LUCAS COSTA BRITO

EXPLAINABLE ARTIFICIAL INTELLIGENCE APPROACHES FOR FAULT DIAGNOSIS IN ROTATING MACHINERY

FEDERAL UNIVERISTY OF UBERLÂNDIA
SCHOOL OF MECHANICAL ENGINEERING
2022

LUCAS COSTA BRITO

EXPLAINABLE ARTIFICIAL INTELLIGENCE APPROACHES FOR FAULT DIAGNOSIS IN ROTATING MACHINERY

Thesis submitted to the Graduate Program in Mechanical Engineering at the Federal University of Uberlândia, in partial fulfillment of the requirements to obtain the degree of **DOCTOR IN MECHANICAL ENGINEERING**.

Concentration area: Mechanics of Solids and Vibrations.

Advisor: Prof. Dr. Marcus Antonio Viana Duarte
Co-Advisor: Prof. Dr. Gian Antonio Susto

Uberlândia - MG
2022

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

B862 Brito, Lucas Costa, 1993-
2022 EXPLAINABLE ARTIFICIAL INTELLIGENCE APPROACHES FOR FAULT DIAGNOSIS IN ROTATING MACHINERY [recurso eletrônico] / Lucas Costa Brito. - 2022.

Orientador: Marcus Antonio Viana Duarte.
Coorientador: Gian Antonio Susto.
Tese (Doutorado) - Universidade Federal de Uberlândia, Pós-graduação em Engenharia Mecânica.
Modo de acesso: Internet.
Disponível em: http://doi.org/10.14393/ufu.te.2022.557
Inclui bibliografia.
Inclui ilustrações.

1. Engenharia mecânica. I. Duarte, Marcus Antonio Viana,1959-, (Orient.). II. Susto, Gian Antonio,1984-, (Coorient.). III. Universidade Federal de Uberlândia. Pós-graduação em Engenharia Mecânica. IV. Título.

CDU: 621

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:
Gizele Cristine Nunes do Couto - CRB6/2091
Nelson Marcos Ferreira - CRB6/3074

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
Coordenação do Programa de Pós-Graduação em Engenharia Mecânica
Av. João Naves de Ávila, nº 2121, Bloco 1M, Sala 212 - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902
Telefone: (34) 3239-4282 - www.posmecanicaufu.com.br - secposmec@mecanica.ufu.br

ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Engenharia Mecânica				
Defesa de:	Tese de Doutorado Acadêmico, nº 340, COPEM				
Data:	21/10/2022	Hora de início:	09:00	Hora de encerramento:	12:20
Matrícula do Discente:	11823EMC011				
Nome do Discente:	Lucas Costa Brito				
Título do Trabalho:	EXPLAINABLE ARTIFICIAL INTELLIGENCE APPROACHES FOR FAULT DIAGNOSIS IN ROTATING MACHINERY				
Área de concentração:	Mecânica dos Sólidos e Vibrações				
Linha de pesquisa:	Dinâmica de Sistemas Mecânicos				
Projeto de Pesquisa de vinculação:	"*Projeto Edge Analytics - Aplicação de técnica de inteligência artificial na captura de dados de máquinas para armazenamento de dados de anormalidades em servidores*"				

Reuniu-se por meio de videoconferência a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Engenharia Mecânica, assim composta: Professores Doutores: Gian Antonio Susto (co-orientador)- Università degli Studi di Padova; Antônio Marcos Gonçalves de Lima - FEMEC/UFU; Gilmar Guimarães - FEMEC/UFU; Kátia Lucchesi Cavalca Dedini - UNICAMP, Robson Pederiva - UNICAMP e Marcus Antonio Viana Duarte - FEMEC/UFU, orientador do candidato.

Iniciando os trabalhos, o presidente da mesa, Dr. Marcus Antonio Viana Duarte, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(às) examinadores(as), que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

Aprovado.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Doutor.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.

Documento assinado eletronicamente por **Gilmar Guimarães**, **Professor(a) do Magistério Superior**, em

21/10/2022, às 12:25, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Katia Lucchesi Cavalca Dedini**, **Usuário Externo**, em 21/10/2022, às 12:26, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Robson Pederiva**, **Usuário Externo**, em 21/10/2022, às 12:26, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Antonio Marcos Gonçalves de Lima**, **Professor(a) do Magistério Superior**, em 21/10/2022, às 12:27, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Gian Antonio Susto**, **Usuário Externo**, em 24/10/2022, às 11:18, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Marcus Antonio Viana Duarte**, **Usuário Externo**, em 25/10/2022, às 08:18, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3967309** e o código CRC **6F0EC479**.

"Don't find fault. Find a remedy".

Henry Ford

Acknowledgements

To the Federal University of Uberlândia (UFU) and the Faculty of Mechanical Engineering for all the infrastructure and assistance.

To the *great master* Prof. D.Sc. Marcus Antônio Viana Duarte, for all the advice along the way.

To Prof. D.Sc. Gian Antonio Susto, for all the contributions, for having accepted this project, and providing a unique opportunity in my life.

To Prof. D.Sc. Aldemir Aparecido Cavalini Júnior, for the opportunities and teachings provided.

To my family for all the support and love. In particular, to my eternal advisor and father, Prof. D.Sc. Jorge Nei Brito, to my mother, Cláudia Borges Costa Brito, for her love and unique wisdom and to *Dona* Madá for always remembering the joy of living.

To my wife and her family for their love, attention and understanding whenever necessary.

To all UFU colleagues and technicians for their contribution to the development of this work.

To my friends for encouraging me to always pursue the goals set.

To Petrobras, CNPq (National Council for Scientific and Techno- logical Development) and CAPES (Federal Agency for the Support and Improvement of Higher Education) for financial support.

Brito, L. C., **Abordagens de Inteligência Artificial Explicáveis para Diagnóstico de Falhas em Máquinas Rotativas.** 2022. 177 f. Tese de Doutorado, Universidade Federal de Uberlândia, Uberlândia-MG, Brasil.

RESUMO

Devido ao crescente interesse pelo aumento da produtividade e redução de custos no ambiente industrial, novas técnicas de monitoramento de máquinas rotativas estão surgindo. A Inteligência Artificial (IA) é uma das abordagens que tem sido proposta para analisar os dados coletados (por exemplo, sinais de vibração) fornecendo um diagnóstico da condição de operação do ativo. Atualmente, vários modelos baseados em aprendizado de máquina e aprendizado profundo têm alcançado excelentes resultados no diagnóstico de falhas. No entanto, para aumentar ainda mais a adoção e difusão de tais tecnologias, usuários e especialistas humanos devem receber explicações dos modelos. Outra questão está relacionada, na maioria dos casos, com a indisponibilidade de dados históricos rotulados que inviabilizam o uso de modelos supervisionados. Para superar esses problemas, esta tese propõe novas metodologias para diagnóstico de falhas em máquinas rotativas baseadas em inteligência artificial explicável e análise de vibração. A possibilidade de reduzir o número de parâmetros monitorados é apresentada. Uma nova estrutura é proposta para identificação automática de bandas de frequência relevantes em sinais de vibração, chamada de Fator de Relevância de Banda (BRF). Além disso, uma identificação não supervisionada da falha em máquinas rotativas através da análise de vibração e classificação não supervisionada do tipo de falha, com base na análise da relevância dos parâmetros é apresentada. Por fim, é desenvolvida uma nova abordagem de classificação baseado na transferência de aprendizado em um conjunto de dados sintético, sem a necessidade de ter sinais de condições reais de falha, denominada Detecção de Falhas usando Inteligência Artificial Explicável (FaultD-XAI). A eficácia das abordagens propostas é mostrada em diferentes conjuntos de dados contendo falhas mecânicas em máquinas rotativas.

Palavras Chave: Inteligência Artificial Explicável, Diagnóstico de Falhas, Máquinas Rotativas, Manutenção Preditiva.

Brito, L. C., **Explainable Artificial Intelligence Approaches for Fault Diagnosis in Rotating Machinery.** 2022. 177 f. Ph.D. thesis, Federal University of Uberlândia, Uberlândia-MG, Brazil.

ABSTRACT

Due to the growing interest for increasing productivity and cost reduction in industrial environment, new techniques for monitoring rotating machinery are emerging. Artificial Intelligence (AI) is one of the approaches that has been proposed to analyze the collected data (e.g., vibration signals) providing a diagnosis of the asset's operating condition. Currently, several machine learning and deep learning-based modules have achieved excellent results in fault diagnosis. Nevertheless, to further increase user adoption and diffusion of such technologies, users and human experts must be provided with explanations and insights by the models. Another issue is related, in most cases, with the unavailability of labeled historical data that makes the use of supervised models unfeasible. To overcome these problems, this thesis proposes new methodologies for fault diagnosis in rotating machinery based on explainable artificial intelligence and vibration analysis. The possibility of reducing the number of monitored features is presented. A novel framework is proposed for automatic identification of relevant frequency bands in vibration signals, called Band Relevance Factor (BRF). Moreover, an unsupervised identification of the fault in rotating machinery through vibration analysis and unsupervised classification of the type of fault, based on the analysis of the features relevance is introduced. Finally, a new transfer learning classification approach based on a synthetic dataset, without the need to have signals of real fault conditions is developed, namely Fault Diagnosis using eXplainable AI (FaultD-XAI). The effectiveness of the proposed approaches is shown on different datasets containing mechanical faults in rotating machinery.

Keywords: Explainable Artificial Intelligence, Fault Diagnosis, Rotating Machinery, Predictive Maintenance.

List of Figures

List of Tables

List of Symbols

α_{gauss}	Gaussian noise coefficient
β	Bearing - contact angle
θ	Phase
A	Peak amplitude of the signal
$ABOD$	Angle-Based Outlier Detection
AD	Anomaly Detection
AE	Neural Network Autoencoder
AI	Artificial Intelligence
ANN	Artificial Neural Network
AS	Anomaly Score
$BPFI$	Ball Pass Frequency Inner
$BPFO$	Ball Pass Frequency Outer
$BPSO$	Binary Particle Swarm Optimization
BRF	Band Relevance Factor
BSF	Ball Spin Frequency
BW	Bandwidth
$CBLOF$	Cluster-based Local Outlier Factor
CNN	Convolutional Neural Network
CVB	Conditional Variance Based
D_1	Bearing - outer diameter
D_2	Bearing - inner diameter
dB	Decibels
df	Frequency Resolution

DL	Deep Learning
E	Energy
EEG	Electroencephalogram
$Eigen-CAM$	Eigenvector-based class activation map
f	Frequency
$FastABOD$	Fast-Angle-Based Outlier Detection
$FaultD-XAI$	Fault Diagnosis using eXplainable AI
FB	Feature Bagging
FBS	Frequency Band Selection
FD	Final Decision
$FLAC$	Fourier Local Autocorrelation
FN	False Negative
FP	False Positive
fr	Rotating Frequency
fs	Sampling Frequency
G	Gaussian Noise
g_{fr}	Running speed of the gear
GI	Gini Index
GMF	Gear Mesh Frequency
GMM	Gaussian Mixture Model
$Grad-CAM$	Gradient-weighted Class Activation Mapping
H	Shannon Entropy
$HBOS$	Histogram-based outlier score
$hist_(x)$	Density estimation
ICA	Independent Component Analysis
IF	Isolation Forest
IFB	Informative Frequency Band
IFD	Intelligent Fault Diagnosis
$IMDE$	Improved Multiscale Dispersion Entropy
IoT	Internet of Things

$ISOMAP$	Isometric Feature Mapping
kNN	k-Nearest Neighbor
LFI	Local Feature Importance
$LocalDIFFI$	Local Depth-based Feature Importance for the Isolation Forest
$LODA$	Lightweight on-line detector of anomalies
LOF	Local Outlier Factor
LRD	Local Reachability Density
MCD	Minimum Covariance Determinant
MCP	Maintenance Planning and Control
MED	Minimum Entropy Deconvolution
ML	Machine Learning
MLP	Multi-Layer Perceptrons
MPO	Leave P Out
$mRMR$	Max-relevance Min-redundancy
N	Multiple of the rotation
N_b	Bearing - number of ball
n_{teeth}	Number of teeth on the gear
n_{total}	Total of signals
$OCSVM$	One-class Support Vector Machines
p	Probability of the time series
P_x	Probability Distribution
P_{noise}	Rms value in volts for the noise
P_{signal}	Rms value in volts for the signal
PA	Position Analysis
PCA	Principal Component Analysis
PdM	Predictive Maintenance
$PR-AUC$	Precision-Recall Area Under the Curve
rms	Root Mean Square
rms_{base}	Root mean square value of the original signal
rms_{diff}	Correction Factor

$rms_{filtered}$	Root mean square value of the filtered band
rpm	Revolution Per Minute
RUL	Remaining Useful Life
s	Anomaly Score
$S(x)$	Spectral Entropy
S_{diff}	Entropy Difference Factor
$SBDS$	Smart Bearing Diagnosis System
$SHAP$	Shapley Additive Explanations
SNR	Signal-to-noise ratio
SVD	Singular Value Decomposition
SVM	Support Vector Machines
t	Time vector
$t-SNE$	t-distributed Stochastic Neighbor Embedding
TN	True Negative
TP	True Positive
VA	Values Analysis
WCM	World Class Maintenance
x	Vibration Signal in time domain
x_a	Augmented Signal
x_e	Enveloped Signal
x_f	Vibration Signal in frequency domain
x_r	Real Signal
x_s	Synthetic Signal
$x_{filtered}$	Filtered Signal
XAI	Explainable Artificial Intelligence

SUMMARY

CHAPTER I

INTRODUCTION

Rotating machinery is commonly used in mechanical systems and plays an important role in industrial applications Lei et al. (2013). Among the various types of rotating machinery, the following popular types are typically adopted in industrial application: aeroengine, steam, gas and wind turbine, automobile transmission, drive trains, fans, blowers, machine tools, compressors, motors, pumps, gearboxes and so on. Even though the rotating machinery is diversified, it generally includes some common essential rotating parts, such as rotors, rolling element bearings, and gears Lei (2017).

The rotor is the rotating part of the equipment and some common faults are: mass unbalance, looseness, bent, misalignment, rub. A rolling element bearing is a component that carries loads by placing rolling elements (such as balls or rollers) between two races. Generally, the rolling bearing consists of four components: outer race, inner race, cage and rolling elements which can suffer from different problems (improper mounting, poor lubrication, entry of foreign matter). All of these problems can cause different bearing faults including: flaking, spalling, peeling, abrasion, scoring, corrosion, pitting, crack, material failure. A gear is a rotating machine part having cut teeth, which meshes with another toothed part to transmit torque and motion. Geared devices can change the speed, torque, and direction of a power source Lei (2017). There are a number of reasons for which a gear may fail in mechanical systems: tooth wear, tooth load, gear eccentricity, backlash, gear misalignment, broken or cracked teeth. Due to the complexity of the systems any fault of

the rotating machinery can cause the complete failure of the system, and therefore must be monitored.

Among the main monitoring techniques currently used, predictive maintenance (PdM) stands out, which include: vibration, oil, thermography analysis, etc. Due to advances in monitoring systems and methods for predicting remaining useful life (RUL), PdM has increasingly become a focus of interest for professionals and researchers, Bousdekis et al. (2018). In addition, since vibration analysis is one of the non-invasive techniques, which presents the greatest amount of information about the monitored component, its use in industry increases every day. It is important to note that, depending on the type of asset being monitored, other predictive techniques are also commonly used to monitor: stator current, stray fluxes, thermal image, oil, noise level, etc.

The study of artificial intelligence (AI) techniques applied in the monitoring of rotating machinery is a topic in continuous development and of great interest by both researchers and industrial engineers. More and more, industries are adopting sophisticated technologies for monitoring to increase the reliability and availability of the machines, and, consequently, remaining competitive in the globalized economy.

Recently, Lei et al. (2020) presented a broad review with more than 400 citations, focused on AI applications for fault detection. The authors provide a historical overview, current developments and future prospects. Among the revised Machine Learning (ML) methods employed in the field, the authors recognized the following as the most commonly adopted: ANNs, Decision Trees, kNN, Probabilistic Graphical Model and Support Vector Machines (SVM). Moreover, the following Deep Learning (DL) approaches are taken into consideration: Autoencoders, Convolutional Neural Networks, Deep Belief Network, Residual Neural Networks. The authors also confirm the dependence on real and labeled data from the machine under analysis. In addition to highlighting the recent and future importance in the Intelligent Fault Diagnosis (IFD) scenario of explainable models, with increasing interest starting from 2017. Finally, they mention that traditional ML models should not be abandoned despite the recent advances of DL: this is because it is still worth investigating statistical learning in IFD with the big data revolution, since the theories of statistical learning have rigorous theoretical bases, which promote the construction of

diagnostic models with parameters, characteristics and results that are easy to understand.

Another important topic in the monitoring of AI-based rotating machines is dimensionality reduction. When using ML algorithms, usually relevant features are extracted from the vibration signal to be used as inputs in the model. Although the extraction is focused on features that are relevant for monitoring the equipment, due to the large number of faults that may exist, many features can be generated, increasing the size of the dataset. To deal with high dimensional datasets, feature selection and dimensionality reduction is often performed as a pre-processing step to achieve efficient storage of the data and robustness of AI algorithms Zocco et al. (2021).

As detailed by Liu et al. (2018) there are three basic tasks of fault diagnosis: (1) determining whether the equipment is normal or not (Anomaly Detection); (2) finding the incipient fault and its reason (Fault Classification); (3) predicting the trend of fault development (Fault Prognosis). It is clear that when determining the type of fault and its reason (task 2), consequently the answer on the condition of the equipment is obtained (task 1). However, the AI models usually used to classify the type of fault require to be trained with labeled data (supervised training), and examples for all conditions, which in most of the cases is not available in the industry Carletti et al. (2019). In addition, motivated by the recent advances in Deep Learning (DL), the vast majority of AI technologies lack of explainability traits and they require a large volume of data labeled for both normal and fault conditions, dramatically limiting their industry application.

An important aspect in the field that has not been fully explored yet is the one related to interpretability of AI-based monitoring solutions in rotating machinery: as argued above, without providing explainable results to the user, even when AI- based modules provide excellent results in historical data, AI models are unlikely to be applied in real-world scenarios Molnar (2020). Moreover, as mentioned by Lei et al. (2020), collecting labeled data from machines generates a high cost, and consequently unlabeled data is the majority in engineering scenarios.

Recently studies are being developed with a focus on Explainable Artificial Intelligence (XAI). In order to explain black-box models, different methods can be used according to the AI model in use Du et al. (2019). In general, the methods provide

information to understand how the model performs fault detection, which can be, for example, a ranking of the most important features, the model weight relevance or the most significant points in the underlying signals Doshi-Velez and Kim (2017). Despite the current interest, the vast majority of studies are focused on explainability for DL models and mostly on fault classification. More information can be found in the articles available on the topic Lei et al. (2016); Zhang et al. (2017); Jia et al. (2018); Li, Zhao, Sun, Cheng, Chen, Yan and Gao (2019); Abid et al. (2020); Li, Zhang and Ding (2019); Chen and Lee (2020); Grezmak et al. (2019, 2020); Saeki et al. (2019).

Therefore, to overcome the limitations presented, the work proposes the development and study of methodologies for monitoring rotating machinery based on explainable artificial intelligence and vibration analysis.

1.1 Objectives

The main goal of this work is to present new methodologies for monitoring rotating machinery based on explainable artificial intelligence and vibration analysis, and consequently enable engineering applications.

The other objectives are:

i) a comparison of different features extracted from the vibration signal and dimensionality reduction techniques, in the unsupervised detection of faults in rotating machinery (anomaly detection);

ii) a novel framework for automatic identification of relevant frequency bands in vibration signals, called Band Relevance Factor (BRF);

iii) unsupervised identification of the fault in rotating machinery through vibration analysis and unsupervised classification of the type of fault in rotating machinery, based on the analysis of the features relevance;

iv) a new explainable transfer learning classification approach based on a synthetic dataset, without the need to have signals of real fault conditions, namely FaultD-XAI (Fault Diagnosis using eXplainable AI);

In addition to the highlighted objectives, a new dataset was created and will be

publicly available, simulating real situations, to study the defects: unbalance, mechanical looseness and misalignment.

1.2 Thesis structure

The work is divided into six chapters. Although all studies are directly related to the monitoring of rotating machinery using artificial intelligence, each methodology involves a specific area of the theme, and therefore, they were divided into separate chapters. Thus, each chapter is composed of a specific introduction to the topic and theoretical foundation. Subsequently, the experimental procedure is presented. The analyzes and results are discussed, and finally, a specific conclusion is presented.

