
- [29] M. Hasan and J.-M. Kim, “Fault Detection of a Spherical Tank Using a Genetic Algorithm-Based Hybrid Feature Pool and k-Nearest Neighbor Algorithm,” *Energies*, vol. 12, no. 6, p. 991, Mar. 2019, doi: 10.3390/en12060991.
- [30] M. M. M. Islam, J. Kim, S. A. Khan, and J.-M. Kim, “Reliable bearing fault diagnosis using Bayesian inference-based multi-class support vector machines,” *The Journal of the Acoustical Society of America*, vol. 141, no. 2, pp. EL89–EL95, Feb. 2017, doi: 10.1121/1.4976038.
- [31] B. Samanta and K. R. Al-Balushi, “ARTIFICIAL NEURAL NETWORK BASED FAULT DIAGNOSTICS OF ROLLING ELEMENT BEARINGS USING TIME-DOMAIN FEATURES,” *Mechanical Systems and Signal Processing*, vol. 17, no. 2, pp. 317–328, Mar. 2003, doi: 10.1006/mssp.2001.1462.
- [32] J. Yu, X. Yang, F. Gao, and D. Tao, “Deep multimodal distance metric learning using click constraints for image ranking,” *IEEE transactions on cybernetics*, vol. 47, no. 12, pp. 4014–4024, 2017.
- [33] M. Gharbi, J. Chen, J. T. Barron, S. W. Hasinoff, and F. Durand, “Deep bilateral learning for real-time image enhancement,” *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, p. 118, 2017.
- [34] Y. Zhang, M. Pezeshki, P. Brakel, S. Zhang, C. L. Y. Bengio, and A. Courville, “Towards end-to-end speech recognition with deep convolutional neural networks,” *arXiv preprint arXiv:1701.02720*, 2017.

-
- [35] G. Brunner, Y. Wang, R. Wattenhofer, and M. Weigelt, “Natural Language Multitasking: Analyzing and Improving Syntactic Saliency of Hidden Representations,” *arXiv preprint arXiv:1801.06024*, 2018.
- [36] S. Poornachandra and C. Naveena, “Pre-processing of MR Images for Efficient Quantitative Image Analysis using Deep Learning Techniques,” in *2017 International Conference on Recent Advances in Electronics and Communication Technology (ICRAECT)*, 2017, pp. 191–195.
- [37] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, “GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification,” *Neurocomputing*, vol. 321, pp. 321–331, Dec. 2018, doi: 10.1016/j.neucom.2018.09.013.
- [38] Q. Meng, D. Catchpole, D. Skillicom, and P. J. Kennedy, “Relational autoencoder for feature extraction,” in *2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK, USA, May 2017, pp. 364–371, doi: 10.1109/IJCNN.2017.7965877.
- [39] G. Ma, X. Yang, B. Zhang, and Z. Shi, “Multi-feature fusion deep networks,” *Neurocomputing*, vol. 218, pp. 164–171, Dec. 2016, doi: 10.1016/j.neucom.2016.08.059.
- [40] Y. Zhang, B. Wen, and A. Y. T. Leung, “Reliability Analysis for Rotor Rubbing,” *Journal of Vibration and Acoustics*, vol. 124, no. 1, p. 58, 2002, doi: 10.1115/1.1423635.
- [41] E. Rubio and J. C. Jáuregui, *Time-frequency analysis for rotor-rubbing diagnosis*. Citeseer, 2011.
- [42] F. Cong, J. Chen, G. Dong, and K. Huang, “Experimental validation of impact energy model for the rub–impact assessment in a rotor system,” *Mechanical Systems and Signal Processing*, vol. 25, no. 7, pp. 2549–2558, Oct. 2011, doi: 10.1016/j.ymsp.2011.04.004.
- [43] Y. Lu, F. Meng, and Y. Li, “Research on Rub Impact Fault Diagnosis Method of Rotating Machinery Based on Wavelet Packet and Support Vector Machine,” 2009, pp. 707–710, doi: 10.1109/ICMTMA.2009.539.
- [44] J. Zhihao, J. Shangwei, J. Wen, and W. Bangchun, “Rubbing Fault Diagnosis of Rotary Machinery Based on Wavelet and Support Vector Machine,” Apr. 2009, pp. 287–290, doi: 10.1109/DBTA.2009.163.