In Chapter I, the general introduction on the thesis theme and its objectives is presented, placing the study in the context of current research. The development of the work is presented in Chapters II to VI. Chapter II addresses the problem of relevance and excess of features in the analysis and the possibility of working with dimensionality reduction techniques (objective - i). Chapter III presents a new automatic methodology (Band Relevance Factor - BRF) to determine the most relevant frequency bands in the signal, contributing to the extraction of features and explainability when used in AI models (objective - ii). In Chapter IV a new contribution to the unsupervised identification of the fault in rotating machinery through vibration analysis and unsupervised classification of the type of fault in rotating machinery, based on the analysis of the features relevance (XAI) is presented (objective - iii). Aiming to solve the lack of real and labeled data to train AI models, in addition to the lack of interpretability of the results, Chapter V presents a new transfer learning classification approach based on a synthetic dataset (objective - iv), without the need to have signals of real fault conditions (Fault Diagnosis using eXplainable AI - FaultD-XAI). Finally, the main conclusions and suggestions for future work are presented in Chapter VI.

CHAPTER II

FAULT DETECTION OF BEARING: AN UNSUPERVISED MACHINE LEARNING APPROACH EXPLOITING FEATURE EXTRACTION AND DIMENSIONALITY REDUCTION

The monitoring of rotating machinery is an essential activity for asset management today. Due to the large amount of monitored equipment, analyzing all the collected signals / features becomes an arduous task, leading the specialist to rely often on general alarms, which in turn can compromise the accuracy of the diagnosis. In order to make monitoring more intelligent, several machine learning techniques have been proposed to reduce the dimension of the input data and also to analyze it. This chapter, therefore, aims to compare the use of vibration features extracted based on machine learning models, expert domain, and other signal processing approaches for identifying bearing faults (anomalies) using machine learning (ML)—in addition to verifying the possibility of reducing the number of monitored features, and consequently the behavior of the model when working with reduced dimensionality of the input data. As vibration analysis is one of the predictive techniques that present better results in the monitoring of rotating machinery, vibration signals from an experimental bearing dataset were used. The proposed features were used as input to an unsupervised anomaly detection model (Isolation Forest) to identify bearing fault. Through the study, it is possible to verify how the ML model behaves in view of the different possibilities of input features used, and their influences on the final result in addition to the possibility of reducing the number of features that are usually monitored by reducing the dimension. In addition

to increasing the accuracy of the model when extracting correct features for the application under study, the reduction in dimensionality allows the specialist to monitor in a compact way the various features collected on the equipment.

2.1 Introduction

Rotating machinery plays an important role in industrial applications Lei et al. (2013). Among the various types of essential rotating parts, rolling element bearings are considered one of the most important. Among the main monitoring techniques currently used, predictive maintenance (PdM) stands out, which include: vibration, oil, thermography analysis, etc. Due to advances in monitoring systems and methods for predicting remaining useful life (RUL), PdM has increasingly become a focus of interest for professionals and researchers Bousdekis et al. (2018). In addition, since vibration analysis is one of the non-invasive techniques, which presents the greatest amount of information about the monitored component, its use in industry increases every day. It is important to note that, depending on the type of asset being monitored, other predictive techniques are also commonly used to monitor: stator current, stray fluxes, thermal image, oil, noise level, etc.

With the reduction in the cost of hardware, more and more sensors are being implemented in the industrial field, considerably increasing the amount of collected signals. Such signals must be analyzed by specialists in order to identify possible faults in the components, so that the planned corrective maintenance can be carried out. However, due to the large amount of generated signals, there are not always enough specialists to analyze all the signals or sufficiently reliable alarm techniques to make the analysis automatic. Thus, several strategies using artificial intelligence have been studied. Among them, the unsupervised fault detection, also called anomaly detection, stands out. Although the field of rotating machinery monitoring is widely developed, a small number of unsupervised approaches have been presented, in relation to the vast majority focused on supervised classification and prognostics, as shown in the review works Liu et al. (2018); Stetco et al. (2019). Anomaly detection consists of identifying unexpected events that vary greatly from normal events, as they usually present different characteristic patterns.The ability to

work unsupervised makes it possible to overcome the problem of obtaining labels in real applications.

When using ML algorithms, usually relevant features are extracted from the vibration signal to be used as inputs in the model. Although the extraction is focused on features that are relevant for monitoring the equipment, due to the large number of faults that may exist, many features can be generated, increasing the size of the dataset. To deal with high dimensional datasets, unsupervised dimensionality reduction is often performed as a pre-processing step to achieve efficient storage of the data and robustness of ML algorithms Zocco et al. (2021).

Dimensionality reduction is an important tool in the use of ML algorithms, which can help to avoid some common problems, such as: (i) Curse of dimensionality: due to the high number of features in relation to the sample, the algorithm tends to suffer overfitting, fitting very well to the training data, but showing a high error rate in the test group. (ii) Occam's Razor: to be used in real applications, the models are intended to be simple and explainable. The greater the number of features present, the greater the difficulty in explaining the model under development. Consequently, real applications become unfeasible. (iii) Garbage In Garbage Out: when using features that do not present significant information for the model, the final result obtained will be lower than desired. In other words, low quality inputs produce bad outputs. To overcome these problems, it is essential to perform feature selection and dimensionality reduction in the dataset.

Therefore, this chapter aims to present a comparison of different features extracted from the vibration signal and dimensionality reduction techniques, in the unsupervised detection of faults in rotating machinery (anomaly detection). In addition to the main objective, different features for monitoring bearing faults are proposed. The proposed methodology has the great advantages over traditional methods of anomaly detection, allowing to work with a reduced number of features, which results in: (i) possibility of follow-up of features by specialists—assists in data visualization (given that some assets can present more than 100 features acquired in real time, which makes detailed monitoring of all impracticable); (ii) avoid introducing irrelevant or correlated features in machine learning models, which would result in a loss of learning quality and, consequently, a reduction in the

success rate; (iii) reduced data storage space; and (iv) less computational time for training the models.

The remainder of this chapter starts with a brief explanation about feature extraction, dimensionality reduction, anomaly detection Section 2.2, and the methodology is presented in Section 2.3. Results and discussion are shown in Section 2.5. Finally, Section 2.6 concludes this chapter.

2.2 Background

2.2.1 Feature Extraction

Monitoring rotating machines has a great advantage over other research fields, which is prior knowledge of the behavior and characteristics of the vast majority of machine failures. Such knowledge allows the application of ML models, and, therefore, it was decided to work with 'classic' ML techniques, exploring the wide knowledge of filtering approaches and features definitions provided by the literature.

Monitoring through the vibration signal is one of the non-invasive techniques that provide the greatest amount of information about the dynamic behavior of the asset, and, for this reason, it is of growing interest Ciabattoni et al. (2018); Wei et al. (2019). In addition to vibration, other techniques and physical variables can be measured for monitoring. Despite this, the vibration analysis technique stands out, among other reasons, because: (i) it does not need to stop the asset to perform the measurement; (ii) easy placement of the sensor for data acquisition (in the most common case of accelerometers); (iii) widespread knowledge about the characteristics of faults; (iv) fast acquisition time (in most cases), enabling the monitoring of a greater amount of assets; and (v) it provides information about mechanical, electrical, and even structural conditions.

As shown in Brito et al. (2022), the features to detect faults in rotating machinery using vibration signals are commonly extracted from:

(i) Time domain: mean, standard deviation, rms (root mean square), peak value, peak-to-peak value, shape indicator, skewness, kurtosis, crest factor, clearance indicator, etc.

(ii) Frequency domain: mean frequency, central frequency, energy in frequency bands, etc.

(iii) Time-frequency domain: entropy are usually extracted by Wavelet Transform, Wavelet Packet Transform, and empirical model decomposition.

2.2.2 Dimensionality Reduction

The higher the number of features, the harder it gets to visualize the training set and then work on it. Sometimes, many of these features are correlated or redundant. Because of this, a fundamental tool for ML applications is dimensionality reduction. Dimensionality reduction can be done in two main ways: (i) keeping only the most relevant features from the original dataset (generally called feature selection); and (ii) reducing the original dataset into a new one through analysis/combinations of the input variables, where the new dataset contains basically the same information as the original (generally called dimensionality reduction). Different techniques can be used to reduce the dimensionality of the data obtained, such as: Principal Component Analysis (PCA), t-distributed Stochastic Neighbor Embedding (t-SNE), Isometric Feature Mapping (ISOMAP), Independent Component Analysis (ICA), and Neural Network Autoencoder (AE). In general, the objective is to reduce the number of features by creating new representative ones and thus discarding the originals. The new set, therefore, should be able to summarize most of the information contained in the original set of features.

The advantages are: (i) reduced data storage space; (ii) less computational time for training the models; (iii) better performance in some algorithms that do not work well in high dimensions; (iv) reduction of correlated variables; and (v) assistance with data visualization. On the other hand, some disadvantages can be mentioned, such as: (i) loss of explainability of the features (when space transformation occurs) and (ii) lack of representativeness of the problem under analysis.

PCA Jolliffe (1986) is a linear transformation that seeks to find the low-dimensional subspace within the data that maximally preserve the covariance up to rotation. This maximum covariance subspace encapsulates the directions along which the data vary the most. Therefore, projecting the data onto this subspace can be thought of as projecting the data onto the subspace that retains the most information Strange and Zwiggelaar (2014).

ISOMAP measures the inter-point manifold distances by approximating geodesics

(rather than Euclidean distance as in Multidimensional scaling—MDS) Tenenbaum et al. (2000). Geodesic distance is the shortest distance between two points on a curve. The use of manifold distances can often lead to a more accurate and robust measure of distances between points so that points that are far away according to manifold distances, as measured in the high-dimensional space, are mapped as far away in the low-dimensional space Strange and Zwiggelaar (2014).

t-SNE computes the probability that pairs of data points in the high-dimensional space are related and then chooses a low-dimensional embedding which produce a similar distribution VanDerMaaten and Hinton (2008). It minimizes the Kullback–Leibler divergence between the two distributions with respect to the locations of the points in the map.

ICA is based on information-theory which transforms a set of vectors into a maximally independent set Jutten and Hérault (1991). It assumes that each sample of data are a mixture of independent components, and it aims to find these independent components. It is based on three main assumptions: (i) Mixing process is linear; (ii) All source signals are independent of each other; and (iii) All source signals have non-Gaussian distribution. The major difference between PCA and ICA is that PCA looks for uncorrelated factors while ICA looks for independent factors. At each step, ICA changes the basis vector (projection directions) and measures the non-Gaussianity of the obtained sources and at each step it takes the basis vectors more towards non-Gaussianity. After some stopping criteria, it reaches an estimation of the original independent sources. ICA extracts hidden factors within data by transforming a set of variables to a new set that is maximally independent and, typically, it is not used for reducing dimensionality but for separating superimposed signals.

Finally, the idea of autoencoders has been part of the historical landscape of neural networks for decades Goodfellow et al. (2016). Autoencoders are a type of artificial neural network that aims to copy their inputs to their outputs. They compress the input into a latent-space representation, and then reconstruct the output from this representation. The output from the bottleneck is used directly as the reduced dimensionality of the input.

2.2.3 Anomaly Detection (AD) and Isolation Forest (IF)

Anomaly detection (also known as outlier detection (The terms 'Anomaly' and 'Outlier' will be treated in the same way in this work)) refers to the task of identifying rare observations which differ from the general ('normal') distribution of a data at hand Zhao, Nasrullah and Li (2019). In other words, they are samples that have values so different from other observations that they are capable of raising suspicions about the mechanism from which they were generated Hawkins (1980). An important parameter of anomaly detection approaches is the ability to summarize a multivariate system in just one indicator, called Anomaly Score (AS) (Other authors refer to the concept of Anomaly Score with various names like for example Health Factor or Deviance Index). While only one study, to the best of our knowledge, has been presented in the field of rotating machinery monitoring using vibration data and state-of-art models Brito et al. (2022), AD approaches have been successfully applied in various areas like fraud detection and oil and gas Barbariol et al. (2020).

Isolation Forest (iForest or IF) is probably the most popular AD approach. It works well in high-dimensional problems that have a large number of irrelevant attributes, and in situations where a training set does not contain any anomalies. Given its high performance and the possibility to parallelize its computation (thanks to its ensemble structure), it was selected for the study. IF Liu et al. (2008) uses the concept of *isolation* instead of measuring distance or density to detect anomalies. The IF exploits a space partitioning procedure: the main idea underlying the approach is that an outlier will require less iterations than an inlier to be isolated. At the end of the partition procedure, an anomaly score is generated. If it is very close to 1, then they are tagged as anomalies; on the other hand, values much smaller than 0.5 are quite safe to classify the normal instances, and if values are close to 0.5, then the entire sample does not really have any distinct anomaly Liu et al. (2008). The complete flowchart exemplifying the steps presented above is shown in Figure 2.1.

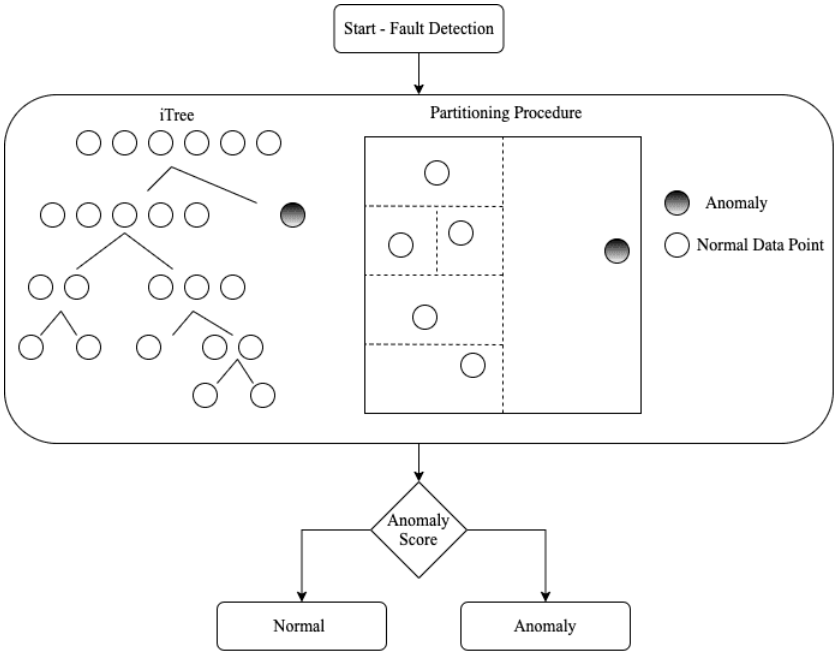


Figure 2.1: Flowchart of the anomaly detection method.

2.3 Proposed approach

The proposed methodology is divided into five main parts: (1) Data Acquisition; (2) Feature Extraction; (3) Dimensionality Reduction; (4) Fault detection: Anomaly Detection; and (5) Feature Trend Analysis, Figure 4.1. The data are acquired. The vibration features are initially extracted based on the type of monitored component. The dimensionality of each extracted feature and the raw signal is reduced. The features are divided into a training and testing group, and the hyperparameters of the anomaly detection models are tuned. The samples are evaluated using the original and reduced features in the fault detection part. Finally, a feature trend analysis is performed.

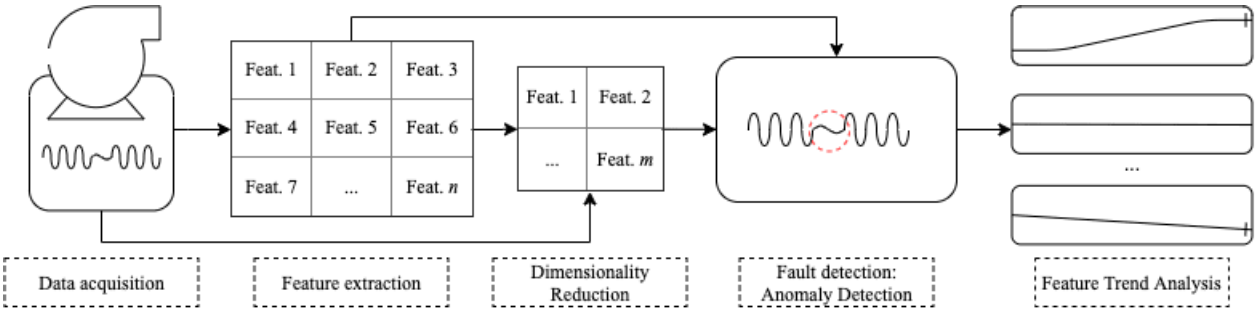


Figure 2.2: General framework of the proposed methodology.

2.3.1 Data Acquisition and Feature Extraction

Accelerometers were used to collect the vibration signal. The 22 features were carefully extracted from the vibration signal for the bearing analysis, taking into account the authors' experience and the most relevant features in the literature on the subject Bolón Canedo et al. (2013); Zhang et al. (2011, 2018); Lei and Zuo (2009); Li, Yang, Li, Xu and Huang (2017); Singh and Shaik (2019), as they are typical choices in signal processing. The selected features are shown in Table 2.1. Three energy features were calculated in the frequency bands of 10–1000 hz, 1000–4000 hz, and 4000–10,000 hz, and nine wavelet sub bands were extracted for entropy calculation.

Table 2.1 – Features extracted from the vibration signal

Features	Description	Features	Description
Absolute Energy	$\int \lvert x(t)\rvert^2 dt$	Root Mean Square (rms)	$\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$
Kurtosis	$\frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^4}{(\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2)^2}$	Skewness	$\frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^3}{(\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2)^{\frac{3}{2}}}$
Global value from envelope analysis peak-to-peak	$max(x_e) - min(x_e)$	Crest Factor	$max(\lvert x\rvert)/rms$
Principal Frequency	$max(x_{fi})$	Wavelet sub band entropy	$-\sum_{i=1}^{N} p_i * log(p_i)$
Ball Pass Frequency Outer (BPFI)	$r/min\frac{N_b}{2}(1+\frac{B_d}{P_d}cos(\beta))$	Ball Pass Frequency Inner (BPFO)	$r/min\frac{N_b}{2}(1-\frac{B_d}{P_d}cos(\beta))$
Ball Spin Frequency (BSF)	$r/min\frac{P_d}{B_d}[(1-\frac{B_d}{P_d}cos(\beta))^2]$		

Where a sampled vibration signal in time domain is defined as $x = \mathrm{x_1}, \mathrm{x_2}, \ldots, \mathrm{x_N}$, in frequency domain $x_f = \mathrm{x_{f1}}, \mathrm{x_{f2}}, \ldots, \mathrm{x_{fN}}$ and after the envelop analysis $x_e = \mathrm{x_{e1}}, \mathrm{x_{e2}}, \ldots, \mathrm{x_{eN}}$. p_i is the probability that each sub band in wavelet transform will be in state i from N possible states, x_{filt} is the filtered signal in the specific band, r/min is the rotation speed, D_1 bearing outer diameter and D_2 bearing inner diameter, N_b the number of ball, $P_d = (\mathrm{D_1}+\mathrm{D_2})/2$ and β the contact angle.

In case of bearing analysis, specific features are those that indicate the type of fault (BPFI, BPFO, and BSF) and the remaining features are those that indicate the presence of a defect. The bearing fault frequencies are important to assess the type of defect and confirm its existence, which is not always noticed by other features. It is also important mentioning that there are cases where the fault does not present the classic defect behavior with the deterministic bearing frequencies in evidence Smith and Randall (2015), which makes it important to use other features. Knowing that bearing faults are generally associated with

impacts, kurtosis is a relevant feature for the study. Impacts generally excite high frequencies, and with the evolution of the fault, new frequencies tend to appear in other bands, which can be noticed in the energy per sub band and in the wavelet frequency sub bands. The principal frequency can vary with the appearance of the defect, stabilizing and suffering changing with the fault evolution due to the random behavior caused by the excessive wear. Crest factor tends to increase as the amplitude of high frequency impacts in the bearing increase compared to the amplitude of overall broadband vibration. Skewness will provide information on how the signal is symmetrical with respect to its mean value. Finally, the *rms* value, global value from envelope analysis, and absolute energy represent the global behavior of the system, indicating a general degradation and accentuation of the defect Brito et al. (2022).

2.3.2 Dimensionality Reduction, Fault Detection, and Feature Trend Analysis

To perform the dimensionality reduction, different methods were used, namely: PCA, t-SNE, ISOMAP, ICA, and AE. Dimension reduction can be performed in different ways, which will consequently impact the final result. Firstly, the dimension of the features extracted from the raw signal was reduced. Furthermore, the dimension of the raw signal was also reduced, in order to verify the possibility of using the reduced signal directly in the ML model. Dimensionality reduction here is taken as a proxy to assess the goodness of the feature's importance and also the possibility of reducing the large number of features that are usually monitored.

Fault detection (AD) was performed using the ML model Isolation Forest. The following were used as inputs in the model: all original features extracted from the vibration signal, original features extracted from the vibration signal and manually selected, features extracted from the vibration signal with reduced dimension, and raw vibration signal with reduced dimension. The features used were plotted in the form of a trend in order to visually verify their respective values and possible deviations from the curve's behavior, helping the specialist to visualize the anomaly. Furthermore, through this analysis, it is possible to verify features that are more relevant to the problem under study.