- [45] S. D. Roy, S. K. Shome, and S. K. Laha, “Impact of wavelets and filter on vibration-based mechanical rub detection using Neural Networks,” in *2014 Annual IEEE India Conference (INDICON)*, 2014, pp. 1–6, Accessed: Dec. 26, 2016. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7030446.
- [46] P. W. Tse, W. Yang, and H. Y. Tam, “Machine fault diagnosis through an effective exact wavelet analysis,” *Journal of Sound and Vibration*, vol. 277, no. 4–5, pp. 1005–1024, Nov. 2004, doi: 10.1016/j.jsv.2003.09.031.

-
- [47] N. Bessous, S. E. Zouzou, W. Bentrach, S. Sbaa, and M. Sahraoui, "Diagnosis of bearing defects in induction motors using discrete wavelet transform," *International Journal of System Assurance Engineering and Management*, vol. 9, no. 2, pp. 335–343, Apr. 2018, doi: 10.1007/s13198-016-0459-6.
- [48] B. Bessam, A. Menacer, M. Boumehraz, and H. Cherif, "Wavelet transform and neural network techniques for inter-turn short circuit diagnosis and location in induction motor," *International Journal of System Assurance Engineering and Management*, vol. 8, no. S1, pp. 478–488, Jan. 2017, doi: 10.1007/s13198-015-0400-4.
- [49] H. Niaoqing, C. Min, and W. Xisen, "THE APPLICATION OF STOCHASTIC RESONANCE THEORY FOR EARLY DETECTING RUB-IMPACT FAULT OF ROTOR SYSTEM," *Mechanical Systems and Signal Processing*, vol. 17, no. 4, pp. 883–895, Jul. 2003, doi: 10.1006/mssp.2002.1470.
- [50] S. Wang, X. Chen, G. Li, X. Li, and Z. He, "Matching Demodulation Transform With Application to Feature Extraction of Rotor Rub-Impact Fault," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 5, pp. 1372–1383, May 2014, doi: 10.1109/TIM.2013.2283552.
- [51] L. Yang, X. Chen, S. Wang, and H. Zuo, "Rub-Impact Detection of Rotor Systems Using Time-Frequency Techniques," in *ASME 2016 International Mechanical Engineering Congress and Exposition*, 2016, p. V04BT05A011-V04BT05A011.
- [52] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 35, no. 1–2, pp. 108–126, Feb. 2013, doi: 10.1016/j.ymssp.2012.09.015.
- [53] Y. Zhang, C. Zhang, J. Sun, and J. Guo, "Improved Wind Speed Prediction Using Empirical Mode Decomposition," *Advances in Electrical and Computer Engineering*, vol. 18, no. 2, pp. 3–10, 2018, doi: 10.4316/AECE.2018.02001.
- [54] J. Zhu, P. Sun, Y. Gao, and P. Zheng, "Clock Differences Prediction Algorithm Based on EMD-SVM," *Chinese Journal of Electronics*, vol. 27, no. 1, pp. 128–132, Jan. 2018, doi: 10.1049/cje.2016.08.039.
- [55] G. Zhao, Q. Liang, and T. S. Durrani, "An EMD Based Sense-Through-Foliage Target Detection UWB Radar Sensor Networks," *IEEE Access*, vol. 6, pp. 29254–29261, 2018, doi: 10.1109/ACCESS.2018.2841900.
- [56] Y. Zhao, E. Liu, J. Zhu, B. Zhang, J. Wang, and H. Tian, "Rub-impact fault diagnosis of rotating machinery based on Hilbert-Huang transform," in *2015 IEEE International Conference on Mechatronics and Automation (ICMA)*, Aug. 2015, pp. 32–36, doi: 10.1109/ICMA.2015.7237452.

-
- [57] Y. Yang, S. Zhang, Q. Han, and Y. Qu, "Vibration Fault analysis for rotor systems by using Hilbert spectrum," in *2010 International Conference on Mechanic Automation and Control Engineering*, Jun. 2010, pp. 2411–2414, doi: 10.1109/MACE.2010.5535813.
- [58] J. Cheng, D. Yu, J. Tang, and Y. Yang, "Local rub-impact fault diagnosis of the rotor systems based on EMD," *Mechanism and Machine Theory*, vol. 44, no. 4, pp. 784–791, Apr. 2009, doi: 10.1016/j.mechmachtheory.2008.04.006.