2.4 Experimental Procedure

Since bearing is one of the most important components in rotating machinery, a bearing dataset (a benchmark for failure prediction in industry 4.0 Diallo et al. (2021)) was chosen for the study. The dataset considered publicly is provided by the University of Cincinnati Center for Intelligent Maintenance Systems (available in Lee et al. (2007) and described in Qiu et al. (2006)), namely Bearing Dataset, is composed by three run-to-failure tests with four bearings in each test and no labels are available. All four bearings were force lubricated. A PCB 353B33 High Sensitivity Quartz ICPs Accelerometer was installed on each bearing housing. To assess the efficiency of the AD model, the data were manually labeled. For the study, bearing 01 of test 02 was used. Due to the proximity of the other bearings, at the end of the test the fault was verified by visual analysis, and confirmed by envelope analysis. The bearing used was a Rexnord ZA-2115, and the speed was kept constant at 2000 r/min. The shaft was driven by an AC motor and coupled by rub belts and a radial load was added to the shaft through a spring mechanism. All failures occurred after exceeding the projected bearing life, which is more than 100 million revolutions. The vibration signals consist of 20,480 points with the sampling rate set at 20 kHz. Vibration data were collected every 20 minutes by a National Instruments DAQCard-6062E data acquisition card. Four thermocouple sensors were placed on the outer race of the bearings to monitoring temperature due to lubrication purposes Qiu et al. (2006). The experimental apparatus (bearing test rig) is shown in Figure 2.3.

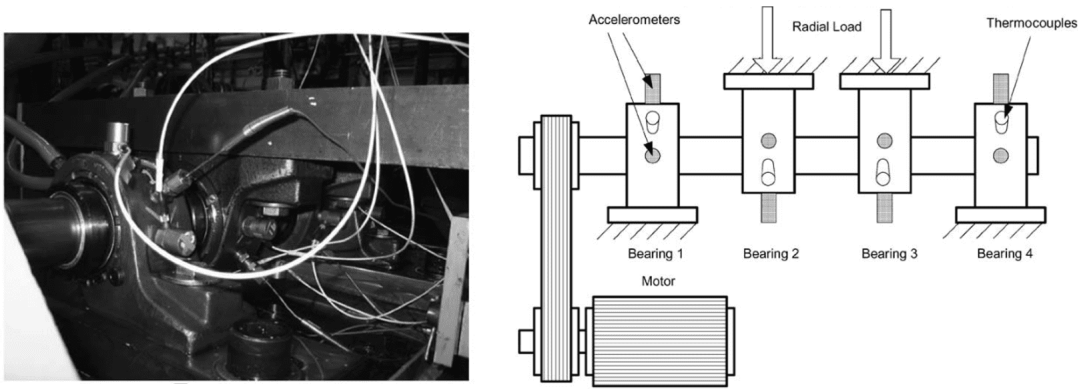


Figure 2.3: Experimental Apparatus (Bearing Test Ring) adapted from Qiu et al. (2006).

2.4.1 Tests and Analysis Approaches

Four tests were performed to evaluate the extracted features and the dimensionality reduction, as follows: (i) Anomaly detection using all features extracted from the raw signal; (ii) Anomaly detection using features selected manually from those extracted in test (i); (iii) Anomaly detection using the sets obtained by reducing the dimensionality of the features extracted in test (i); and (iv) Anomaly detection using the raw signal with reduced dimension through the PCA method.

For the fault detection using the Isolation Forest model, a dynamic condition was considered with the data collected in sequence, where a temporal relationship and fault evolution are presented. For the study, a sliding window was used, where the training group was updated with each new sample, in case it was considered normal. Twenty-five samples were initially used for the training group and, after each iteration, if the sample was considered normal, it was added to the training group, in order to ensure stability in the model. This approach was used in order to minimize as much as possible the amount of initial samples needed for the method to work, also ensuring its stability. For this situation, as the model was started together with the machine under normal conditions (e.g., after maintenance or a new machine), there are no anomalies in the training group. It is worth noting that this approach can also be used if there are anomalies in the training group (e.g., cases of continuous monitoring where the machine was repaired after a fault, and it is desired to use all the signals to increase the amount of data in the model).

2.4.2 Hyperparameter Tuning and Evaluation Metrics

The hyperparameters for the Isolation Forest model were adjusted based on the training group to obtain the best performance (100 estimator and 128 the maximum number of samples). A cross-validation procedure was applied, using a Leave P Out (LPO) approach for the dynamic condition, where 5% of the training samples were removed in each new update of the training group. The hyperparameters are presented in relation to the library used Zhao, Nasrullah and Li (2019).

As for the dimensionality reduction methods, two main components were used. The choice was to guarantee the possibility of data visualization, based on the PCA technique

(state-of-the-art), where 100% of the variance explained was obtained. In PCA, the full SVD (Singular Value Decomposition) was used. t–SNE was set using perplexity = 30 and early exaggeration = 10. ISOMAP was calculated using 5 near neighbors. For ICA, the functional form of the G function used in the approximation, the neg-entropy it used was 'logcosh' and 0.0001 of tolerance. For the AE, layers with dimensions 128 and 64, sigmoid optimization function, optimized 'adam', and error metric 'mse' were used. The hyperparameters are presented in relation to the scikit-learn library used.

An unsupervised methodology is proposed for the fault detection. The anomaly score is calculated, where samples with high anomaly score values are usually anomalies. Threshold values were defined based on the training group. The results are presented using the F1-Score and the average confusion matrix of the iterations with respective standard deviations. The tests were performed using 2.2 GHz Intel Core i7 Dual-Core, 8 GB 1600 MHz DDR3, Intel HD Graphics 6000 1536 MB.

2.5 Results and Discussion

2.5.1 Data Exploration

The data used in this work show evidence of incipient defect from sample 531, identified from the analysis of the signals, as shown in Figure 4.3a. The signal is present according to the sample (x-axis).

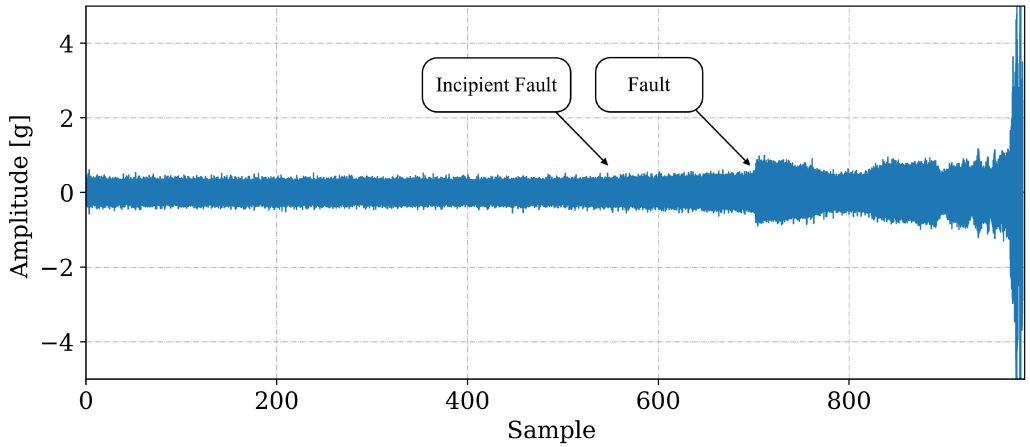


Figure 2.4: Complete waveform for the bearing under test, from start of monitoring to complete failure. The arrows indicate the instant where, through the analysis of the signals, the beginning of the fault (Incipient Fault) was noticed and the moment where the incipient fault progressed to a fault (Fault).

Despite indicating in Figure 4.3a that the incipient fault cannot be identified only

through temporal signal analysis, it is necessary to use other signal processing techniques such as envelope analysis, Figure 2.5. The use of the technique makes it possible to filter out other excitations that can hide the evidence of fault frequencies in bearings due to their low amplitude. This is also another factor that makes necessary to use relevant features in ML models. Figure 2.5 shows the moment of the beginning of the incipient fault indicated by the presence of fault frequencies (BPFO). Therefore, based on the analysis of the 984 observations, 531 were labeled as normal and 453 as anomalies (fault).

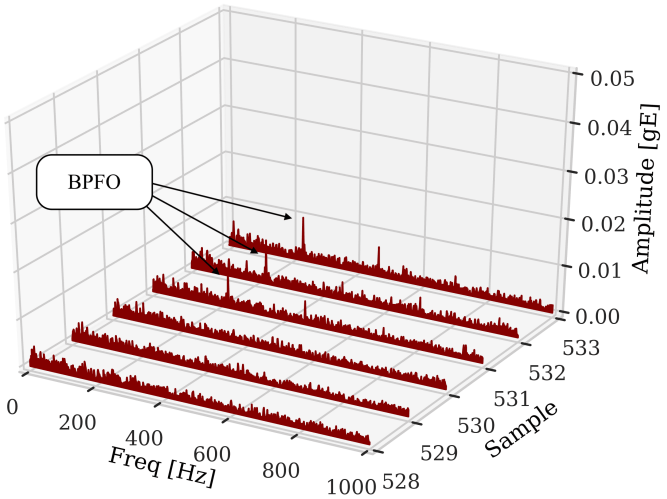


Figure 2.5: Waterfall envelope spectrum of the signals in normal operating condition (528,529,530) and after the beginning of the incipient fault (531) with the arrows at the characteristic frequencies of fault of the outer race in the bearing under analysis.

2.5.2 Fault Detection: Anomaly Detection

Using the proposed methodology, the F1-Score results with the respective standard deviation (in percentage) obtained for the fault detection are presented in Figure 2.6.

Among all the sets of proposed features, the set with features manually selected based on the type of fault under study was the one with the best result (test ii). This is due to the fact that the selected features were carefully chosen to present a high correlation with the fault, being extremely relevant for the detection of anomaly (e.g., BPFO which is exclusive for the type of fault present). The fact that the set has a small dimension also contributes to the result. As it is the test that showed the best result, its confusion matrix in percentage and the respective standard deviations are presented as: TN (True Negative) = 52.78 (0.06), TP (True Positive) = 45.43 (0.56), FN (False Negative) = 1.74 (0.56), and

FP (False Positive) = 0.08 (0.06)%. Analyzing the confusion matrix, it is possible to observe that there is a small amount of false negatives, which means that the method is able to identify all faults/anomalies, thus avoiding future breakdown. Another important point to be highlighted is that the errors happened at the beginning of the fault, and, with its progression, they no longer exist. Such phenomenon can occur even with human specialists due to the difficulty in detection of incipient faults. However, with the progression of the fault, it is possible to correctly identify and avoid a breakdown of the equipment.

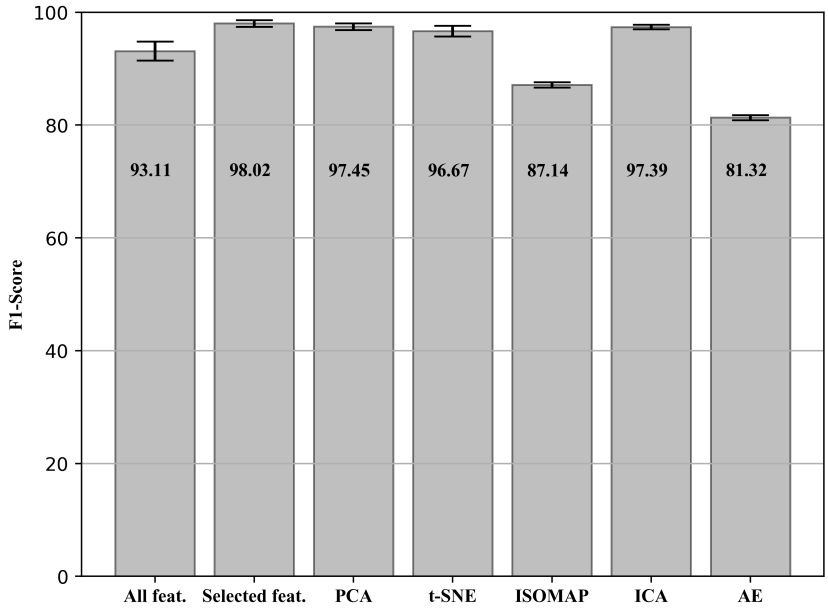


Figure 2.6: F1-Score and standard deviation (in percentage) obtained for the fault detection.

It is possible to notice that, when using all the extracted features (test i), the fault detection method presents a result inferior to the best obtained, and a larger standard deviation. Using all extracted features, features that were not relevant to the fault under study were intruded, resulting in information that tends to 'confuse' the model, and therefore reducing its assertiveness. Furthermore, a greater number of features tends to introduce bias and variance in the system, increasing the standard deviation of the results, and consequently reducing their robustness. The dimensionality reduction methods (test iii) showed good results, close to the set of manually selected features, mainly: PCA, t-SNE, and ICA. The result shows that the methods were able to reduce the size of the data in a representative way. In addition, it is also possible to use the methods for similar situations where there are several features to be monitored, and the analysis of it all is impracticable. Thus, it is possible to

proceed with the dimensionality reduction and follow only the obtained main components. If there is any detected variation, all available features should be analyzed. It is noteworthy that, despite being extremely useful tools for monitoring rotating machinery and artificial intelligence applications, when performing the domain transformation, the explainability of these features is lost, which can be harmful for real applications.

2.5.3 Trend Analysis: Extracted Features

Analyzing Figure 2.7, it can be noticed that the features present different behaviors in relation to the analyzed signals. Ideally, a variation of the features is expected in sample number 531, when it is possible to notice the appearance of fault frequencies through the envelope analysis, which in turn is a good tool to detect incipient defects in bearings.

As expected, Figure 2.7, frequency range, and wavelet features are good indicators to monitor bearing RUL (except for low frequency bands/wavelet, due to the fault type), following the concept of the four stages of life. In general, for bearing faults, initially, excitation at very high frequency (>aprox.20khz) is possible to be detected by specific techniques such as: acoustic emission—followed by the second stage with high frequency (>approx.1–2khz), capable of being detected by envelope analysis, for example. In the third stage, there is an increase in the amplitudes related to the fault frequencies, which can be seen in the acceleration and velocity spectrum. Finally, where the failure is imminent, the spectrum floor is raised, and the spectrum does not have the harmonics, but the noise floor is considerably higher, and very high frequency vibration may trend downwards (smoothing of metal reduces sharp impacts).

As features conventionally used to detect failures in rotating machines were extracted in general, not all parameters show indication of the defect since its incipient phase, as can be seen in Figure 2.8 for the BPFO, which is the type of fault present in the bearing. Thus, as expected, this parameter is able to characterize the failure from its incipient stage.

The features were extracted to cover different types of failures in rotating machinery; therefore, not all of them present an indication of the defect since its incipient stage. On the other hand, specific features for fault identification, as can be seen in Figure 2.8

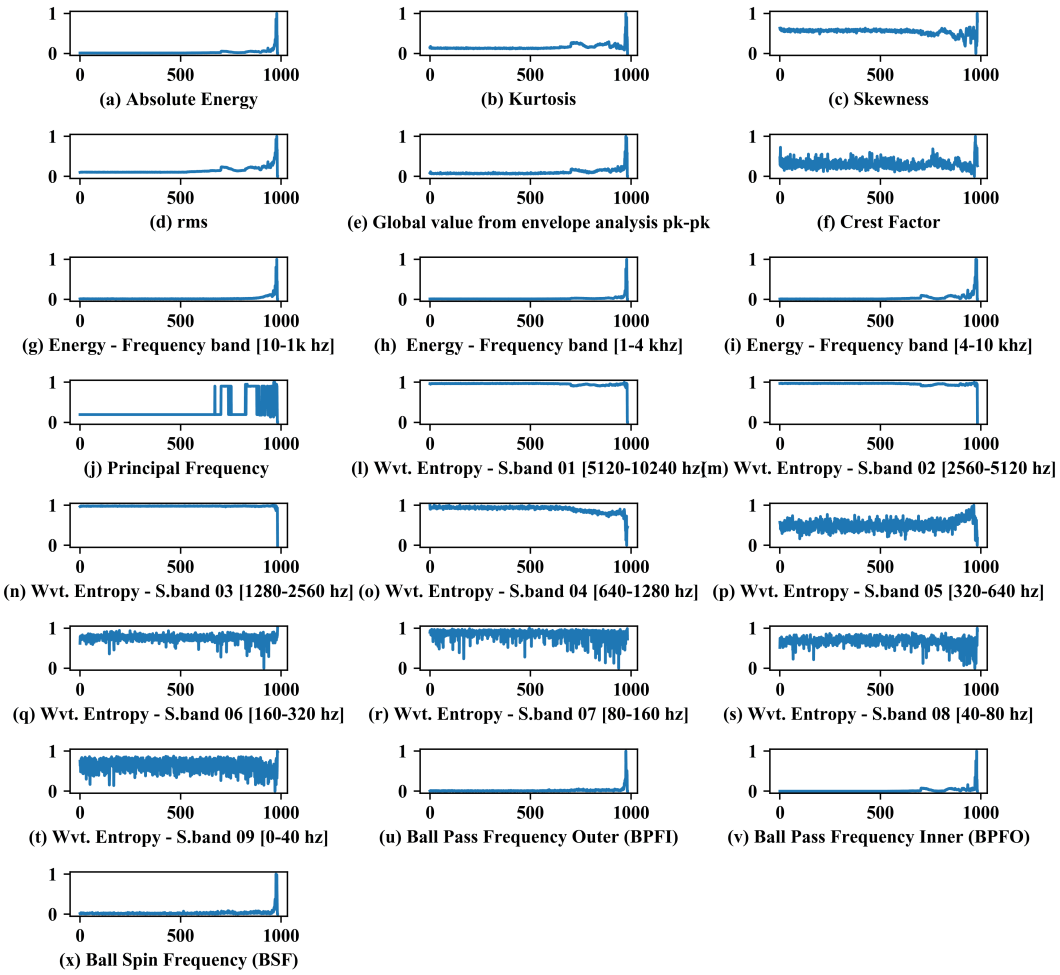


Figure 2.7: Trend analysis and behavior of extracted features during testing.

for BPFO, were able to characterize the fault since its incipient stage.

Features such as absolute energy, kurtosis, skewness, rms, and global value from envelope analysis pk-pk can be characterized as relevant for identifying the fault under study, given the variation in the trend with respect to fault progression. The crest factor and other characteristic frequencies of bearing failures (BPFI and BSF) presented variations but less significant in relation to the others. Crest factor showed a noisy variation. BPFI and BSF, for not showing correlation with bearing fault, showed significant variation only towards the end of life. The principal frequency feature presented variations, mainly after the fault aggravation, which can be explained by the increase in the noise floor, and consequently variations in the main frequencies for each sample due to its random behavior.

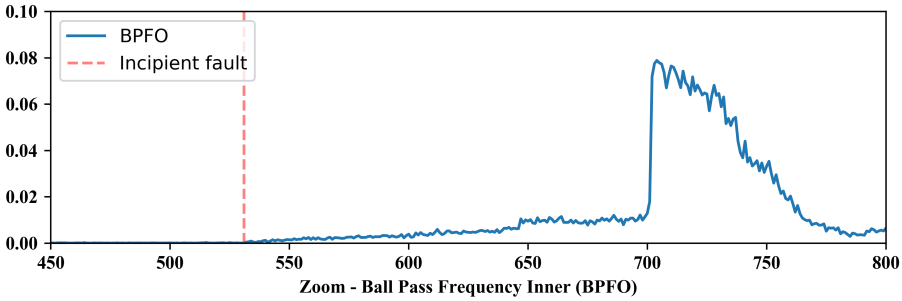


Figure 2.8: Zoom of BPFO features.

2.5.4 *Trend Analysis: Extracted Features with Reduced Dimension*

The results obtained for each method (PCA—97.45%, t-SNE—96.67%, ISOMAP—87.14%, ICA—97.39 % and AE—81.32%) can be compared with the behavior of its principal components, Figure 2.9. With the best result among the dimensionality reduction techniques studied, PCA presents significant variations in the first principal component in the anomaly region, and a slight inclination from sample 531 that can be observed by zooming in on the second principal component. This inclination is responsible for helping the Isolation Forest to identify the incipient fault, considering that it is not present in the first main component. The same behavior for the principal components can be noticed in the ICA method.