- [59] L. Xiang, G. Tang, and A. Hu, "Analysis of Rotor Rubbing Fault Signal Based on Hilbert-Huang Transform," 2009, pp. 586–589, doi: 10.1109/ICMTMA.2009.507.
- [60] L. Yibo, M. Fanlong, and L. Yanjun, "Research on rub impact fault diagnosis method of rotating machinery based on EMD and SVM," in *2009 International Conference on Mechatronics and Automation*, 2009, pp. 4806–4810, Accessed: Dec. 26, 2016. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5246424.
- [61] X. Qin, Q. Li, X. Dong, and S. Lv, "The Fault Diagnosis of Rolling Bearing Based on Ensemble Empirical Mode Decomposition and Random Forest," *Shock and Vibration*, vol. 2017, pp. 1–9, 2017, doi: 10.1155/2017/2623081.
- [62] K. Yu, T. R. Lin, and J. W. Tan, "A bearing fault diagnosis technique based on singular values of EEMD spatial condition matrix and Gath-Geva clustering," *Applied Acoustics*, vol. 121, pp. 33–45, Jun. 2017, doi: 10.1016/j.apacoust.2017.01.023.
- [63] Y. Lei, Z. He, and Y. Zi, "Application of the EEMD method to rotor fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 1327–1338, May 2009, doi: 10.1016/j.ymsp.2008.11.005.
- [64] Z. Li and B. Shi, "Research of Fault Diagnosis Based on Sensitive Intrinsic Mode Function Selection of EEMD and Adaptive Stochastic Resonance," *Shock and Vibration*, vol. 2016, pp. 1–12, 2016, doi: 10.1155/2016/2841249.
- [65] Y. Lei and M. J. Zuo, "Fault diagnosis of rotating machinery using an improved HHT based on EEMD and sensitive IMFs," *Measurement Science and Technology*, vol. 20, no. 12, p. 125701, Dec. 2009, doi: 10.1088/0957-0233/20/12/125701.
- [66] C. Yi, D. Wang, W. Fan, K.-L. Tsui, and J. Lin, "EEMD-Based Steady-State Indexes and Their Applications to Condition Monitoring and Fault Diagnosis of Railway Axle Bearings," *Sensors*, vol. 18, no. 3, p. 704, Feb. 2018, doi: 10.3390/s18030704.
- [67] Y. Liao, P. Sun, B. Wang, and L. Qu, "Extraction of repetitive transients with frequency domain multipoint kurtosis for bearing fault diagnosis," *Measurement Science and Technology*, vol. 29, no. 5, p. 055012, May 2018, doi: 10.1088/1361-6501/aaae99.
- [68] D. Wang, P. W. Tse, and K. L. Tsui, "An enhanced Kurtogram method for fault diagnosis of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 35, no. 1–2, pp. 176–199, Feb. 2013, doi: 10.1016/j.ymsp.2012.10.003.

-
- [69] M. Kang, J. Kim, J.-M. Kim, A. C. C. Tan, E. Y. Kim, and B.-K. Choi, "Reliable Fault Diagnosis for Low-Speed Bearings Using Individually Trained Support Vector Machines With Kernel Discriminative Feature Analysis," *IEEE Transactions on Power Electronics*, vol. 30, no. 5, pp. 2786–2797, May 2015, doi: 10.1109/TPEL.2014.2358494.
- [70] M. M. M. Islam, S. A. Khan, and J.-M. Kim, "Multi-fault Diagnosis of Roller Bearings Using Support Vector Machines with an Improved Decision Strategy," in *Advanced Intelligent Computing Theories and Applications*, vol. 9227, D.-S. Huang and K. Han, Eds. Cham: Springer International Publishing, 2015, pp. 538–550.
- [71] Z. Xia, S. Xia, L. Wan, and S. Cai, "Spectral Regression Based Fault Feature Extraction for Bearing Accelerometer Sensor Signals," *Sensors*, vol. 12, no. 10, pp. 13694–13719, Oct. 2012, doi: 10.3390/s121013694.
- [72] D. He, R. Li, and J. Zhu, "Plastic Bearing Fault Diagnosis Based on a Two-Step Data Mining Approach," *IEEE Transactions on Industrial Electronics*, pp. 3429–3440, 2012, doi: 10.1109/TIE.2012.2192894.
- [73] F. Chu and W. Lu, "Experimental observation of nonlinear vibrations in a rub-impact rotor system," *Journal of Sound and Vibration*, vol. 283, no. 3–5, pp. 621–643, May 2005, doi: 10.1016/j.jsv.2004.05.012.
- [74] S.-H. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," *City*, vol. 1, no. 2, p. 1, 2007.
- [75] T. Han, B.-S. Yang, W.-H. Choi, and J.-S. Kim, "Fault Diagnosis System of Induction Motors Based on Neural Network and Genetic Algorithm Using Stator Current Signals," *International Journal of Rotating Machinery*, vol. 2006, pp. 1–13, 2006, doi: 10.1155/IJRM/2006/61690.
- [76] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mechanical Systems and Signal Processing*, vol. 104, pp. 799–834, May 2018, doi: 10.1016/j.ymsp.2017.11.016.
- [77] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, Jul. 2009, doi: 10.1016/j.ipm.2009.03.002.
- [78] G. E. Hinton and S. T. Roweis, "Stochastic Neighbor Embedding," p. 8.
- [79] L. van der Maaten and G. Hinton, "Visualizing Data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579–2605, 2008.
- [80] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, no. 1, pp. 1–37, Jan. 2008, doi: 10.1007/s10115-007-0114-2.
- [81] B. Cai, Y. Zhao, H. Liu, and M. Xie, "A Data-Driven Fault Diagnosis Methodology in Three-Phase Inverters for PMSM Drive Systems," *IEEE Transactions on Power Electronics*, vol. 32, no. 7, pp. 5590–5600, Jul. 2017, doi: 10.1109/TPEL.2016.2608842.

-
- [82] B. Cai, H. Liu, and M. Xie, "A real-time fault diagnosis methodology of complex systems using object-oriented Bayesian networks," *Mechanical Systems and Signal Processing*, vol. 80, pp. 31–44, Dec. 2016, doi: 10.1016/j.ymsp.2016.04.019.
- [83] A. Prosvirin, J. Kim, and J.-M. Kim, "Bearing Fault Diagnosis Based on Convolutional Neural Networks with Kurtogram Representation of Acoustic Emission Signals," in *Advances in Computer Science and Ubiquitous Computing*, vol. 474, J. J. Park, V. Loia, G. Yi, and Y. Sung, Eds. Singapore: Springer Singapore, 2018, pp. 21–26.
- [84] F. Piltan and J.-M. Kim, "Bearing Fault Diagnosis by a Robust Higher-Order Super-Twisting Sliding Mode Observer," *Sensors*, vol. 18, no. 4, p. 1128, Apr. 2018, doi: 10.3390/s18041128.
- [85] Y. Lu, F. Meng, and Y. Li, "Research on Rub Impact Fault Diagnosis Method of Rotating Machinery Based on Wavelet Packet and Support Vector Machine," in *2009 International Conference on Measuring Technology and Mechatronics Automation*, Zhangjiajie, Hunan, China, 2009, pp. 707–710, doi: 10.1109/ICMTMA.2009.539.
- [86] S. D. Roy, S. K. Shome, and S. K. Laha, "Impact of wavelets and filter on vibration-based mechanical rub detection using Neural Networks," in *2014 Annual IEEE India Conference (INDICON)*, Pune, India, Dec. 2014, pp. 1–6, doi: 10.1109/INDICON.2014.7030446.
- [87] P. W. Tse, W. Yang, and H. Y. Tam, "Machine fault diagnosis through an effective exact wavelet analysis," *Journal of Sound and Vibration*, vol. 277, no. 4–5, pp. 1005–1024, Nov. 2004, doi: 10.1016/j.jsv.2003.09.031.
- [88] Li Yibo, Meng Fanlong, and Lu Yanjun, "Research on rub impact fault diagnosis method of rotating machinery based on EMD and SVM," in *2009 International Conference on Mechatronics and Automation*, Changchun, China, Aug. 2009, pp. 4806–4810, doi: 10.1109/ICMA.2009.5246424.
- [89] A. Prosvirin, J. Kim, and J.-M. Kim, "Efficient Rub-Impact Fault Diagnosis Scheme Based on Hybrid Feature Extraction and SVM," in *Advances in Computer Communication and Computational Sciences*, vol. 759, S. K. Bhatia, S. Tiwari, K. K. Mishra, and M. C. Trivedi, Eds. Singapore: Springer Singapore, 2019, pp. 405–415.