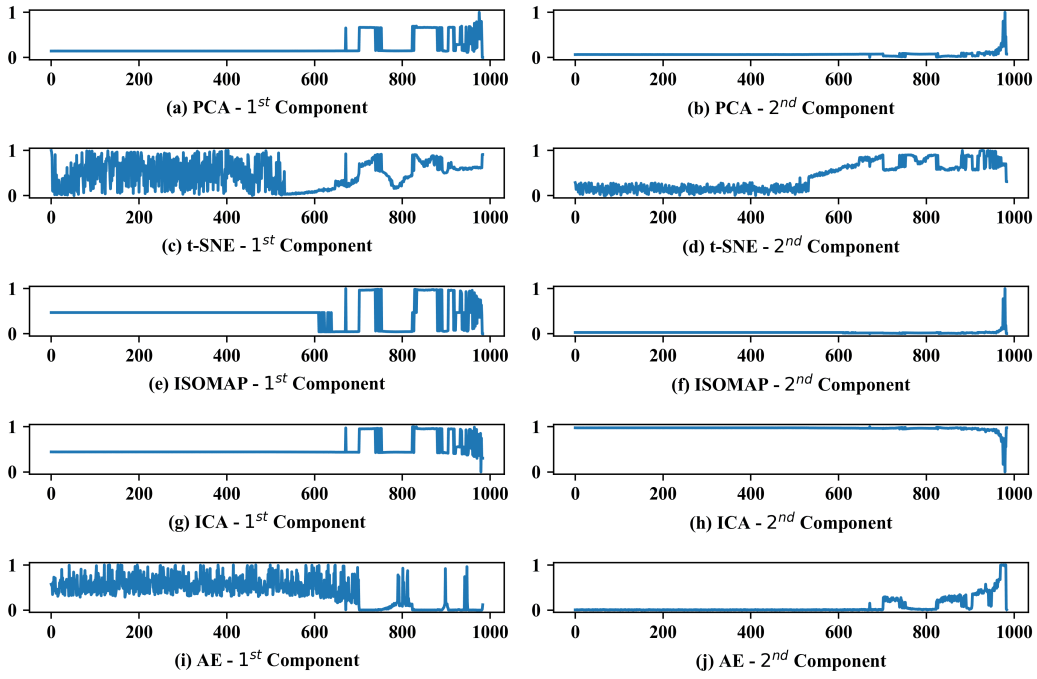


Figure 2.9: Reduced dimensionality of extracted features using different methods.

t-SNE showed a slope in the trend for the two principal components since the

incipient fault, resulting in a high hit rate, despite the noisy behavior in the first component. Similar to t-SNE, AE showed great variation where the samples were in a normal state, and it was not able to identify incipient faults which justify the lower result. The same problem occurred using ISOMAP, which was not able to identify the incipient faults, considering that the principal components showed variations, approximately, only after the 600 sample.

2.5.5 Dimensionality Reduction in the Raw Signal

The purpose of the analysis (test iv) is to verify the possibility of reducing the dimension of the raw signal, and using the principal components in an anomaly detection algorithm, in order to avoid the need to extract features beforehand.

As the PCA was the technique with the best result among the studied methods, it was used. The initial signal with 20,480 points was reduced to 2, 300 and 800 principal components were chosen based on the number of samples available, 2 for comparison with the analyses performed previously, 300, the quantity with the best result and 800, a value close to the maximum acceptable by the method. The F1-Score results with the respective standard deviation (in percentage) obtained for PCA using the reduced features of the raw signal are presented in Figure 2.10.

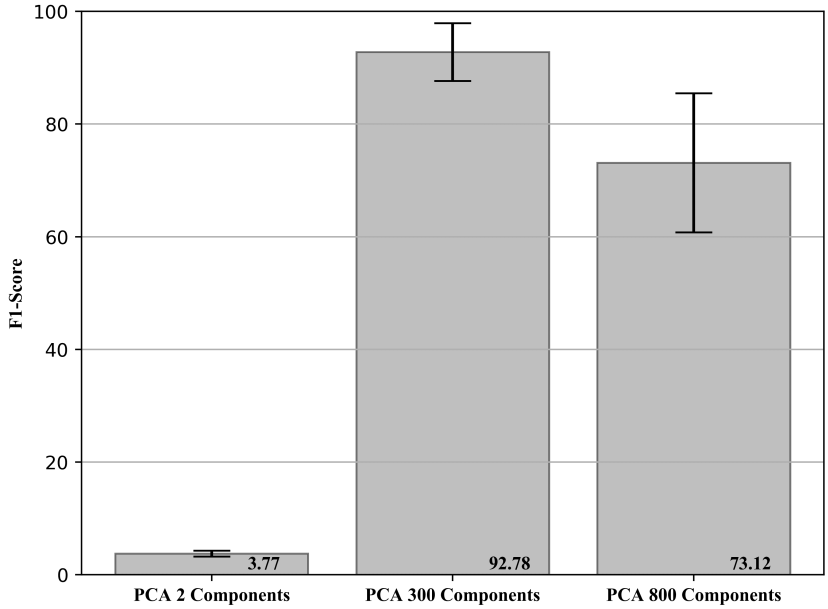


Figure 2.10: F1-Score and standard deviation (in percentage) obtained for PCA using the reduced features of the raw signal.

The results obtained show that it is possible to reduce the dimension of a raw

signal through the PCA method and obtain similar results using previously extracted features (going from 20,480 points to 300 components). On the other hand, the number of features needed to represent the problem well is high when compared with the extracted features (test ii), which leads to an increase in the computational cost, and makes it impossible to use certain machine learning algorithms that do not present good results with high dimensional space.

When using only two principal components, a variance explained of 6.86% was obtained while for two components using the extracted features (test i and ii), 100% was obtained. For 300 and 800 principal components, a variance explained of 76.92 and 96.36% was obtained. The tests using the proposed anomaly detection methodology showed results similar to the features extracted with only 300 principal components, and a significant reduction in the metrics for the other quantities studied. This fact was expected, since, as it is a raw signal, using only a few principal components is not able to represent the signal under analysis well. On the other hand, a high amount of components introduces irrelevant correlations and features to ML models, reducing their efficiency and robustness (increase in the standard deviation).

It can be concluded that, for the studied data, it is possible to reduce the raw signal size and obtain good results in the anomaly detection using Isolation Forest. However, on the other hand, the final dimension obtained is still extremely high when compared with the dimension obtained based on the features previously extracted, making the model less efficient and less robust. The trend analysis of the features, Figure 2.11 (It is noteworthy that all components are represented overlaid due to the quantity in each analysis), confirms that only two principal components are not able to represent the variations shown above. On the other hand, with the increase of the main components, a greater amount of variations can be noticed, contributing to the increase in the efficiency of the model, which results in a trade-off.

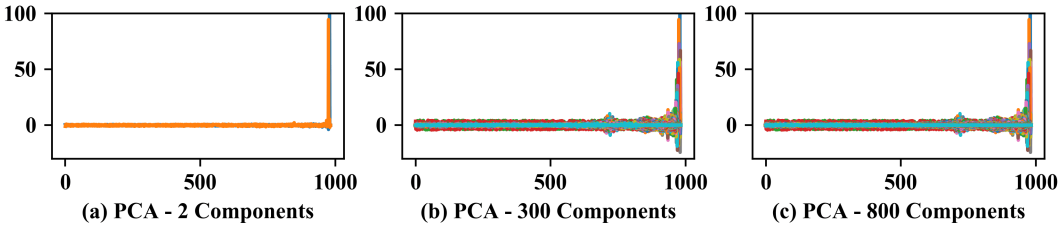


Figure 2.11: Reduced dimensionality of raw signal using PCA with a different quantity of principal components.

2.6 Conclusions

This chapter presents a comparison of different features extracted from the vibration signal and dimensionality reduction techniques, in the unsupervised detection of fault in rotating machinery, especially bearings (anomaly detection). A framework composed of five main steps is proposed, namely: (1) Data Acquisition; (2) Feature Extraction; (3) Dimensionality Reduction; (4) Fault detection: Anomaly Detection; and (5) Feature Trend Analysis.

The results show that it is possible to use vibration signals for unsupervised fault detection in rotating machinery, especially bearings. Furthermore, the feature extraction process is fundamental for the success of ML models. Techniques to reduce the size of input features have been proposed, showing their feasibility. Among the studied techniques, PCA had the best performance, being second only to the manual selection of features based on the expert's knowledge.

The raw signal can also be reduced in size, and later used in ML models for fault detection. It is noteworthy, however, that the amount of principal components needed to represent the problem tends to be greater, which can lead to a reduction in the robustness and assertiveness of the system. Feature trend analysis is an interesting tool to visually verify variations in the system, and can be used by the specialist in conjunction with the anomaly score obtained by the model for monitoring.

The work contributes to the development of monitoring of rotating machinery through machine learning, avoiding the introduction of irrelevant or correlated features in ML models, reducing data storage space and computational time to train the models. This, in addition to allowing the monitoring of anomalies using artificial intelligence, allows the

specialist to monitor the features in a summarized manner, which is not always possible when there are many monitoring variables. Thus, decision-making is supported by solid indicators, essential for application in an industrial scenario, a concept that is further explored in the new area of study in artificial intelligence, called Explainable Artificial Intelligence (XAI). New studies will focus on understanding how the model relates input features with significance to perform anomaly detection, and thus be able to work optimized on dimensionality reduction.

CHAPTER III

BAND RELEVANCE FACTOR (BRF): A NOVEL AUTOMATIC FREQUENCY BAND SELECTION METHOD BASED ON VIBRATION ANALYSIS FOR ROTATING MACHINERY

The monitoring of rotating machinery has now become mandatory in the industry, given the high criticality in production processes. Extracting relevant information from the signals is the key point for the success of the activity. Studies in the areas of Informative Frequency Band selection (IFB) and Feature Extraction/Selection have demonstrated effective approaches to extracting information. However, in general, it focuses on identifying bands where impulsive excitations are present or analyzing the relevance of the features after its signal extraction, still counting on the lack of process automation. Just as other frequencies relevant to a vibration analysis of a rotating machinery are not identified, features can be extracted from irrelevant bands, inserting unnecessary information into the analysis. To overcome this problem, the study proposes a new approach called BRF (Band Relevance Factor). The approach aims to perform an automatic selection of all relevant frequency bands for a vibration analysis of a rotating machine based on spectral entropy. The results are presented through a relevance ranking and can be visually analyzed through a heatmap. The efficiency of the approach is validated in a synthetically created dataset and two real dataset, showing that the BRF is able to identify the bands that present relevant information for the analysis of rotating machinery.

3.1 Introduction

Rotating machinery is one of the most widely used mechanical equipment of modern industry Li, Wang, Liu, Liang and Si (2018). Due to its criticality in the processes,

several techniques are used to monitor the integrity of each asset, among which the vibration analysis stands out Brito et al. (2022).

The vibration signal is composed of characteristics that refer to the behavior of the system under analysis Schmidt et al. (2021). Vibration signals from rotating machinery usually present frequency bands where the dynamic behavior of the equipment and certain faults can be analyzed. Determining the regions that present relevant information is essential to assist the specialist in identifying defects, evaluating the quality of the collected signal, allowing the extraction of significant features for use in statistical and machine learning (ML) models, or even helping to explain the artificial intelligence models, an area of recent interest called: Explainable Artificial Intelligence (XAI).

Despite the amount of information that can be obtained by analyzing the signals, the collected vibration signals of rotating machinery are usually weakened and disturbed by the strong environment noises and other neighboring components Li, Wang, Liu, Liang and Si (2018). This makes it necessary to develop methods to assist in extracting relevant information.

Two major areas stand out in the study of identifying relevant information in signals: i) Feature Extraction/Selection ii) Informative Frequency Band selection (IFB). Feature extraction/selection is one of the fundamental areas of study for signal analysis and applications of statistical and machine learning models. Performing the extraction and selection of relevant features allows knowing the signal under study and avoiding the introduction of non-significant or redundant features that tend to reduce the assertiveness of the model Brito et al. (2021). For the study of rotating machinery, as shown in Brito et al. (2022), the features to detect faults and to analyze the dynamic behavior of rotating machinery using vibration signals are commonly extracted from time, frequency and time-frequency domain.

Correctly identifying the relevant frequency bands for feature extraction/selection is critical to successful application. When features are extracted directly from the raw signal without a previous relevance analysis, there is a high chance of introducing a lot of irrelevant information into the analysis. As presented by Liu et al. (2022) if the extracted features are good enough, any classification algorithm may perform very well. Unfortunately, it is a

really challenging task that is almost impossible to fulfill in the real application.

Studies have been developed, and can be divided into three categories, i.e., filters, wrappers, and embedded methods. Yan and Jia (2019) propose a fault diagnosis strategy based on improved multiscale dispersion entropy (IMDE) and max-relevance min-redundancy (mRMR). The analysis show that the proposed method can extract effectively fault information. Zhang et al. (2018) use the two-stage feature selection, where Relief is applied for preliminarily selection and in the reselection process, Binary Particle Swarm Optimization (BPSO). Wang, Zhao, Zhang and Ning (2021) proposes an end-to-end feature selection and diagnosis method for rotating machinery combining with dimensionality reduction. Li, Yang, Li, Xu and Huang (2017) analyzed the multi-scale symbolic dynamic entropy and mRMR feature selection for fault detection in planetary gearboxes. Other works can be found at Bolón Canedo et al. (2013), where a general review on the topic is presented. Also in Liu et al. (2022) where an analysis of the importance of extracting features is performed.

In most of the works presented, the main focus is to obtain relevant information from the extraction of features, which, combined with artificial intelligence methods, are capable of diagnosing the operating condition of the equipment. Despite the good results presented, it is noted that the choice of frequency bands still needs to be automated, since in general, the methods use the raw signal or equally divided bands, without a prior analysis of relevance. Or, when the relevance is analyzed, it is related to the feature already extracted from the signal. In engineering scenarios, however, the machine users would like an automatic method to shorten the maintenance cycle and improve the diagnosis accuracy Lei et al. (2020).

Another area of study is called: Informative Frequency Band Obuchowski et al. (2014) (some other terminologies are: Frequency Band Selection (FBS) Liu, Jin, Zuo and Peng (2019), Optimal Band Selection Barszcz and JabŁonski (2011); Cui et al. (2019). Unlike feature extraction, the area aims to study mainly methods to distinguish regions where impulsive excitations occur caused by faults such as bearings and gear defects, given the difficulty of identifying the bands excited by impulses, in noisy signals, or in incipient defects.

Miao et al. (2022) present a practical framework of Gini Index (GI) in the

application of rotating machinery, comparing state-of-the-art and new methods, such as spectral kurtosis-based methods, decomposition methods, deconvolution methods: Kurtogram Antoni (2006, 2007), Protrugram Barszcz and JabŁonski (2011), Autogram Moshrefzadeh and Fasana (2018), GI derivations. At the end of the tests, using the Envelope spectrum kurtosis-based method (Protrugram) it was not possible to analyze the impulsive excitations in the presence of harmonic interference. As for the Kurtosis-based method (minimum entropy deconvolution (MED) Endo and Randall (2007) and Kurtogram) it was also not possible to perform analysis in the presence of random impulse. Hebda-Sobkowicz et al. (2020) present a novel approach to detecting cyclic impulses in the presence of non-cyclic impulses, using conditional variance based (CVB) statistic/selector and compare the result to the state-of-arts methods: spectral kurtosis l2/l1 norm Wang (2018*b*), Alpha selector Zak et al. (2015), kurtogram, spectral Gini index Miao et al. (2017), spectral smoothness index Bozchalooi and Liang (2007); Wang (2018*a*), and the infogram Antoni (2016). In summary, for a signal with only Gaussian noise, no technique indicated any band, as expected. For the cyclic and non-cyclic impulsive signal with Gaussian noise background, all methods showed good results. For the case in which the presence of the non-cyclic impulsive signal was much greater in relation to the cyclic impulse, the CVB selector was able to identify the frequency band corresponding to the cyclic impulses.

As can be seen, IFB works are focused on determining frequency regions where impulsive excitations are present, so that the signal can be filtered and a signal analysis technique such as envelope applied. The studies are extremely relevant, in view of the great difficulty in identifying such frequency bands, due to the low amplitudes presented when the defect is in the incipient stage. On the other hand, other frequencies present in the signal are also related to faults and/or dynamic behavior of the machine, which, because they present behavior different from the impulsive (eg, cyclic/harmonic/random) are not addressed by the aforementioned methods. Such frequency bands are fundamental for a vibration analysis performed by a human expert, and also for the automatic extraction of features for AI models. This is because it contains information regarding the dynamic behavior of the machine (e.g., rotation frequency), and various faults such as: unbalance, mechanical backlash, misalignment, etc.

Because most of the time, they are evident in the signal, due to their greater amplitude, such frequencies are not the focus of the works on IFB. However, aiming to support the human expert in decision making and Artificial Intelligence (AI) frameworks in the automatic extraction of relevant features and XAI, developing a method that can automatically identify such relevant frequency bands in the signal, contributes to the monitoring studies of rotating machinery and possible industrial applications.

Thus, the study proposes a new approach called BRF (Band Relevance Factor). The approach aims to perform an automatic selection of all relevant frequency bands for a vibration analysis of a rotating machinery based on spectral entropy.

Initially, the method allows the automatic evaluation of the quality of the collected signal, assessing whether it only presents noise, or if there is the presence of characteristic frequencies of the equipment. Furthermore, the method automatically points the vibration analyst to all relevant frequency bands in the signal. Finally, the use of the method as a pre-feature extraction stage allows the extraction of features in automatically selected relevant bands, for use in statistical and machine learning models.

Because it has a strong relationship with the behavior of rotating machinery, since the vast majority of faults are related to harmonic components and their multiples, spectral entropy was used. Due to its behavior, a signal from a healthy machine will show larger entropy value due to its high irregularity, while a faulted machine will have low entropy due to its low irregularity caused by the localized damage Wang et al. (2016); Li, Yang, Li, Xu and Huang (2017), which allows mapping the most relevant bands. In addition to entropy, the approach combines the use of the root-mean-square value (rms), a parameter widely used in vibration analysis of rotating machinery, as it provides an overview of the energy present in the signal.

In summary, the main contributions of this chapter are: i) A novel framework is proposed for automatic identification of relevant frequency bands in vibration signals; ii) A relevance ranking is proposed, to define among the selected bands the most important ones; iii) The heatmap is proposed to facilitate the visual analysis of the relevance ranking; iv) Possibility of applying the method in different rotating machinery and faults; v) Industrial application.

Due to the main characteristics, the work can be considered a contribution to two major areas: feature extraction/selection and IFB. Initially it shares the main objective of feature extraction, which is to extract relevant information from the signal, adding an automatic framework. On the other hand, it uses informative frequency band analysis more broadly (not focused only on impulsive defects), to automatically determine the relevant frequency bands in the signals. Such bands can be evaluated by the specialist during the analysis, or selected by an ML framework to evaluate the dynamic behavior of rotating machinery.

The remainder of this chapter starts with a brief explanation about Entropy. The proposed method is presented in Section 3. Experimental procedure is shown in Section 4. Results and discussion are given in Section 5. Finally, Section 6 concludes this chapter.

3.2 Background

3.2.1 Entropy

Entropy, as a statistical measure, is able to quantify and detect changes in time series taking into account their nonlinear behavior Li, Wang, Liu, Liang and Si (2018). Due to its characteristics, recently several entropy-based methods have been studied for rotating machinery. Li, Li, Yang, Liang and Xu (2018); Zheng et al. (2018) proposed the multi-scale permutation entropy for fault diagnosis in planetary gearboxes and rolling bearing. Zheng et al. (2017) studied a fuzzy entropy approach for fault detection in bearing. In general, studies show promising results for the use of techniques in monitoring rotating machinery. Cheng et al. (2016); Luo et al. (2016); Ai et al. (2017) presented studies involving Spectral Entropy in the detection of faults in rotating machines. Other studies can be found in Li, Wang, Liu, Liang and Si (2018); Liu., Zhi, Zhang, Guo, Peng and Liu (2019).

The definition of entropy is proposed by Shannon to evaluate the irregularity and self-similarity of time series in information theory Shannon (1948). For the discrete data series {*x1, x2,..., xn*}, the Shannon entropy H is defined as follows in Equation 3.1.

$$H(x) = -\sum_{i=1}^{n} p(x_i) log(p(x_i)) \tag{3.1}$$

where p represents the probability of the time series $\{x_{\mathrm{i}}\}$

A bigger entropy indicates a more uncertainty or irregularity of time series and if a probability can be divided into the sum of several individual values, so does the Shannon entropy Li, Wang, Liu, Liang and Si (2018). For a given time series, if the probability values of different states are similar, it is difficult to determine the future status, thus the time series has its maximum entropy value. In contrast, if there is only one state, the time series has its minimum entropy Rostaghi and Azami (2016).

Among the different methods of using Shannon's Entropy proposed in the fault diagnosis of rotating machinery Li, Wang, Liu, Liang and Si (2018), Spectral Entropy stands out. The spectral entropy of a signal is a measure of its spectral distribution, and it is a normalized form of Shannon entropy, calculated in Equation 3.2.

$$S(x) = -\frac{\sum_{i=1}^{n} P_x log(P_x)}{log(n)} \tag{3.2}$$

where, P_{x} is the probability distribution and n is the total frequency points used to normalized between 0 and 1. P_{x} is described in Equation 3.3.

$$P_x = \frac{E_i}{\sum_i^n E_i} \tag{3.3}$$

E_{i} is the energy in each frequency i.