- [90] S.-H. Cha, "Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions," vol. 1, no. 4, p. 8, 2007.
- [91] A. Prosvirin, M. Islam, J. Kim, and J.-M. Kim, "Rub-Impact Fault Diagnosis Using an Effective IMF Selection Technique in Ensemble Empirical Mode Decomposition and Hybrid Feature Models," *Sensors*, vol. 18, no. 7, p. 2040, Jun. 2018, doi: 10.3390/s18072040.
- [92] A. E. Prosvirin, M. M. M. Islam, and J.-M. Kim, "An Improved Algorithm for Selecting IMF Components in Ensemble Empirical Mode Decomposition for Domain of Rub-Impact

- Fault Diagnosis,” *IEEE Access*, vol. 7, pp. 121728–121741, 2019, doi: 10.1109/ACCESS.2019.2938367.
- [93] A. Komaty, A.-O. Boudraa, B. Augier, and D. Dare-Emzivat, “EMD-Based Filtering Using Similarity Measure Between Probability Density Functions of IMFs,” *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 1, pp. 27–34, Jan. 2014, doi: 10.1109/TIM.2013.2275243.
- [94] M. M. Suarez-Alvarez, D.-T. Pham, M. Y. Prostov, and Y. I. Prostov, “Statistical approach to normalization of feature vectors and clustering of mixed datasets,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 468, no. 2145, pp. 2630–2651, Sep. 2012, doi: 10.1098/rspa.2011.0704.
- [95] L. Zhan and C. Li, “A Comparative Study of Empirical Mode Decomposition-Based Filtering for Impact Signal,” *Entropy*, vol. 19, no. 1, p. 13, Dec. 2016, doi: 10.3390/e19010013.
- [96] L. Bai, Z. Han, Y. Li, and S. Ning, “A Hybrid De-Noising Algorithm for the Gear Transmission System Based on CEEMDAN-PE-TFPF,” *Entropy*, vol. 20, no. 5, p. 361, May 2018, doi: 10.3390/e20050361.
- [97] S. Madhavan, R. Jain, C. Sujatha, and A. S. Sekhar, “Vibration based damage detection of rotor blades in a gas turbine engine,” *Engineering Failure Analysis*, vol. 46, pp. 26–39, Nov. 2014, doi: 10.1016/j.engfailanal.2014.07.021.
- [98] M. Willsch, T. Bosselmann, and N. M. Theune, “New approaches for the monitoring of gas turbine blades and vanes,” in *Proceedings of IEEE Sensors, 2004.*, Vienna, Austria, 2004, pp. 20–23, doi: 10.1109/ICSENS.2004.1426089.
- [99] F. K. Choy and J. Padovan, “Non-linear transient analysis of rotor-casing rub events,” *Journal of Sound and Vibration*, vol. 113, no. 3, pp. 529–545, Mar. 1987, doi: 10.1016/S0022-460X(87)80135-9.
- [100] F. Chu and W. Lu, “Experimental observation of nonlinear vibrations in a rub-impact rotor system,” *Journal of Sound and Vibration*, vol. 283, no. 3–5, pp. 621–643, May 2005, doi: 10.1016/j.jsv.2004.05.012.
- [101] Y. Lei, J. Lin, Z. He, and M. J. Zuo, “A review on empirical mode decomposition in fault diagnosis of rotating machinery,” *Mechanical Systems and Signal Processing*, vol. 35, no. 1–2, pp. 108–126, Feb. 2013, doi: 10.1016/j.ymsp.2012.09.015.
- [102] Z. Huo, Y. Zhang, L. Shu, and M. Gallimore, “A New Bearing Fault Diagnosis Method Based on Fine-to-Coarse Multiscale Permutation Entropy, Laplacian Score and SVM,” *IEEE Access*, vol. 7, pp. 17050–17066, 2019, doi: 10.1109/ACCESS.2019.2893497.
- [103] F. Piltan, A. E. Prosvirin, I. Jeong, K. Im, and J.-M. Kim, “Rolling-Element Bearing Fault Diagnosis Using Advanced Machine Learning-Based Observer,” *Applied Sciences*, vol. 9, no. 24, p. 5404, Dec. 2019, doi: 10.3390/app9245404.

-
- [104] Y. Lu and Y. Liu, "Recognition of rotor rubbing fault types based on BP neural networks," in *The 27th Chinese Control and Decision Conference (2015 CCDC)*, Qingdao, China, May 2015, pp. 4614–4618, doi: 10.1109/CCDC.2015.7162739.
- [105] Z. Gao, T. Breikin, and H. Wang, "Discrete-time proportional and integral observer and observer-based controller for systems with both unknown input and output disturbances," *Optimal Control Applications and Methods*, vol. 29, no. 3, pp. 171–189, May 2008, doi: 10.1002/oca.819.
- [106] D. W. C. Ho and Z. Gao, "Proportional multiple-integral observer design for descriptor systems with measurement output disturbances," *IEE Proceedings - Control Theory and Applications*, vol. 151, no. 3, pp. 279–288, May 2004, doi: 10.1049/ip-cta:20040437.
- [107] C. B. Njima and T. Garna, "PIO Output Fault Diagnosis by ARX-Laguerre Model Applied to 2nd Order Electrical System," *IEEE Access*, vol. 8, pp. 83052–83061, 2020, doi: 10.1109/ACCESS.2020.2990696.
- [108] F. Piltan, A. E. Prosvirin, M. Sohaib, B. Saldivar, and J.-M. Kim, "An SVM-Based Neural Adaptive Variable Structure Observer for Fault Diagnosis and Fault-Tolerant Control of a Robot Manipulator," *Applied Sciences*, vol. 10, no. 4, p. 1344, Feb. 2020, doi: 10.3390/app10041344.
- [109] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. N. L. Benders, and I. Isgum, "Automatic Segmentation of MR Brain Images With a Convolutional Neural Network," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1252–1261, May 2016, doi: 10.1109/TMI.2016.2548501.
- [110] R. Liu, G. Meng, B. Yang, C. Sun, and X. Chen, "Dislocated Time Series Convolutional Neural Architecture: An Intelligent Fault Diagnosis Approach for Electric Machine," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1310–1320, Jun. 2017, doi: 10.1109/TII.2016.2645238.
- [111] D. K. Appana, A. Prosvirin, and J.-M. Kim, "Reliable fault diagnosis of bearings with varying rotational speeds using envelope spectrum and convolution neural networks," *Soft Computing*, vol. 22, no. 20, pp. 6719–6729, Oct. 2018, doi: 10.1007/s00500-018-3256-0.
- [112] X. Wu, Z. Peng, J. Ren, C. Cheng, W. Zhang, and D. Wang, "Rub-Impact Fault Diagnosis of Rotating Machinery Based on 1-D Convolutional Neural Networks," *IEEE Sensors Journal*, vol. 20, no. 15, pp. 8349–8363, Aug. 2020, doi: 10.1109/JSEN.2019.2944157.
- [113] Z. Chen and Z. Li, "Research on fault diagnosis method of rotating machinery based on deep learning," in *2017 Prognostics and System Health Management Conference (PHM-Harbin)*, Harbin, China, Jul. 2017, pp. 1–4, doi: 10.1109/PHM.2017.8079279.
- [114] S. A. Khan, A. E. Prosvirin, and J.-M. Kim, "Towards bearing health prognosis using generative adversarial networks: Modeling bearing degradation," in *2018 International*

- Conference on Advancements in Computational Sciences (ICACS)*, Lahore, Feb. 2018, pp. 1–6, doi: 10.1109/ICACS.2018.8333495.
- [115] Q. Liu, G. Ma, and C. Cheng, “Data Fusion Generative Adversarial Network for Multi-Class Imbalanced Fault Diagnosis of Rotating Machinery,” *IEEE Access*, vol. 8, pp. 70111–70124, 2020, doi: 10.1109/ACCESS.2020.2986356.
- [116] W. Mao, Y. Liu, L. Ding, and Y. Li, “Imbalanced Fault Diagnosis of Rolling Bearing Based on Generative Adversarial Network: A Comparative Study,” *IEEE Access*, vol. 7, pp. 9515–9530, 2019, doi: 10.1109/ACCESS.2018.2890693.