When the distribution of values is flat with equal or close amplitudes for each frequency component, Spectral entropy will result in high values, close to 1. Distribution similar to a signal with only noise. On the other hand, if the amplitudes are concentrated in a few frequency components, especially if only a few frequencies have non-zero amplitudes, the spectral entropy value will be close to 0 Pan et al. (2008); Liu et al. (2013).

It is known that the vast majority of faults in rotating machines are related to harmonic components and their multiples. For this reason, it is possible to use entropy to analyze the signals, knowing that the vibration signal collected from a healthy machine has a larger entropy value due to its high irregularity, while that collected from a faulty rotating machinery has a smaller entropy value due to its low irregularity caused by the localized

damage Wang et al. (2016); Li, Yang, Li, Xu and Huang (2017).

3.3 Proposed approach

The proposed approach is presented in Fig. 3.1. Initially the original signal is transformed from the time domain to frequency domain and its entropy calculated. Signals with entropy greater than or equal to -3 decibels (dB) are considered to be composed only of noise and, therefore, irrelevant for the analysis. On the other hand, signals with entropy less than -3 dB are considered relevant, and likely to identify important frequency bands.

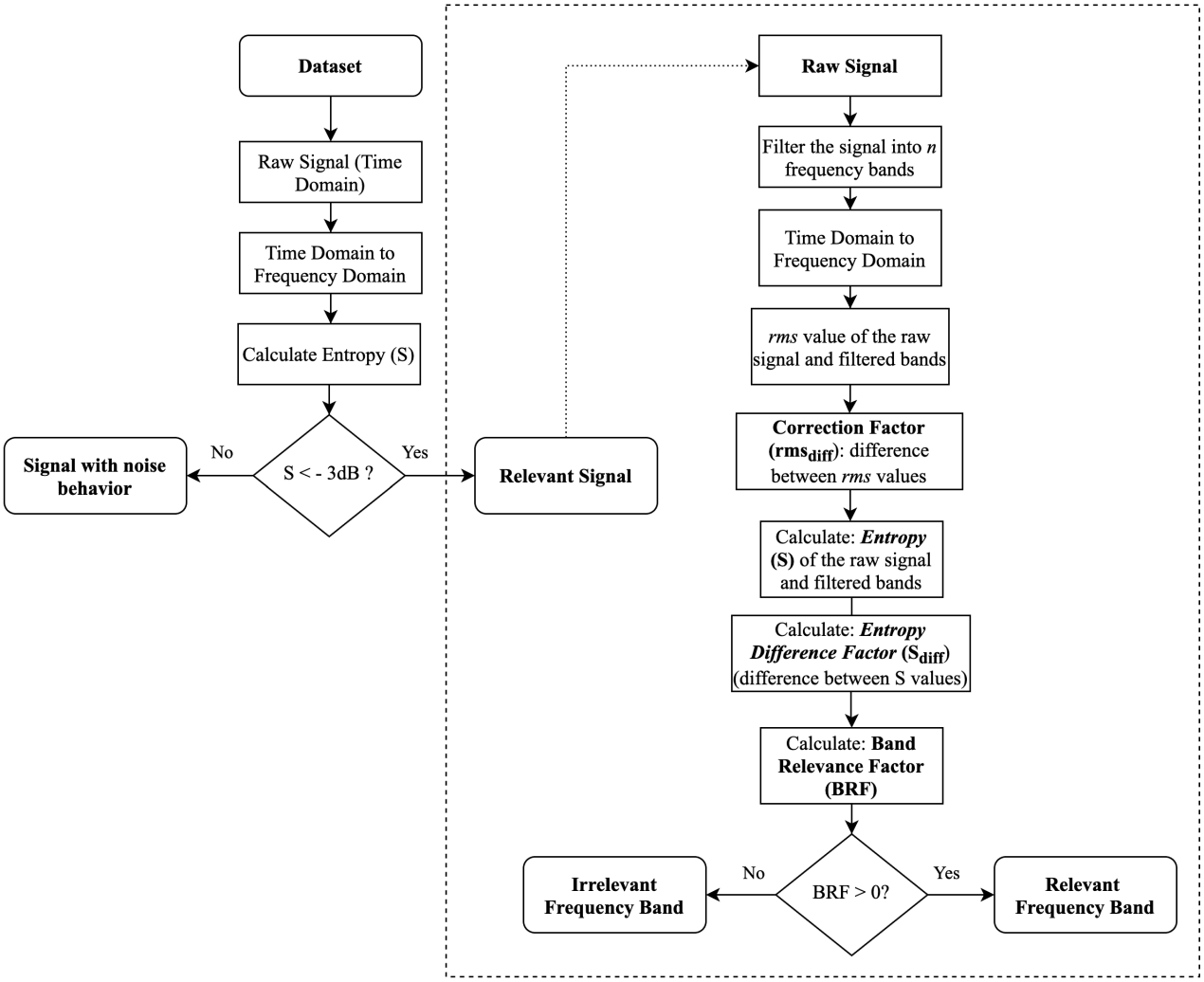


Figure 3.1: General framework of the proposed methodology.

After being identified as relevant, the original signal is filtered in k-level (bands of interest), in order to verify which bands are relevant in relation to the complete signal (being the amount of level defined by the user). Subsequently, the signal is transformed from time to frequency domain. The rms value of the filtered and complete signal is calculated,

and the difference between them is called the Correction Factor (rms_{diff}). The entropy of the complete signal and each band is calculated (S). The difference between the entropy of the complete signal and each band is calculated, called the Entropy Difference Factor (S_{diff}). Finally, the division between the Entropy Difference Factor and the Correction Factor is calculated to obtain the new proposed feature, called Band Relevance Factor (BRF). Positive values indicate that the band under analysis is relevant to the original signal, and negative values that the band is irrelevant to the original signal.

For the methodology, a so-called relevant signal is a signal that presents deterministic frequencies distinguishable from the background noise, without the need to perform advanced signal processing to remove it. Likewise, relevant frequency bands are regions that present relevant information for analysis and that can be identified through a spectral analysis, not involving advanced processing techniques. In addition, they allow the specialist to carry out an analysis regarding its characteristic, in order to understand the dynamic behavior of a system and/or identify related faults.

3.3.1 Raw Signal Entropy

The first step of the methodology consists of calculating the entropy of the raw (original) signal in order to measure the irregularity of the system. If the entropy value in dB is greater than or equal to -3 dB, the signal is considered noise. In this case no band is relevant in relation to the others, and therefore the following procedure does not need to be applied and the signal is not analyzed.

The -3 dB point is commonly used with filters. It indicates the frequency at which the associated power drops to half of its 50% maximum value. A regular/noise (equiprobable) signal in the frequency domain tends to have roughly the same amplitude value at all frequencies. Therefore, when calculating the value of Spectral Entropy (S), regular signals (equiprobable) will have S equal to or close to 1, and ordered signals (harmonics) will have S less than 1 and closer to 0. Analyzing the entropy value (assuming S = 1) for regular dB signals, it is known that S = $10log_{10}(1)$ is equal to 0.

Thus, the value for regular/noise signals (equiprobable) analyzed on the dB scale will present values of entropy zero or close to zero. Assuming that some signals may be

combinations of harmonics and noise, it is necessary to define a maximum entropy value from which it is possible to analyze the signals for fault detection (i.e., the signals are not purely noise). Assuming that 3 dB means a gain or reduction of 50% of power in the associated frequency, and approximately 0 dB being the entropy value for a regular signal/noise, it is defined that for the entropy value of - 3dB, the system is proportionately more ordered, with less uncertainty and less variability. Being possible to separate the deterministic frequencies from the noise, and consequently perform the analysis of the behavior of the machine, and identify faults.

Therefore, if the S of the original signal is less than -3 dB, there is the presence of relevant frequencies in the signal, which may be related to some fault characteristic or the dynamic behavior of the equipment, and therefore it is possible to continue the analysis. For higher S values the signal is considered irrelevant.

3.3.2 Correction Factor

After analyzing the relevance of the complete signal, filtering the signal in k-level (bands of interest) and transforming it to the frequency domain, the Correction Factor, proposed as a weighting in the BRF, is calculated in Equation 3.4.

$$rms_{diff} = rms_{base} - rms_{filtered} \tag{3.4}$$

where rms_{base} is a statistical measure defined as the root mean square value of the original signal and rms_{filtered} the root mean square value of the filtered band, both in dB. rms measures the overall value of energy present in the signal, the greater the magnitude, the greater the rms value from the global point of view.

A signal with many spectral components (e.g., excitation of resonances in bearings, mechanical looseness, cavitation etc.), will result in a larger entropy value due to its high irregularity, which may have amplitudes low or high. Consequently, if the BRF used only entropy as an analysis parameter, the ranking of the most relevant bands in the signal could present an error, since the amplitude would not be directly taken into account. Therefore, the Correction Factor is proposed to quantify the importance of the amplitude present in

the frequency band through the difference between the *rms* value of the original and filtered signal.

3.3.3 Entropy Difference Factor

The Entropy Difference Factor is proposed to evaluate the order, uncertainty and variability of the signal, calculated in Equation 3.5.

$$S_{diff} = 3 + S_{base} - S_{filtered} \tag{3.5}$$

To provide decision making based on the S_{diff} (+ or -), 3 dB was added to the S_{diff} value. This allows to replace the decision threshold which was previously -3 dB to 0 dB. Thus, if the S_{diff} difference is positive or equal to zero, the band under analysis is relevant, otherwise the band is not relevant to the analysis.

In the new scale, the value for regular/noise bands (equiprobable) analyzed in dB will present positive entropy values. Considering that the signal was approved in the first test, the value of S_{base} has a negative value, so when applied to the equation we will have: (-) - (+) = -, therefore irrelevant for the signal.

In a signal with only one harmonic, the S_{filtered} value of the band in which the harmonic is present will be less than or approximately the same as the full signal, resulting in a S_{diff} equal to or approximately zero. In a signal with harmonic and noise, when filtering in the harmonic region, entropy tends to reduce, as it is a more ordered region (since the harmonic is present), and less uncertainty, due to the noise limitation, in this way, the S_{filtered} entropy will be lower, ensuring a positive value.

3.3.4 Band Relevance Factor (BRF)

Although entropy quantifies the regularity of the signal, that is, the presence of harmonics or random frequencies, not all component failures present harmonic behavior, which can lead to an error in the identification ranking of the relevance of the band if only the Entropy Difference Factor is used. To overcome this problem, the Entropy Difference Factor is divided by the Correction Factor, resulting in the Band Relevance Factor (BRF).

The more energy in the band, the lower the value of the Correction Factor, and

consequently the higher the value of the BRF, since it is in the denominator. In this way, it is possible to determine, among the selected bands, which is the one with the greatest relevance.

The Band Relevance Factor is obtained through the Equation 3.6.

$$BRF = \frac{S_{diff}}{rms_{diff}} \tag{3.6}$$

Being, S_{diff} the Entropy Difference Factor and rms_{diff} the Correction Factor.

As shown, positive values indicate that the band under analysis presents relevant information for analysis. In addition, the higher the BRF value, the more information that band presents for analysis, thus making it possible not only to define whether the band is relevant or not, but also to obtain a relevance ranking.

3.3.5 Heatmap and Relevance Ranking

Through BRF it is possible to obtain a ranking with the most relevant frequency bands in each analyzed k-level. To facilitate the visualization of the frequency bands, a heatmap is proposed, where the values are normalized on a scale from 1 to -1, with only positive values being relevant.

Each level represents the division of the signal into 2^{k}, where k is the level and 2^{k} is the amount of bands present in each level. Example: a complete signal with a frequency range from 0 to 10240 hz, for level 0, the bandwidth (BW) will be from 0 to 10240 hz (2^0), that is, the complete signal. For level 1, the signal will be divided into two bands (2^1), 0 to 5120 hz and 5120 to 10240 hz and so on for the other levels.

It is worth mentioning that the analyzes must be performed per level, that is, each level will present a relevance ranking, since the calculations are performed separately, and only combined to obtain all the results in a single graph.

The same happens for the *rms* value used for comparison, however the values are normalized from 0 to 1, as there are no negative *rms* values in the frequency signal.

3.4 Experimental procedure

3.4.1 Data description

Three dataset were used, one synthetically created to exemplify the proposed approach (Case 1), one publicly available (Case 2 - Bearing dataset) and one developed by the author for the study (Case 3 - Mechanical faults dataset). The dataset were chosen because they approach different faults present in the components of rotating machinery, in addition to presenting the differences between normal and fault conditions for each rotating machinery. The use of different dataset allows validating the proposed methodology under different conditions, and exemplifying the strengths and limitations.

Since the bearing dataset (Case 2) is publicly available and already explored in Brito et al. (2022), it will be commented shortly. For more details, please refer to Qiu et al. (2006); Brito et al. (2022). All real dataset are vibration signals collected using accelerometers.

Case 1: Synthetic dataset

The proposed methodology aims to evaluate, mainly, the frequency bands that have relevance in the signal, thus allowing the analysis by the expert, or extraction of features for use in machine learning models. In this way, a synthetic dataset was generated, aiming to replicate real conditions in rotating machinery.

The proposed signal is an oscillatory signal whose waveform is given in the time domain by Equation 3.7.

$$x(t) = A\,sin(2\pi\,ft + \theta\,) \tag{3.7}$$

where A indicates the peak amplitude of the signal, f frequency [Hz], t time vector [s] and θ phase [rad], because monoaxial accelerometers without tachometer were used, the phase reference was not considered, although it could help in the identification of some defects. t = 1s and fs = 20480 hz were used.

The harmonic signal was chosen to compose the dataset, due to its similarity with characteristic defects in rotating machinery, such as the force produced by the unbalance of

a rotor. Furthermore, the harmonic signal represents the fundamental rotation frequency of the machine (1x), which is one of the most relevant parameters for the analysis of rotating machinery signals.

In real signals, in addition to the influence of other frequencies related to nearby machines, external excitations etc., some vibrations/defects excite more than one frequency in the signal, such as mechanical looseness, misalignment, gear defects. To simulate a condition closer to reality, the synthetic signal was composed of $n$different frequencies, Equation 3.8.

$$x_t(t) = A_1\, sin(2\pi\, f_1 t + \theta\,) + A_2\, sin(2\pi\, f_2 t + \theta\,) + ... + A_n\, sin(2\pi\, f_n t + \theta\,) \tag{3.8}$$

where $x_t(t)$ is a sum of harmonic signals $x(t)$ Equation 3.7, of different frequencies, f_1, f_2,..., f_{n}. The frequencies were defined in such a way that when performing the division by bands, some bands contained more than one frequency, others only one and others none, namely: 30, 120, 500, 700, 750, 2300, 2450, 2600, 2700, 2800 and 3450 hz. Amplitudes were randomly determined for a range of 0 - 1.

In real applications, every signal will present a noise level, and for this reason, it was added to the synthetic signal, Equation 3.9.

$$x_s = x_t + \alpha_{\mathrm{gauss}} G,\; G \sim N(0,1) \tag{3.9}$$

where $\alpha_{\mathrm{gauss}} > 0$ is the Gaussian noise coefficient, x_{s} is the final synthetic signal, x_{t} is the raw signal and G the gaussian noise.

As the proposed method aims to identify relevant frequency bands, and knowing that they can be influenced by the noise present in the signal, four noise levels were added.

As it is a synthetic and known signal, the different noise levels were obtained based on the signal-to-noise ratio (SNR). SNR is the ratio between the desired information or the power of a signal and the undesired signal or the power of the background noise, and its unit of expression is typically decibels (dB). Equation 3.10.

$$SNR = 10\log_{10}(\frac{P_{\text{signal}}}{P_{\text{noise}}}) \tag{3.10}$$

where, P_{signal} and P_{noise} are the *rms* value in volts for the signal and noise respectively.

Observing Equation 3.10, note that SNR greater than 1:1 (greater than 0 dB) indicates more signal than noise. Therefore, the SNR levels were defined as: 24, 12, 6 and 0 dB. The SNR relationships were also designed in such a way to facilitate the visualization of the signals in the frequency domain, allowing for a better validation of the methodology, with the proposal being:

i) 24 dB (SNR): the harmonic frequencies present in the signal are easily evidenced in spectral analysis, with a small noise.

ii) 12 dB (SNR): harmonic frequencies continue to be visible in a spectral analysis, but with an increase in background noise, which does not preclude a correct identification.

iii) 6 dB (SNR): harmonic frequencies of greater amplitude are identifiable in a conventional spectral analysis, and may have some frequencies close to noise, which would raise doubts as to whether it is noise or harmonic frequency.

iv) 0 dB (SNR): harmonic frequencies and noise are completely mixed in a conventional spectral analysis, not allowing the identification of machine faults.

To facilitate the description in the text, the noises will be called low, medium, high and mixed, referring to SNR 24, 12, 6 and 0 dB, respectively.

Case 2: Bearing dataset

The dataset Qiu et al. (2006) is composed of remaining useful life (RUL) test on bearings, with 03 tests performed and 4 bearings in each test. The end of life reached for each bearing occurred after 100 million revolutions which is the designed life time. Each file consists of 20,480 points with the sampling rate set at 20 kHz. The signals were manually labeled, based on knowledge about fault diagnosis using vibration analysis, and 2 signals were selected to validate the applicability of the proposed method. For the study, the bearing 01 of test 02 was used, with signal 150 representing the normal operating condition and signal 905 with fault in the outer race.

Case 3: Mechanical faults dataset

The last dataset was developed by the authors. The faults were introduced on a test bench, Fig. 5.2, being: motor, frequency inverter, bearing house, two bearings, two pulleys, belt and rotor (disc).

Figure 3.2: Bench test.

20 tests were performed, 5 for each condition: unbalance, misalignment, mechanical looseness and normal operating condition. Each test consists of 4 sets of 420 signals collected continuously. Each file consisting of 25,000 points with the sampling rate set at 25 kHz (420 signals per accelerometer). Resulting in the end of the tests, in a total of 8400 signals per accelerometer. The rotation was kept constant with a value measured on the axis of approximately 1238 rpm. The sequence of tests was randomly defined. Before starting any test, the bench was dismantled and returned to normal operating condition, to later introduce the fault. The experimental procedure allows variations to occur, making the tests closer to industrial reality.

Because it is a small bench, the measurement points are close, and in order to reduce the computational cost of the tests, only the signals from the horizontal position of the accelerometer present in the coupled side bearing (near the pulley) were used in the analyses.

In addition, only two conditions were chosen to facilitate the reading and understanding of the methodology, in view of the other dataset used, namely: normal operating condition and unbalance.

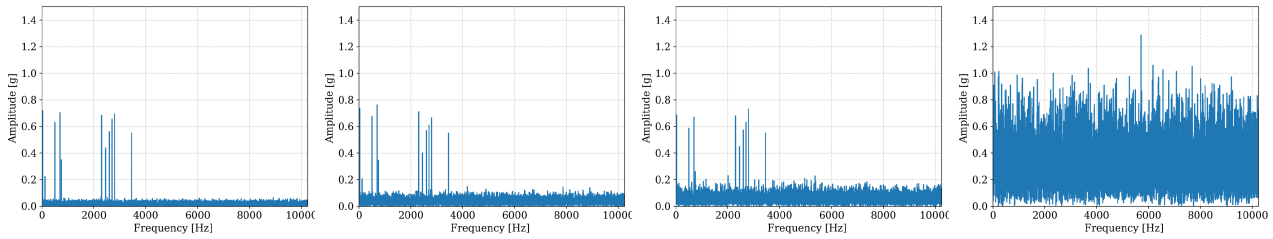
3.5 Results and discussion

3.5.1 Data exploration

In this section the synthetically generated samples and the two dataset used are analyzed and discussed. As the objective is just to visualize the signal characteristic, the signal in the frequency domain was plotted with $df = 1$ hz.

Case 1: Synthetic dataset

In Fig.3.3 the original signals for the two conditions (Normal and BPFO) and the synthetic signals for the seven created conditions (Normal, BPFO, BPFI, Unbalance, Misalignment, Looseness and Gear Fault) from Case 1 are presented.



(a) Low noise (b) Medium noise (c) High noise (d) Mixed noise

Figure 3.3: Examples of vibration signals for Case 1.

In Fig.3.3a,3.3b,3.3c it can be verified that the generated frequencies are highlighted as expected, since the SNR relationship is positive, presenting more signal than noise in the sample. On the other hand, in Fig.3.3d there is a mixture of the signal and noise not allowing to identify the frequencies of interest originally inserted.