- [117] D. Charte, F. Charte, S. García, M. J. del Jesus, and F. Herrera, “A practical tutorial on autoencoders for nonlinear feature fusion: Taxonomy, models, software and guidelines,” *Information Fusion*, vol. 44, pp. 78–96, Nov. 2018, doi: 10.1016/j.inffus.2017.12.007.
- [118] G. Jiang, P. Xie, H. He, and J. Yan, “Wind Turbine Fault Detection Using a Denoising Autoencoder With Temporal Information,” *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 1, pp. 89–100, Feb. 2018, doi: 10.1109/TMECH.2017.2759301.
- [119] E. Principi, D. Rossetti, S. Squartini, and F. Piazza, “Unsupervised electric motor fault detection by using deep autoencoders,” *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 2, pp. 441–451, Mar. 2019, doi: 10.1109/JAS.2019.1911393.
- [120] G. Jiang, H. He, P. Xie, and Y. Tang, “Stacked Multilevel-Denoising Autoencoders: A New Representation Learning Approach for Wind Turbine Gearbox Fault Diagnosis,” *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 9, pp. 2391–2402, Sep. 2017, doi: 10.1109/TIM.2017.2698738.
- [121] Y. Bengio, A. Courville, and P. Vincent, “Representation Learning: A Review and New Perspectives,” *arXiv:1206.5538 [cs]*, Apr. 2014, Accessed: Jul. 16, 2020. [Online]. Available: <http://arxiv.org/abs/1206.5538>.
- [122] I. Guyon and A. Elisseeff, “An Introduction to Feature Extraction,” in *Feature Extraction*, vol. 207, I. Guyon, M. Nikravesh, S. Gunn, and L. A. Zadeh, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 1–25.
- [123] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*, vol. 72. Cham: Springer International Publishing, 2015.
- [124] M. B. Ali, “USE OF DROPOUTS AND SPARSITY FOR REGULARIZATION OF AUTOENCODERS IN DEEP NEURAL NETWORKS,” p. 99.
- [125] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, “Self-Normalizing Neural Networks,” *arXiv:1706.02515 [cs, stat]*, Sep. 2017, Accessed: Jul. 16, 2020. [Online]. Available: <http://arxiv.org/abs/1706.02515>.
- [126] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” p. 8.

-
- [127] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv:1412.6980 [cs]*, Jan. 2017, Accessed: Mar. 04, 2020. [Online]. Available: <http://arxiv.org/abs/1412.6980>.
- [128] J. Cheng, D. Yu, J. Tang, and Y. Yang, "Local rub-impact fault diagnosis of the rotor systems based on EMD," *Mechanism and Machine Theory*, vol. 44, no. 4, pp. 784–791, Apr. 2009, doi: 10.1016/j.mechmachtheory.2008.04.006.
- [129] L. Deng and R. Zhao, "Fault feature extraction of a rotor system based on local mean decomposition and Teager energy kurtosis," *Journal of Mechanical Science and Technology*, vol. 28, no. 4, pp. 1161–1169, Apr. 2014, doi: 10.1007/s12206-013-1149-9.
- [130] W. Yan, Z. Zhang, and S. Wang, "Time-frequency distribution decomposition with application to detection of rotor rub-impact fault," in *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, Pisa, Italy, May 2015, pp. 858–862, doi: 10.1109/I2MTC.2015.7151381.
- [131] L. Yungong, Z. Jinping, W. Liqiang, and C. Yang, "A fault feature extraction method for rotor rubbing based on load identification and measured impact response," *Procedia Engineering*, vol. 24, pp. 793–797, 2011, doi: 10.1016/j.proeng.2011.11.2739.
- [132] Z. Li, M. Ye, and M. Zhu, "Generalized S transform and its application in mechanical fault diagnosis," in *2016 Prognostics and System Health Management Conference (PHM-Chengdu)*, Chengdu, China, Oct. 2016, pp. 1–7, doi: 10.1109/PHM.2016.7819849.
- [133] S. Wang, X. Chen, G. Li, X. Li, and Z. He, "Matching Demodulation Transform With Application to Feature Extraction of Rotor Rub-Impact Fault," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 5, pp. 1372–1383, May 2014, doi: 10.1109/TIM.2013.2283552.
- [134] S. Liu, G. Tang, X. Wang, and Y. He, "Time-Frequency Analysis Based on Improved Variational Mode Decomposition and Teager Energy Operator for Rotor System Fault Diagnosis," *Mathematical Problems in Engineering*, vol. 2016, pp. 1–9, 2016, doi: 10.1155/2016/1713046.