In a conventional vibration analysis, focusing on the identification of faults (predictive maintenance), it can be concluded that Fig.3.3a,3.3b,3.3c allow a diagnosis of the signal, while Fig.3.3d would be irrelevant since it is not possible to distinguish the characteristic fault frequencies from the noise. The same can be analyzed in relation to the frequency bands relevant to the diagnosis. The bands in which the frequencies are highlighted are usually related to the excitations in the machines and, consequently, are relevant for the diagnosis of the defect.

In a visual vibration analysis, stand out the bands containing the frequencies: 30, 120, 500, 700, 750, 2300, 2450, 2600, 2700, 2800 and 3450 Hz (which were synthetically inserted in the signal). Other regions in the signals represent only noise. Thus, the

methodology must be able to identify the bands in which such frequencies are present.

Case 2: Bearing Fault

In Fig.3.4, the two randomly selected signals from the Case 2, representing normal operating condition Fig.3.4a and bearing outer race fault Fig.3.4b are presented.

(a) Normal condition (b) Outer race fault

Figure 3.4: Examples of vibration signals for Case 2.

In Fig.3.4a it can be noted that only one frequency related to the dynamic behavior of the equipment is highlighted since there is no fault in the signal. On the other hand, in Fig.3.4b other frequencies become present, indicating to the specialist a change in equipment behavior, which in turn is related to the fault. The frequencies highlighted in both cases are relevant for diagnosis and should be analyzed by the specialist to determine the current condition of the equipment. Therefore, as in the analysis, the methodology must be able to identify such frequencies as relevant.

Case 3: Mechanical faults dataset

In Fig.3.8, the two randomly selected signals from the Case 3, representing normal operating condition Fig.3.5a and unbalance Fig.3.5b are presented.

(a) Normal condition (b) Unbalance

Figure 3.5: Examples of vibration signals for Case 3.

The dataset was purposely generated to simulate severe industrial conditions, where the vibration signal has a rich amount of frequencies, even if they are not related to

equipment faults (e.g., instrumentation interference, nearby equipment etc.). Thus, even for the normal operating condition, it can be noted that in Fig.3.5a there are several highlighted frequencies. In Fig.3.5b where the unbalance was inserted, it is verified that the frequency related to the unbalance starts to dominate the signal, presenting greater relevance for the analysis (21.9 hz). Through a vibration analysis performed by a specialist, the rotation frequency will be related to the characteristic of the unbalance defect in the machine. Therefore, the methodology must be able to identify such frequency as relevant even with the presence of other frequencies in the signal.

3.5.2 BRF Analysis

In this subsection, the results obtained with the proposed methodology and *rms* value are presented. Due to the lack of space, for each dataset a signal was selected and the heatmap with the most relevant bands per level is presented. For the analysis of the other signals, the ranking with the five most relevant bands are presented.

Case 1: Synthetic dataset

To represent the synthetic dataset, the heatmap with medium noise was selected, Fig. 3.6.

It can be noted that the BRF was able to identify as relevant the bands that contain the frequencies inserted in the signal. It is also verified that the smaller the size of the band, the BRF tends to classify regions that present only noise as relevant. This occurs because, as it is limited in a small region, only one or a few dominant frequencies can occur and the entropy may assumes a value close to a system with equiprobable signal. This limitation is circumvented by analyzing the ranking with the most relevant bands determined by the method (or heatmap color scale), which points out to the harmonic frequencies inserted in the signal.

Comparing with the *rms*, the methodology was able to better distinguish the relevant bands for analysis, as can be observed in levels 1, 2, 3, 4 and 5, mainly. According to the BRF, bands with a value less than/equal to zero (scale from white to red) are irrelevant for the analysis. Analyzing the heatmap, the regions with a white to red scale are bands

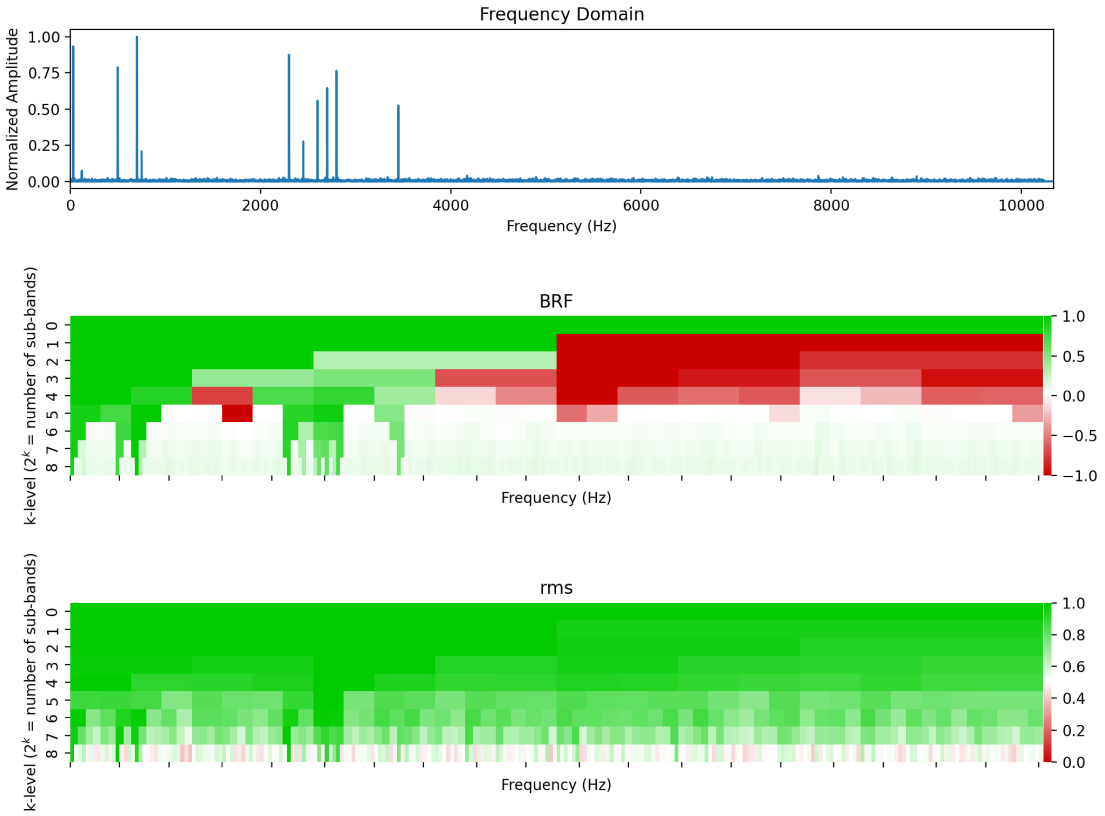


Figure 3.6: Heatmap - Case 1.

where there is only the presence of noise and no harmonic frequency. On the other hand, the same bands are classified with maximum value or close to maximum by the *rms* method.

It is worth mentioning that the analysis by level is independent, not being one level dependent on the other in the method. Then the analyst can determine the size of the band (BW) on which to perform an analysis.

The rankings with the five most relevant bands according to BRF and *rms* for the four signals from Case 1 are shown in Table 3.1.

Analyzing the signals (low, medium and high noise) it is noted that, as expected, no band above 5120 hz was selected in the top 5 of relevance of the BRF methodology. In this region only noise was inserted without the presence of any harmonic frequency, and therefore should not be selected. On the other hand, several bands above 5120 hz were selected by *rms*.

It can also be noted that in all rankings for signals low, medium and high noise,

Table 3.1 – Relevance Ranking - Case 1

Signal	Feat.	Rank/ k-level	1 (BW 5120 hz)	2 (BW 2560 hz)	3 (BW 1280 hz)	4 (BW 640hz)	5 (BW 320 hz)	6 (BW 160 hz)	7 (BW 80 hz)	8 (BW 40 hz)
		1	0:5120	0:2560	0:1280	2560:3200	2560:2880	0:160	0:80	0:40
		2		2560:5120	2560:3840	0:640	0:320	2720:2880	2800:2880	2800:2840
	BRF	**3**			1280:2560	640:1280	640:960	640:800	640:720	680:720
		4				1920:2560	2240:2560	2240:2400	2240:2320	2280:2320
Low		**5**				3200:3840	320:640	2560:2720	2640:2720	2680:2720
Noise		**1**	0:5120	0:2560	0:1280	0:640	2560:2880	2560:2720	640:720	680:720
		2	5120:10240	2560:5120	2560:3840	2560:3200	2240:2560	640:800	0:80	0:40
	RMS	**3**		5120:7680	1280:2560	1920:2560	640: 960	0:160	2640:2720	2680:2720
		4		7680:10240	8960:10240	640:1280	0:320	480:640	480:560	2280: 2320
		5			5120:6400	9600:10240	320:640	2240:2400	2240: 2320	480:520
		1	0:5120	0:2560	0:1280	0:640	640:960	640:800	640:720	680:720
		2		2560:5120	2560:3840	640:1280	2560:2880	0:160	0:80	0:40
	BRF	**3**			1280:2560	2560:3200	0:320	2240:2400	2240:2320	2280:2320
		4				1920:2560	2240:2560	480:640	480:560	480:520
Medium		**5**				3200:3840	320:640	2720:2880	2800:2880	2800:2840
Noise		**1**	0:5120	0:2560	2560:3840	0:640	2560:2880	0:160	2240: 2320	480:520
		2	5120:10240	2560:5120	0:1280	2560:3200	2240:2560	2560:2720	480:560	2280:2320
	RMS	**3**		5120:7680	1280:2560	1920:2560	320:640	2720:2880	640:720	0:40
		4		7680:10240	5120:6400	3200:3840	0:320	640:800	0:80	680:720
		5			3840:5120	5120:5760	640: 960	2240:2400	2640:2720	2680:2720
		1	0:5120	0:2560	2560:3840	2560:3200	2560:2880	2720:2880	2800:2880	2800:2840
		2		2560:5120	0:1280	0:640	2240:2560	2560:2720	0:80	0:40
	BRF	**3**			1280:2560	1920:2560	0:320	0:160	2240: 2320	2280:2320
		4				640:1280	640: 960	2240:2400	640:720	680:720
High		**5**				3200:3840	320:640	640:800	2640:2720	2680:2720
Noise		**1**	0:5120	2560:5120	2560:3840	0:640	2560:2880	2560:2720	480:560	2280:2320
		2	5120:10240	0:2560	0:1280	2560:3200	4160:4480	640:800	2560:2640	480:520
	RMS	**3**		5120:7680	3840:5120	5120:5760	0:320	3680:3840	0:80	0:40
		4		7680:10240	5120:6400	3840: 4480	5440:5760	0:160	720:800	680:720
		5			6400:7680	4480:5120	640: 960	4160:4320	4240:4320	2600:2640
		1	-	-	-	-	-	-	-	-
		2								
	BRF	**3**								
		4								
Mixed		**5**								
Noise		**1**	5120:10240	5120:7680	5120:6400	0:640	6080:6400	6240:6400	6320:6400	5680:5720
		2	0:5120	2560:5120	0:1280	2560:3200	2880:3200	4880:5040	4560:4640	4560:4600
	RMS	**3**		7680:10240	2560:3840	5760:6400	4480:4800	0:160	5680:5760	1400:1440
		4		5120:7680	8960:10240	7040:7680	960:1280	5600:5760	8480:8560	6360:6400
		5			3840:5120	5120:5760	0:320	3040:3200	3280:3360	8520:8560

Table 3.2 – Ranking comparison - Case 1

Signal	*k-level/ Type of Analysis*	0 (BW 10240 hz)	1 (BW 5120 hz)	2 (BW 2560 hz)	3 (BW 1280 hz)	4 (BW 640hz)	5 (BW 320 hz)	6 (BW 160 hz)	7 (BW 80 hz)	8 (BW 40 hz)
Low Noise	VA	100 %	50 %	50 %	60 %	80 %	100 %	80 %	80 %	80 %
	PA	100 %	50 %	50 %	60 %	0 %	60 %	0 %	0 %	20 %
Medium Noise	VA	100 %	50 %	50 %	60 %	80 %	100 %	80 %	80 %	80 %
	PA	100 %	50 %	50 %	20 %	20 %	0 %	0 %	0 %	0 %
High Noise	VA	100 %	50 %	50 %	40 %	40 %	60 %	60 %	20 %	60 %
	PA	100 %	50 %	0 %	40 %	0 %	40 %	0 %	0 %	20 %
Mixed Noise	VA	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
	PA	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %

the bands address some of the harmonic frequencies inserted in the signal: 30, 120, 500, 700, 750, 2300, 2450, 2600, 2700, 2800 and 3450 hz. What does not occur using only *rms*.

From the analysis it can also be verified that when there is a positive signal-to-noise ratio (SNR), that is, more signal than noise in the sample, even with different levels, the BRF is able to identify the relevant frequency bands of the signal.

Finally, analyzing the condition in which the SNR is equal to zero (amount of noise and signal is equal), it is verified that the methodology presents null ranking. That is, in the first analysis performed, the method identifies that it is not possible to separate (for conventional analysis) relevant bands in the signal, since it is completely mixed with noise. Using only the *rms* value, the analyst would be induced to check all bands since there is energy coming from the noise. In an automatic methodology for extracting features for an artificial intelligence method (e.g., machine learning), it would result in the extraction of irrelevant features and consequently a reduction in the model's accuracy rate.

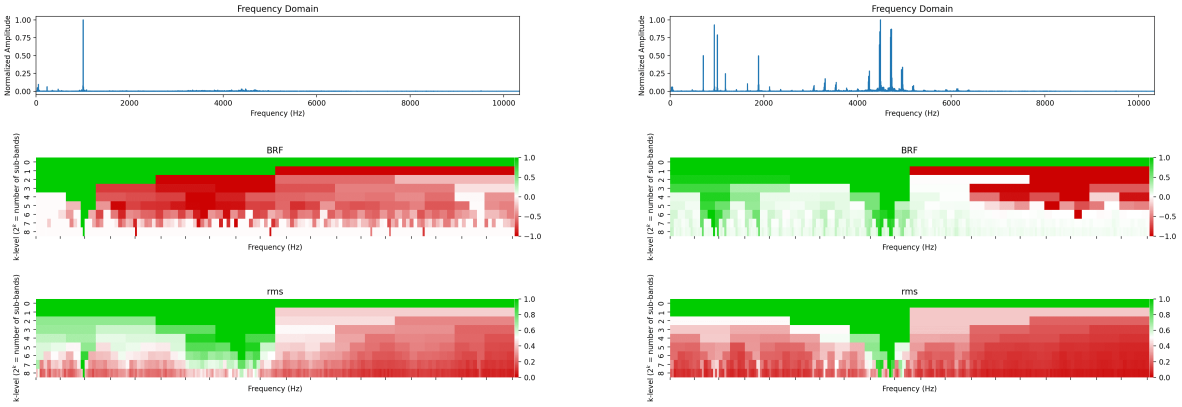
To make a comparison between the 5 most relevant bands obtained in the ranking with BRF and rms, two analyzes were performed. The first, called Values Analysis (VA), calculates the number of times that the same band appeared in the two methodologies without taking into account the order, and the second, called Position Analysis (PA), checks how many times the positions (order) of the selected bands were equal in the ranking, Table 3.2.

Analyzing Table 3.2, it can be verified that the signal where the noise is totally mixed, all the values were different, since according to the BRF, no band was considered relevant. For the other signals, it can be noted that the percentage of values that are repeated in both methodologies is similar for conditions: low and medium noise, where the amount of noise is lower, and reduces for high noise condition where the amount of noise is greater than the previous ones. This happens because the BRF is more robust in the presence of

noise, being able to better quantify the bands that are relevant for the analysis when there is higher noise. Regarding the ranking positions (order), it can be noted that the methodologies presented different sequences in most cases.

Case 2: Bearing dataset

The heatmap for the two selected signals, normal and fault condition on the outer race, are shown in Fig.3.7a and Fig.3.7b, respectively.



(a) Normal condition (b) Fault in outer race

Figure 3.7: Heatmap - Case 2.

For the normal condition, Fig.3.7a, both methodologies highlighted the band in which the frequency of approximately 1000 Hz was present as the most relevant. In the BRF, only one band was selected as relevant at all levels. However, using the *rms* value, other bands also presented energy close to the maximum value.

In Fig.3.7b where the fault is present, both methodologies identified the high frequency region (excitation of the resonances due to the impact generated by the fault) as relevant. On the other hand, it is known that with the progression of the fault, the excitations present only at high frequency begin to appear in the region of medium and low frequency, associated with the characteristics of the bearing. In this case, the BRF was able to identify not only high frequency excitations as relevant, but also excitations in the region below 2000 Hz. The *rms* value considered only the high frequency region as the most relevant. Knowing that this behavior is associated with the progression of the fault, using only the *rms* the analyst could miss an evolution of the defect, since the low and medium frequencies would not be as prominent as the high ones.

The rankings with the five most relevant bands according to BRF and *rms* for the two signals from Case 2 are shown in Table 3.3.

Table 3.3 – Relevance Ranking - Case 2

Signal	**Feat.**	**Rank/ k-level**	**1 (BW 5120 hz)**	**2 (BW 2560 hz)**	**3 (BW 1280 hz)**	**4 (BW 640hz)**	**5 (BW 320 hz)**	**6 (BW 160 hz)**	**7 (BW 80 hz)**	**8 (BW 40 hz)**
Normal	**BRF**	**1**	0:5120	0:2560	0:1280	640:1280	960:1280	960:1120	960:1040	1000:1040
		2								
		3								
		4								
		5								
	RMS	**1**	0:5120	2560:5120	3840:5120	3840: 4480	4160:4480	4320:4480	960:1040	1000:1040
		2	5120:10240	0:2560	2560:3840	3200:3840	4480:4800	960:1120	4400:4480	4400:4440
		3		5120:7680	0:1280	4480:5120	3520:3840	4640:4800	4320:4400	4360:4400
		4		7680:10240	1280:2560	640:1280	3200:3520	4480:4640	4640:4720	4680:4720
		5			5120:6400	0:640	3840:4160	3520:3680	0:80	40:80
Fault	**BRF**	**1**	0:5120	2560:5120	3840:5120	4480:5120	4480:4800	800:960	880:960	920:960
		2		0:2560	0:1280	640:1280	640:960	4640:4800	4480:4560	4480:4520
		3			1280:2560	3840: 4480	960:1280	4480:4640	4640:4720	4680:4720
		4				1280:1920	4160:4480	960:1120	4720:4800	4720:4760
		5				3200:3840	1600:1920	4320:4480	4400:4480	4440:4480
	RMS	**1**	0:5120	2560:5120	3840:5120	4480:5120	4480:4800	4640:4800	4640:4720	4680:4720
		2	5120:10240	0:2560	2560:3840	3840: 4480	4160:4480	4480:4640	4480:4560	4480:4520
		3		5120:7680	0:1280	3200:3840	4800:5120	4320:4480	4400:4480	4440:4480
		4		7680:10240	5120:6400	640:1280	3520:3840	4160:4320	4720:4800	4720:4760
		5			1280:2560	5120:5760	3200:3520	4800:4960	4880:4960	4920:4960

As in Case 1, the ranking of the BRF confirms the selection of bands related to the most relevant frequencies in the signal, associated with the fault diagnosis. For the normal situation, only one band was selected as relevant, and for the fault condition, different bands in low, medium and high frequencies were selected, according to the fault progression characteristic mentioned above. On the other hand, the *rms* value no longer points to frequencies that are directly associated with the defect present in the signal.

The comparison between the top 5 values of the relevance rankings obtained with BRF and *rms* is presented in Table 3.4.

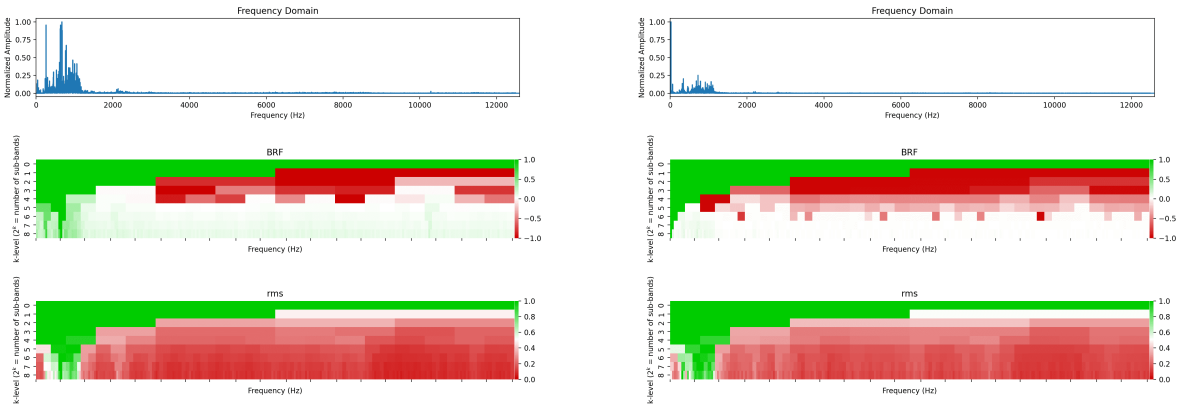
It can be noted that in the normal condition, the methodologies presented more different values, since the *rms* considered the high frequency region relevant in this situation. On the other hand, the fault situation presented some similar values (mainly related to high frequency), and the main difference is related to the low frequency region considered by BRF and not by the *rms*.

Table 3.4 – Ranking comparison - Case 2

Signal	*k-level/ Type of Analysis*	0 (BW 10240 hz)	1 (BW 5120 hz)	2 (BW 2560 hz)	3 (BW 1280 hz)	4 (BW 640hz)	5 (BW 320 hz)	6 (BW 160 hz)	7 (BW 80 hz)	8 (BW 40 hz)
Normal	VA	100 %	50 %	25 %	20 %	40 %	0 %	20 %	40 %	40 %
	PA	100 %	50 %	0 %	0 %	0 %	0 %	0 %	20 %	20 %
Fault	VA	100 %	50 %	50 %	60 %	80 %	40 %	60 %	80 %	80 %
	PA	100 %	50 %	50 %	20 %	20 %	20 %	0 %	40 %	40 %

Case 3: Mechanical fault dataset

The heatmap for the two selected signals, normal and unbalance, are shown in Fig.3.8a and Fig.3.8b, respectively.



(a) Normal condition (b) Unbalance

Figure 3.8: Heatmap - Case 3.

For the normal condition, Fig.3.8a both methodologies selected similar frequency bands, mainly at levels 5, 6 and 7. For the fault condition, Fig.3.8b, as it is an unbalance, the most relevant frequency in the signal is related to the rotation frequency, which in the case it is at 21.9 hz. Analyzing the result obtained by the BRF, it is noted that the band that contained the frequency was selected as the most relevant in each of the levels, validating the methodology. On the other hand, using the *rms*, the region presented lower or no relevance when analyzing each level.

The rankings with the five most relevant bands according to BRF and *rms* for the two signals from Case 3 are shown in Table 3.5.

The ranking of the BRF for the unbalance condition confirms the presence of the rotation frequency (associated with fault) in the selected band as the most relevant in each of the levels.

The comparison between the top 5 values of the relevance rankings obtained with BRF and *rms* is presented in Table 3.6.

It can be noted that there are differences in the values and positions selected in the top 5 of each methodology, with greater similarity for levels 5, 6 and 7 in relation to values (normal condition, mainly). As in the other Cases, the positions of the rankings present greater differences.

Table 3.5 – Relevance Ranking - Case 3

Signal	Feat.	Rank/ k-level	1 (BW 6250 hz)	2 (BW 3125 hz)	3 (BW 1562 hz)	4 (BW 781 hz)	5 (BW 390 hz)	6 (BW 195 hz)	7 (BW 97 hz)	8 (BW 48 hz)
Normal	BRF	1	0:6250	0:3125	0:1562	0:781	390:781	585:781	585:683	634:683
		2				781:1562	0:390	195:390	195:293	244:292
		3					781:1171	781:976	781:878	585:633
		4					1171:1561	976:1171	683:781	732:812
		5					10156:10546	0:195	878:976	781:829
	RMS	1	0:6250	0:3125	0:1562	0:781	781:1171	585:781	585:683	634:683
		2	6250:12500	6250:9375	1562:3124	781:1562	390:781	781:976	878:976	878:927
		3		3125:6250	6250:7812	1562:2343	0:390	976:1171	976:1074	585:633
		4		9375:12500	7812:9375	2343:3125	1171:1561	390:585	781:878	927:976
		5			3125:4687	7812:8593	1953:2343	195:390	683:781	1025:1074
Fault	BRF	1	0:6250	0:3125	0:1562	0:781	0:390	0:195	0:97	0:48
		2						195:390	683:780	683:731
		3							293:390	293:341
		4								878:927
		5								585:633
	RMS	1	0:6250	0:3125	0:1562	0:781	781:1171	585:781	683:781	293:341
		2	6250:12500	6250:9375	1562:3124	781:1562	390:781	781:976	781:878	634:683
		3		3125:6250	6250:7812	1562:2343	0:390	976:1171	976:1074	731:781
		4		9375:12500	4687: 6250	2343:3125	1171:1561	195:390	585:683	781:829
		5			7812:9375	7812:8593	1953:2343	390:585	293:390	976:1024

Table 3.6 – Ranking comparison - Case 3

Signal	*k-level/ Type of Analysis*	0 (BW 12500 hz)	1 (BW 6250 hz)	2 (BW 3125 hz)	3 (BW 1562 hz)	4 (BW 781 hz)	5 (BW 390 hz)	6 (BW 195 hz)	7 (BW 97 hz)	8 (BW 48 hz)
Normal	VA	100 %	50 %	25 %	20 %	40 %	80 %	80 %	80 %	40 %
	PA	100 %	50 %	25 %	20 %	40 %	20 %	20 %	20 %	40 %
Fault	VA	100 %	50 %	25 %	20 %	20 %	20 %	20 %	40 %	0 %
	PA	100 %	50 %	25 %	20 %	20 %	0 %	0 %	0 %	0 %

3.6 Conclusions

A new approach that allows to automatically select the relevant frequency bands for an analysis of rotating machinery signals, and to obtain the respective ranking of importance is presented. Through the analysis of the entropy and *rms* value of the signal, the BRF is obtained allowing the classification of relevance of the band. In automatic systems (e.g., Artificial Intelligence frameworks) the use of the method allows analyzing only bands that contain relevant information, and in manual analyses, it provides the vibration specialist with guidance on the main focuses of the analysis.

The results obtained for the different dataset show that the BRF is able to identify the bands that present relevant information for the analysis of rotating machinery. The methodology was able to identify the relevance of bands both for synthetically generated data, with and without the presence of noise, and for two real dataset.

Due to its unsupervised characteristic, the methodology can be applied within frameworks as a pre-feature extraction method in data analytics and artificial intelligence applications, avoiding extracting features from irrelevant bands of the signal. The BRF also makes a brief contribution to the development of XAI (when applied together with AI

techniques), by providing the ranking of relevance of the bands. It can also be used to verify the correct operation of sensors (quality of the acquired signal), contributing to the reliability of wireless/remote and IoT (Internet of Things) monitoring systems.

Future works will explore the possibility of using the BRF value as an anomaly detection / drift concept detection method. Considering the variation in the classification of a band from relevant to irrelevant during a time series may indicate variations or anomalies in the system. Or suggesting that the model needs to be retrained due to a new distribution of the data.

In addition, the BRF will be studied as a monitoring feature over time (trend analysis) in order to assess its sensitivity as an indicator of the component's end-of-life. Future developments also include studying the behavior of the method in time series of different applications (e.g., electroencephalogram (EEG) etc).

CHAPTER IV

AN EXPLAINABLE ARTIFICIAL INTELLIGENCE APPROACH FOR UNSUPERVISED FAULT DETECTION AND DIAGNOSIS IN ROTATING MACHINERY

The monitoring of rotating machinery is an essential task in today's production processes. Currently, several machine learning and deep learning-based modules have achieved excellent results in fault detection and diagnosis. Nevertheless, to further increase user adoption and diffusion of such technologies, users and human experts must be provided with explanations and insights by the modules. Another issue is related, in most cases, with the unavailability of labeled historical data that makes the use of supervised models unfeasible. Therefore, a new approach for fault detection and diagnosis in rotating machinery is here proposed. The methodology consists of three parts: feature extraction, fault detection and fault diagnosis. In the first part, the vibration features in the time and frequency domains are extracted. Secondly, in the fault detection, the presence of fault is verified in an unsupervised manner based on anomaly detection algorithms. The modularity of the methodology allows different algorithms to be implemented. Finally, in fault diagnosis, Shapley Additive Explanations (SHAP), a technique to interpret black-box models, is used. Through the feature importance ranking obtained by the model explainability, the fault diagnosis is performed. Two tools for diagnosis are proposed, namely: unsupervised classification and root cause analysis. The effectiveness of the proposed approach is shown on three datasets containing different mechanical faults in rotating machinery. The study

also presents a comparison between models used in machine learning explainability: SHAP and Local Depth-based Feature Importance for the Isolation Forest (Local-DIFFI). Lastly, an analysis of several state-of-art anomaly detection algorithms in rotating machinery is included.

4.1 Introduction

Rotating machinery is commonly used in mechanical systems and plays an important role in industrial applications Lei et al. (2013). Among the various types of rotating machinery, the following popular types are typically adopted in industrial application: aeroengine, steam, gas and wind turbine, automobile transmission, drive trains, fans, blowers, machine tools, compressors, motors, pumps, gearboxes and so on. Even though the rotating machinery is diversified, it generally includes some common essential rotating parts, such as rotors, rolling element bearings, and gears Lei (2017).

The rotor is the rotating part of the equipment and some common faults are: mass unbalance, looseness, bent, misalignment, rub. A rolling element bearing is a component that carries loads by placing rolling elements (such as balls or rollers) between two races. Generally, the rolling bearing consists of three components: outer race, inner race and rolling elements which can suffer from different problems (improper mounting, poor lubrication, entry of foreign matter). All of these problems can cause different bearing faults including: flaking, spalling, peeling, abrasion, scoring, corrosion, pitting, crack, material failure. A gear is a rotating machine part having cut teeth, which meshes with another toothed part to transmit torque. Geared devices can change the speed, torque, and direction of a power source Lei (2017). There are a number of reasons for which a gear may fail in mechanical systems: tooth wear, tooth load, gear eccentricity, backlash, gear misalignment, broken or cracked teeth. Due to the complexity of the systems any fault of the rotating machinery can cause the complete failure of the system, and therefore must be monitored.

The study of artificial intelligence (AI) techniques applied in the monitoring of rotating machinery is a topic in continuous development and of great interest by both researchers and industrial engineers. More and more, industries are adopting more sophisticated technologies for monitoring to increase the reliability and availability of the

machines, and, consequently, remaining competitive in the globalized economy.

As detailed by Liu et al. (2018) there are three basic tasks of fault diagnosis: (1) determining whether the equipment is normal or not; (2) finding the incipient fault and its reason; (3) predicting the trend of fault development. It is clear that when determining the type of fault and its reason (task 2), consequently the answer on the condition of the equipment is obtained (task 1). However, the AI models usually used to classify the type of fault require to be trained with labeled data (supervised training), and examples for all conditions, which in most of the cases is not available in the industry Carletti et al. (2019). In addition, motivated by the recent advances in Deep Learning (DL), the vast majority of AI technologies lack of explainability traits and they require a large volume of data labeled for both normal and fault conditions, dramatically limiting their industry application.

Although the field of rotating machinery monitoring is widely developed, a small number of approaches have been presented based on unsupervised anomaly detection, in relation to the vast majority focused on classification and prognostics, as shown in the review works Liu et al. (2018); Kumar and Hati (2020); Lei et al. (2020); Stetco et al. (2019). A detailed review of AI for fault detection in rotating machines is presented in Liu et al. (2018): most of the 100 references cited there refer mainly to the fault classification. In their review, the main models present in literature were: Artificial Neural Networks (ANNs), k-Nearest Neighbor (kNN), Naives Bayes, Support Vector Machines (SVM) and DL-based approaches. In Kumar and Hati (2020) a review of the main machine learning (ML) and DL techniques applied in the monitoring of induction motors is presented, aiming to detect faults such as: broken bars, bearings, stator faults and eccentricity. Among more than 100 references cited, the vast majority refers to classification and the main models used are: ANNs, Decision Trees, k-NN, SVM and DL-based approaches.

More recently, a broad review with more than 400 citations, focused on AI applications for fault detection is presented Lei et al. (2020). The authors provide a historical overview, in addition to current developments and future prospects. Among the revised ML methods employed in the field, the authors recognized the following as the most commonly adopted: ANNs, Decision Trees, kNN, Probabilistic Graphical Model and SVM. Moreover, the following DL approaches are taken into consideration: Autoencoders, Convolutional

Neural Networks, Deep Belief Network, Residual Neural Networks. As the other cited review works, the references in Lei et al. (2020) mainly focused on classification of the type of fault. The authors also confirm the dependence on real and labeled data from the machine under analysis. In addition to highlighting the recent and future importance in the Intelligent Fault Diagnosis (IFD) scenario of explainable models, with increasing interest starting from 2017. Finally, they mention that traditional ML models should not be abandoned despite the recent advances of DL: this is because it is still worth investigating statistical learning in IFD with the big data revolution, since the theories of statistical learning have rigorous theoretical bases, which promote the construction of diagnostic models with parameters, characteristics and results that are easy to understand.

Anomaly detection is the process of identifying unexpected events in the dataset, which are different from normal. In general, the signals generated by a fault have characteristic patterns that are different from normal and indicate a change in the behavior of the machine. Using a method that indicates changes in the current condition of the equipment, does not need a labeled historical dataset for training and provides explainability of the results can be the solution to the mass dissemination of artificial intelligence methods in the industrial environment for monitoring rotating machinery.

Among the references studied, there are very few studies involving anomaly detection with unsupervised approaches in the monitoring of rotating machinery. Authors in Ogata and Murakawa (2016 in Bilbao, Spain.) used Fourier local autocorrelation (FLAC) and Gaussian Mixture Model (GMM) approach (based on class cluster) to extract features in the time-frequency domain and to detect faults respectively; in the same work, vibration signals were used to detect faults in wind turbine components. The authors showed that the use of features extracted from FLAC improves the model's performance, making it possible to detect anomalies in even more complicated cases, such as low speeds, where conventional features do not present such satisfactory results. In von Birgelen et al. (2018) an application of the Self Organizing Maps for anomaly detection was presented, showing its effectiveness in detecting variations when the component fails: uses case related to cyber-physical system components (bearings and blades) were exploited. In Amruthnath and Gupta (2018) the authors proposed the use of different methods of ML using vibration

signals from a fan for unsupervised detection of incipient fault. The algorithms used were: PCA T2 statistic, Hierarchical clustering, K-Means, Fuzzy C-Means clustering. Finally, they presented a comparison of the models, showing the feasibility of implementing them in the monitoring of machines. Other studies Zhang et al. (2019); Hasegawa et al. (2018, 2017) propose anomaly detection approaches in the monitoring of rotating machinery based on combinations of different techniques with variations of GMM. To the best of our knowledge, the vast majority of state-of-the-art ML unsupervised anomaly detection, have never being used, e.g., Isolation Forest (IF), Local Outlier Factor (LOF), Angle-Based Outlier Detection (ABOD) etc.

Another important aspect in the field that has not been fully explored yet is the one related to interpretability of ML-based monitoring solutions in equipment machinery: as argued above, without providing explainable results to the user, even when ML-based modules provide excellent results in historical data, AI models are unlikely to be applied in real-world scenarios Molnar (2020). Moreover, as mentioned by Lei et al. (2020), collecting labeled data from machines generates a high cost, and consequently unlabeled data is the majority in engineering scenarios. Therefore, using an AD (anomaly detection) model that works with unlabeled data and provides explainability is essential to enable large-scale implementation of AI in the monitoring of rotating machinery.

Recently studies are being developed with a focus on Explainable Artificial Intelligence (XAI). In order to explain black-box models, different methods can be used according to the ML model in use Du et al. (2019). In general, the methods provide information to understand how the model performs fault detection, which can be, for example, a ranking of the most important features, the model weight relevance or the most significant points in the underlying signals Doshi-Velez and Kim (2017). Despite the current interest, the vast majority of studies are focused on explainability for DL models and mostly on fault classification. More information can be found in the articles available on the topic Lei et al. (2016); Zhang et al. (2017); Jia et al. (2018); Li, Zhao, Sun, Cheng, Chen, Yan and Gao (2019); Abid et al. (2020); Li, Zhang and Ding (2019); Chen and Lee (2020); Grezmak et al. (2019, 2020); Saeki et al. (2019). Among the references researched, only Saeki et al. (2019); Hendrickx et al. (2020) address the explainability of the model in anomaly detection,

being Saeki et al. (2019) based on DL. Hendrickx et al. (2020) presented a methodology for detecting anomalies in electric motors (voltage unbalance) using a set of similar equipment through electrical and vibration signature. The authors use generic building blocks and present advantages of not needing historical data, incorporating human knowledge. Despite the interesting approach, it is noted that for its use it is necessary to have data from more than one machine, so that they can be compared, making applications on single machines unfeasible.

In this chapter, a new approach for fault detection and diagnosis in rotating machinery is proposed. In the first part, the vibration features in the time and frequency domains are extracted. Secondly, in the fault detection, the presence of fault is verified in an unsupervised manner based on anomaly detection algorithms. Finally, in fault diagnosis, Shapley Additive Explanations (SHAP), a technique to interpret black-box models, is used. Through the feature importance ranking obtained by the model's explainability, the fault diagnosis is performed. Two approaches of diagnosis are proposed, namely: unsupervised classification and root cause analysis. Due to the importance of rotors, bearings and gears for rotating machinery and especially their respective faults, three datasets are proposed for validation of the methodology, each one being related to a respective component and its possible faults.

The main contributions of the proposed approach are: i) unsupervised identification of the fault in rotating machinery through vibration analysis; ii) unsupervised classification of the type of fault in rotating machinery, based on the analysis of the features relevance; iii) possibility of performing root cause analysis when the features may be related to more than one fault and the unsupervised classification is not feasible; iv) a new contribution to the study of XAI and novel application in fault diagnosis for rotating machinery is presented based on SHAP and Local-DIFFI; v) possibility to be applied in different types of faults; vi) possibility to change models according to the dataset; vii) industrial applications.

To the best of the authors' knowledge, this is the first study to compare and analyze unsupervised state-of-the-art anomaly detection algorithms for monitoring rotating machinery. In addition to providing explainability about the ML models used and proposing

a new approach to perform unsupervised classification or root causes analysis.

The remainder of this chapter starts with a brief explanation about the machine learning and XAI methods used in Section 2. The proposed approach is presented in Section 3. Experimental procedure is shown in Section 4. Analysis of the experimental results are given in Section 5. Finally, Section 6 concludes this chapter.

4.2 Background

4.2.1 Anomaly Detection Algorithms

In this sub-Section we will provide a brief overview on the data-driven unsupervised Anomaly Detection (AD) algorithms compared in this work.

Anomaly detection (also known as outlier detection[1]) refers to the task of identifying rare observations which differ from the general ('normal') distribution of a data at hand Zhao, Nasrullah and Li (2019). Anomaly Detection approaches have the capability of summarizing the status of a multivariate systems with a unique quantitative indicator, that is typically called *Anomaly Score* (AS)[2]: while many approaches provide guidelines on how to define outliers based on the AS, the quantitative nature of the AS indeces allowed to implement different strategies that allow to govern the trade-off between false positives and false negatives depending on the application at hand. While no applications to the best of our knowledge have been presented in the field of rotating machinery monitoring using vibration data and state-of-art models (that will be introduced in the rest of the Section), anomaly detection approaches have been successfully applied in various areas like biomedical engineering Meneghetti et al. (2018), fraud detection Rai and Dwivedi (2020), oil and gas Barbariol et al. (2020) and additive manufacturing Li et al. (n.d.).

Algorithms are arranged by increasing year of presentation. In view of the large number of tested models, only the algorithms achieving the best performance in this study are discussed. The mathematical formulation and details for Lightweight on-line detector of anomalies (LODA), Angle-based Outlier Detector (ABOD) / Fast-ABOD and Cluster-based

[1]The terms 'Anomaly' and 'Outlier' will be treated in the same way in this work.

[2]Other authors refer to the concept of Anomaly Score with various names like for example Health Factor or Deviance Index.

Local Outlier Factor (CBLOF) can be retrieved at Pevn'y (2016); Hans-Peter Kriegel and Zimek (2008); He et al. (2003).

k-Nearest Neighbors (kNN)

k-nearest neighbor (kNN) is a simple and popular method used for supervised tasks of classification and regression. In the context of AD, kNN can be also employed: given a sample, the distance to its kth-nearest neighbor can be considered as AS Knorr and Ng (1998). More formally, the anomaly score Ramaswamy and K.S. (2000) is then defined in Equation 4.1.

$$s_{\text{kNN}}(x) = D^k(x) \tag{4.1}$$

where $D^k(x)$ denotes the distance of the k^{th} nearest neighbor from observation x. The distance function can be any metric distance function. The most common methods for selecting distance function are: largest distance, where the distance to the k^{th} neighbor is used as the AS; mean distance, where the AS is the average of all k neighbors; median distance, which uses the median of the distance to k neighbors as AS.