- [135] X. Hong, Z. Xu, and Z. Zhang, "Abnormal Condition Monitoring and Diagnosis for Coal Mills Based on Support Vector Regression," *IEEE Access*, vol. 7, pp. 170488–170499, 2019, doi: 10.1109/ACCESS.2019.2955249.
- [136] M. Kordestani, M. F. Samadi, M. Saif, and K. Khorasani, "A New Fault Diagnosis of Multifunctional Spoiler System Using Integrated Artificial Neural Network and Discrete Wavelet Transform Methods," *IEEE Sensors Journal*, vol. 18, no. 12, pp. 4990–5001, Jun. 2018, doi: 10.1109/JSEN.2018.2829345.
- [137] J. Gu *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, May 2018, doi: 10.1016/j.patcog.2017.10.013.

-
- [138] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
- [139] R. Liu, G. Meng, B. Yang, C. Sun, and X. Chen, “Dislocated Time Series Convolutional Neural Architecture: An Intelligent Fault Diagnosis Approach for Electric Machine,” *IEEE Trans. Ind. Inf.*, vol. 13, no. 3, pp. 1310–1320, Jun. 2017, doi: 10.1109/TII.2016.2645238.
- [140] W. Zhang, G. Peng, and C. Li, “Bearings Fault Diagnosis Based on Convolutional Neural Networks with 2-D Representation of Vibration Signals as Input,” *MATEC Web Conf.*, vol. 95, p. 13001, 2017, doi: 10.1051/mateconf/20179513001.
- [141] X. Wu, Z. Penga, J. Renc, C. Chenga, W. Zhanga, and D. Wang, “Rub-impact Fault Diagnosis of Rotating Machinery Based on 1-D Convolutional Neural Networks,” *IEEE Sensors J.*, pp. 1–1, 2019, doi: 10.1109/JSEN.2019.2944157.
- [142] J. Antoni, “Fast computation of the kurtogram for the detection of transient faults,” *Mechanical Systems and Signal Processing*, vol. 21, no. 1, pp. 108–124, Jan. 2007, doi: 10.1016/j.ymsp.2005.12.002.
- [143] C. Wu and J. Xiao, “Application of variational mode decomposition to diagnose rub-impact fault of hydraulic turbine,” in *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, Chennai, Mar. 2017, pp. 799–802, doi: 10.1109/WiSPNET.2017.8299871.
- [144] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, “A complete ensemble empirical mode decomposition with adaptive noise,” in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011, pp. 4144–4147, doi: 10.1109/ICASSP.2011.5947265.
- [145] F. Piltan and J.-M. Kim, “Nonlinear Extended-state ARX-Laguerre PI Observer Fault Diagnosis of Bearings,” *Applied Sciences*, vol. 9, no. 5, p. 888, Mar. 2019, doi: 10.3390/app9050888.
- [146] F. Piltan, C.-H. Kim, and J.-M. Kim, “Advanced Adaptive Fault Diagnosis and Tolerant Control for Robot Manipulators,” *Energies*, vol. 12, no. 7, p. 1281, Apr. 2019, doi: 10.3390/en12071281.
- [147] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, “Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification,” *Signal Processing*, vol. 130, pp. 377–388, Jan. 2017, doi: 10.1016/j.sigpro.2016.07.028.
- [148] K. Bouzrara, T. Garna, J. Ragot, and H. Messaoud, “Decomposition of an ARX model on Laguerre orthonormal bases,” *ISA Transactions*, vol. 51, no. 6, pp. 848–860, Nov. 2012, doi: 10.1016/j.isatra.2012.06.005.
- [149] P. E. Hart, N. J. Nilsson, and B. Raphael, “A Formal Basis for the Heuristic Determination,” *IEEE TRANSACTIONS ON SYSTEMS SCIENCE AND CYBERNETICS*, p. 8, 1968.

-
- [150] Y. Qi, C. Shen, D. Wang, J. Shi, X. Jiang, and Z. Zhu, "Stacked Sparse Autoencoder-Based Deep Network for Fault Diagnosis of Rotating Machinery," *IEEE Access*, vol. 5, pp. 15066–15079, 2017, doi: 10.1109/ACCESS.2017.2728010.