Minimum Covariance Determinant (MCD)

The minimum covariance determinant (MCD) is a robust estimator of multivariate locations and its goal is to find n instances (out of N) whose covariance matrix has the lowest determinant Rousseeuw and Driessen (1999). In the context of AD, MCD is used with Mahalanobis distance (MD), a well-known distance metric of a point from a distribution: first a minimum covariance determinant model is fitted and then the Mahalanobis distance is used as AS. Since the parameters required by MD are unknown (mean and covariance matrix), the MCD model is used to estimate them, and then the MD can be calculated as follows, in Equation 4.2.

$$s_{\text{MCD}}(x) = d(x, \bar{x}, Cov(X)) = \sqrt{(x - \bar{x})' Cov(X)^{-1} (x - \bar{x})} \tag{4.2}$$

where $\bar{x}$ is the sample mean and $Cov(X)$ is the sample covariance matrix. If data are assumed centered not normalized, the robust location and covariance are directly computed with the FastMCD algorithm without additional treatment. Otherwise, the support of the robust location and the covariance estimate are computed, and a covariance estimate is recomputed from it, without centering the data Zhao, Nasrullah and Li (2019).

Local Outlier Factor (LOF)

LOF Breunig et al. (2000) is a density-based approach for AD; such class of approaches are based on the study of local neighborhoods of the data points under exam: an observation is a dense region is considered as a normal data point (also referred in the literature as an *inlier*), while observations in low-density regions are anomalies.

The LOF procedure involves two steps: (i) evaluating the so-called Local Reachability Density; (ii) evaluating the AS s_{LOF}. the Local Reachability Density of a data point x in its k-neighborhood $\mathcal{N}_k(x)$ (the space where the k other data points closest to x are living) is defined in Equation 4.3.

$$\text{LRD}_k(x) = \frac{k}{\sum_k (y \in \mathcal{N}_k(x)) r_k(x, y)} \tag{4.3}$$

where $r_x(x, y) = \max\{d_k(x), d(x, y)\}$ is the so-called reachability distance and $d_k(x)$ is the distance from x of its k-th nearest neighbor. The reachability distance just defined is used instead of the distance $d(x, y)$ in order to reduce statistical fluctuations/noise in the evaluation of the AS s_{LOF}; the AS is in fact defined in Equation 4.4.

$$s_{\text{LOF}}(x) = \frac{1}{k} \sum_{y \in \mathcal{N}_k(x)} \frac{\text{LRD}_k(y)}{\text{LRD}_k(x)} \tag{4.4}$$

The above defined anomaly score can assume values between 0 and ∞, however, a value around 1 (or lower than 1) indicates that the data point x is somehow similar to its neighbors and it can be therefore considered as an inlier; a value of s_{LOF} larger than 1 indicates instead a case in which the data point under exam can be considered as an outlier. For more details we refer the interest readers to Breunig et al. (2000). LOF is a classic approach to AD and extended versions of the algorithms have been proposed over the years Kriegel et al.

(2009); Schubert et al. (2012): in this work we also consider the popular Cluster-based LOF (CBLOF) He et al. (2003).

One-class Support Vector Machines (OCSVM)

One-Class Support Vector Machine Schölkopf et al. (2001) is an extension for AD of the popular approach for classification known as Support Vector Machine. The training data is projected to a high-dimensional space and the hyperplane that best separates the points from the origin is determined. When evaluating a new sample, if it lays within the frontier-delimited subspace, it is considered to come from the same population and therefore it is considered as an inlier; otherwise, the data point is considered as an anomaly by the approach.

As in SVM, kernel functions are used to produce non-linear hyperplanes; different kernels can be used: linear, polynomial, sigmoid, gaussian. In this work, the kernel coefficient for gaussian, polynomial and sigmoid will be called *gamma* and the parameter to define an upper bound on the fraction of training errors and a lower bound of the fraction of support vectors, nu.

Feature Bagging (FB)

Feature Bagging is the combination of multiple outlier detection algorithms using different set of features Lazarevic and Kumar (2005). Every outlier detection algorithm uses a small subset of features that are randomly selected from the original feature set. Any AD approach can be used as the base estimator. Using a cumulative sum approach, each AS generated by each outlier detector used is combined in order to find a final AS and described in Equation 4.5.

$$s_{\text{final}}(x) = \sum_{t=1}^{T} s_t(x) \tag{4.5}$$

where the final anomaly score $s_{\text{final}}(x)$is the sum of all anomaly scores s_t from all T iterations on each outlier detector used. The number of base estimators in the ensemble and the number of features to draw from X to train each base estimator can be adjusted. Moreover, the final

combination of the AS can be performed by the averaging all models or taking the maximum scores.

Isolation Forest (IF)

Isolation Forest (iForest or IF), Liu et al. (2008, 2012), uses the concept of *isolation* instead of measuring distance or density to detect anomalies. The IF exploits a space partitioning procedure: the main idea underlying the approach is that an outlier will require less iterations than an inlier to be isolated, i.e., to find through the partitioning procedure a region of the space where only such observation lies in.

The partitioning procedure used by the IF is achieved through the creation of iTrees, binary trees that are the result of a *random* partitioning procedure obtained by splitting the data based on one of their features at each iteration of the algorithm. Following the above stated fundamental idea of IF, it is expected that the path to reach a leaf node from the root of an iTree will be shorter for outliers than for inliers; the anomaly score will be related to this path length: the shorter the more anomalous the data point. We underline that this procedure is done randomly: to achieve fast computation the features and the splitting points are chosen randomly; the drawback of this approach is that a single tree can give an estimate of the path length that has high variance: thus, similar to the popular Random Forest (that we remark is a supervised approach), an ensemble of T trees is constructed in order to provide a low-variance estimation. More in detail, an iTree is built as follows.

1. A subsample of data $S \in X$ is randomly selected.

2. A feature $v \in \{1, \ldots, p\}$ is randomly selected: a node in the tree is created and at this node the value of v is used;

3. A random threshold $\bar{v}$ on v is chosen within the domain of the variable;

4. Two children nodes are generated: one associated to the points with values for variable v below $\bar{v}$ and one for those with value above;

5. The points from 2 to 4 of this procedure are repeated until either a data point is isolated or a threshold on the maximum tree length is reached.

After the iTrees are constructed, the AS score for a data point x, Equation 4.6.

$$s_{\mathrm{IF}}(x) = 2^{\frac{-E(h(x))}{c}} \tag{4.6}$$

where $h(x)$ is the length of the path for a data point from its leaf to the root, $E(h(x))$ is the average of $h(x)$ in iTrees collection of iTrees and c is an adjustment factor which is set to the average path length of unsuccessful searches in a binary search tree procedure. Using the AS just defined, if instances return s_{IF} very close to 1, then they are tagged as anomalies; on the other hand, values much smaller than 0.5 are quite safe to classify as normal instances, and values close to 0.5 then the entire sample does not really have any distinct anomaly Liu et al. (2012).

iForest works well in high dimensional problems which have a large number of irrelevant attributes, and in situations where training set does not contain any anomalies. Given its high performance and the possibility to parallelize its computation (thanks to its ensemble structure), IF is probably the most popular AD approach: for this reason, we will consider, as it will be detailed in Section 4.2.2, a dedicated approach for providing interpretable traits to IF.

Histogram-based outlier score (HBOS)

HBOS is an AD approach based on histograms that was introduced for providing fast computation of an AS w.r.t. previously proposed AD methods.

The HBOS algorithm can be summarized as follows: univariate histograms for each single feature are computed (in case of numerical data a set of k bins of equal size are used for each histograms). The number of bins k is an hyper-parameter that needs to be tuned; histograms are normalized to $[0, 1]$ for each single feature; frequency (relative amount) of samples in a bin is used as density estimation; AS for each instance x is computed as a

product of the inverse of the estimated density, Equation 4.7.

$$s_{\text{HBOS}}(x) = \sum_{i=0}^{p} log\left(\frac{1}{hist_i(x)}\right) \tag{4.7}$$

where p is the number of features and $hist_i(x)$ is the density estimation. With such definition of the AS, with HBOS the outliers correspond to high values of $s_{\text{HOBS}}(x)$, while inliers to low values. In this algorithm, two parameters are still employed and need to be tuned, being α and the tolerance (tol). α is a regulation factor to avoid overfitting and tol adjusts the flexibility while dealing the samples falling outside the bins.

Ensemble

The ensemble method combines different algorithms to obtain a single final result. Knowing that ML models are sensitive to the types of data, ensemble methods are commonly used to increase the efficiency and robustness of the final result. Being H_i the result of each i^{th} base model, the sum of the k selected ones is the final result (FR) of the ensemble method, and the final decision (FD) is obtained by a majority voting, both described as:

$$FD = \begin{cases} 1, \text{ if } FR > k/2 \\ 0, \text{ otherwise.} \end{cases} \text{, where } FR = \sum_{i=1}^{k} H_i \tag{4.8}$$

Where in this case, 1 indicates that the sample is an anomaly and 0 that the sample is normal.

4.2.2 Explainable Artificial Intelligence (XAI)

One of the main limitations of Machine Learning models application in real situations is related to their lack of explainability. In many real-world applications it is necessary to understand how the model made a certain decision, allowing the specialist to trust the predicted result. For such reason, many XAI approaches have been presented in the past recent years.

In order to explain black-box models, different methods can be used according to the ML model in use Du et al. (2019). These methods can be classified according to various

criteria. The first difference is whether the methods are intrinsic or post-hoc: in intrinsic methods, explainability is performed by restricting the complexity of the model, while post-hoc are methods that analyze the model after training. Another important concept is whether the explainability models are model-specific or model-agnostic: model-specific approaches are limited to a specific model, on the other hand, model-agnostic can be used on any machine learning model and is applied after the model has been trained (post-hoc) Molnar (2020). Model-specific approaches in general access the structure of the ML models or their weights to make inferences, while model-agnostic techniques usually works by analyzing the input features and their respective outputs of the ML model.

Local and global explanations are other different approaches to XAI. Global explainability is usually used to describe the something regarding the model as a whole, while local explainability aims to understand the reason behind the result obtained for a specific condition/observation.

In this subsection the XAI approaches adopted in this work are revised.

Shapley Additive Explanations (SHAP)

Shapley Additive Explanations, M and Su-In (2017) is a state-of-art and model-agnostic (it can be applied to any algorithm) for interpreting ML predictions, both in unsupervised and supervised tasks.

Based on Shapley values from coalitional game theory, SHAP provides a feature importance ranking which can be used to explain the ML model to the individual data point level: in the context of anomaly detection, having an ordered list of features can be really helpful for domain expert to enable an effective troubleshooting. The feature importance ranking is the result of the contribution of each feature to the final prediction of the model.

Since the Shapley values are expensive to obtain, SHAP approximates them of a conditional expectation function of the original model. The detailed mathematical formulation of SHAP can be retrieved at M and Su-In (2017).

Local Depth-based Feature Importance for the Isolation Forest (Local-DIFFI)

Given the increased interest and popularity of IF, we chose to consider in this work also a model-specific approach for providing, like in SHAP, a feature importance ranking.

Local Depth-based Feature Importance for the Isolation Forest is the first model-specific method for interpretability in IF Carletti et al. (2020). While IF is one of the most commonly adopted AD algorithms, its structure and prediction lack in interpretability. To overcome this problem the Local-DIFFI method proposes an effective and computationally inexpensive approach to define local feature importance (LFI) in IF, Equation 4.9.

$$LFI = \frac{I_o}{C_o}, \tag{4.9}$$

where C_o is the features counter for the single predicted outlier x_o and Io is updated by adding the quantity Carletti et al. (2020) while iterating over all the trees in the forest, Equation 4.10.

$$\Delta = \frac{1}{h_t(X_o)} - \frac{1}{h_{max}} \tag{4.10}$$

The model is a post-hoc method, which, due to its operation, preserves the performance of an established and effective AD algorithm (IF). An interesting property of Local-DIFFI is that, while achieving comparable results w.r.t. SHAP, its computing time is orders of magnitudes smaller than SHAP. The method proposes to provide additional information about a trained instance of the IF model with the main objective of increasing the users' confidence in the result obtained. Besides the local feature importance provided by Local-DIFFI, the method can also be used to provide global feature importance, namely DIFFI.

4.3 Proposed Approach

The proposed methodology is depicted in Fig. 4.1 and it is divided into three parts: 1) Feature extraction; 2) Fault detection: Anomaly Detection; 3) Fault diagnosis: Unsupervised classification / Root cause analysis. The vibration features are initially extracted based on the type of monitored component. The extracted features are divided

into a training and testing group, and the hyperparameters of the anomaly detection models are tuned. The samples are evaluated in the fault detection part: if a fault (anomaly) is not detected, the analysis is completed; on the other hand, if the sample is a fault (anomaly), the most relevant features used to generate the result are evaluated through the model's explainability. In the fault diagnosis part, the features that indicate only the presence of fault, but do not indicate the type / location are disregarded (called general features, e.g., rms and kurtosis). For components that have unique fault specific features (e.g., bearing, gearbox), it is possible to perform an unsupervised classification based on the most relevant feature for the result. On the other hand, for analysis where the features may be related to more than one fault (e.g., misalignment and mechanical looseness), the most relevant features (feature ranking) for identifying the sample as an anomaly are presented, allowing the specialist to analyze the problem in more detail, namely root cause analysis.

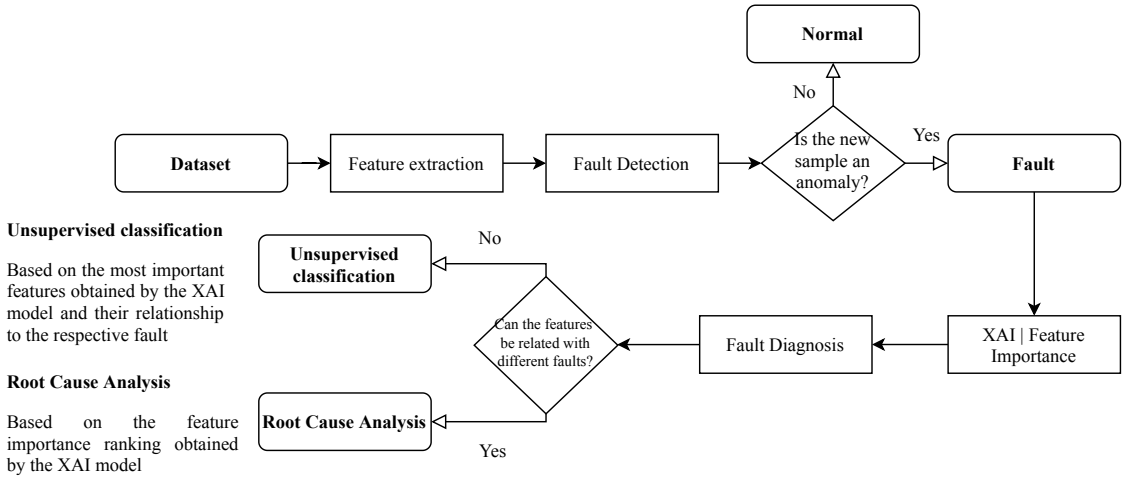


Figure 4.1: General framework of the proposed methodology.

4.3.1 Feature extraction

One of the main reasons for the wide use of DL models in many tasks is that DL approaches implicitly implement a feature extraction procedure due to the DL architectures ability to learn discriminating features through non-linear relations performed within the model: avoiding the time consuming task of feature extraction is a captivating property for ML technologies developers. However, in the domain of rotating machinery, the vast majority of faults have already been studied and ad-hoc defined features that are informative

for fault detection (that can be computed directly on the raw signals or after dedicated signal processing filters) have been developed by researchers over the years. For this reason, we have decided to base our approach on 'classic' ML techniques exploiting the wide knowlegde of filtering approaches and feature definitions provided by the literature.

Among the sensors used for monitoring rotating machinery, the vibration-based diagnostic method is the most popular and researched. The interest is justified by the fact that the vibration signals directly represent the dynamic behavior of the equipment Ciabattoni et al. (2018); Samuel and Pines (2005); Dalvand et al. (2017); Wei et al. (2019) and are a non-invasive technique. The features to detect faults in rotating machinery using vibration signals are commonly extracted from the time, frequency and time-frequency domains Lei et al. (2020).

(i) Among the most used in the time domain are: mean, standard deviation, rms (root mean square), peak value, peak-to-peak value. According to Lei et al. (2007, 2010), these features can be affected by the speed and load of the machines, therefore, other features are also commonly used to fill this gap: shape indicator, skewness, kurtosis, crest factor, clearance indicator, etc., which are robust to the machine's operating conditions.

(ii) The features in the frequency domain are extracted from the frequency spectrum, for example: mean frequency, central frequency, energy in frequency bands, etc. Different information can be obtained that is not found or is hardly extracted in the time domain Lei et al. (2020).

(iii) For the time-frequency domain, features such as entropy are usually extracted by Wavelet Transform, Wavelet Packet Transform and empirical model decomposition. These features are capable of reflecting the machine's health states in non-stationary operating conditions Lei et al. (2020).

In this study, two approaches were combined in relation to the types of features. Firstly, general features were selected to indicate the presence of system fault and degradation. This approach does not allow the identification / location of the fault, but it allows to detect variations in the system in a global way, avoiding that a fault is not identified. Secondly, specific features commonly associated with the type of defect in the respective components were used to enable the identification / location of the fault. During

the extraction of specific features, it must be defined whether the features are related to different faults or are unique, enabling the fault diagnosis through unsupervised classification or root cause analysis. The choice of the selected features for this study was based on those most used industrially in predictive maintenance, in order to provide a better understanding and to enable future applications without the need for deep signal processing. It is worth mentioning that different features (for example the ones proposed in Sánchez et al. (2020, 2018)) can also be applied in the proposed framework, that has also the advantage of allowing for a certain flexibility in the features choice.

4.3.2 Fault detection: Anomaly detection

Identifying the fault is extremely important for production processes, and even more important when performed in an unsupervised manner, that is, without the presence of labeled data related to the fault modes in training set. In this part, the extracted features are divided into a training and testing group, and the hyperparameters of each AD model are adjusted. The samples are evaluated in unsupervised manner and identified as normal or fault (anomaly). If an anomaly is not considered, the analysis is completed. On the other hand, if the sample is an anomaly, the root cause analysis or fault classification based on the model's explainability is performed.

Different models used in the field of AD were studied. As is common knowledge, the performance of AD models are strongly related to the type of data available. While we report in the following the best approaches in our studies, the approach is generic and the user can modify the AD model in use if the expected performance is not achieved, without affecting its structure.

The different AD algorithms evaluated were the ones reported in Section 5.2.2: Clustering Based Local Outlier Factor (CBLOF), Local Outlier Factor (LOF), Isolation Forest (IF), Lightweight on-line detector of anomalies (LODA), Histogram-based Outlier Detection (HBOS), k-Nearest Neighbors (kNN), Fast - Angle-based Outlier Detector (FastABOD), Outlier Detection with Minimum Covariance Determinant (MCD), One-Class Support Vector Machine (OCSVM), Feature Bagging (FB) and Ensemble (combination of all models) available in Zhao, Nasrullah and Li (2019).

4.3.3 Fault diagnosis: Unsupervised Classification / Root Cause Analysis

Despite the advances in ML applications for fault diagnosis in rotating machines, the vast majority of methods are performed in a supervised manner. In other words, the methods use labeled data in the training to ensure that the model is able to distinguish between different classes of faults. The proposed methodology presents an approach where no training labels are necessary. The fault diagnosis is performed in an unsupervised manner based on the importance ranking obtained by the model explainability. Two different analysis are possible depending on the type of component being monitored, namely: unsupervised classification and root cause analysis. For faults that have unique characteristic features (e.g., bearings, gearbox) unsupervised classification can be performed directly. On the other hand, for analysis where the features may be related to more than one fault (e.g., misalignment and mechanical looseness), the most relevant features (feature ranking) to identifying the sample as an anomaly are presented, allowing the specialist to analyze the problem in more detail, called root cause analysis.

The methodology is based on the feature importance ranking for each new sample identified as an anomaly, as presented in Algorithm 1 and Fig.4.2. After identifying the anomaly in the previous part, the most relevant features are analyzed through the model's explanability. SHAP is used to obtain the feature importance ranking. The general features that only indicate the presence of a fault, but do not indicate the type / location are disregarded (e.g. rms and kurtosis). A new ranking of importance is obtained using only the specific features. As each specific feature is related to a potential unique type/location of component fault, the most relevant feature is considered as the type/location of the fault present in the system. For root cause analysis, since the features may be related to more than one fault, the feature importance ranking is presented, assisting the specialist in identifying the type of fault. For example, assuming that based on the importance score calculated by SHAP, the most relevant features in order are: rms, Ball Pass Frequency Outer (BPFO), Ball Pass Frequency Inner (BPFI), kurtosis, Ball Spin Frequency (BSF). Applying the methodology, the new ranking of importance would be: BPFO, BPFI and BSF. After that, according to the type of procedure applied, the result is obtained. For unsupervised

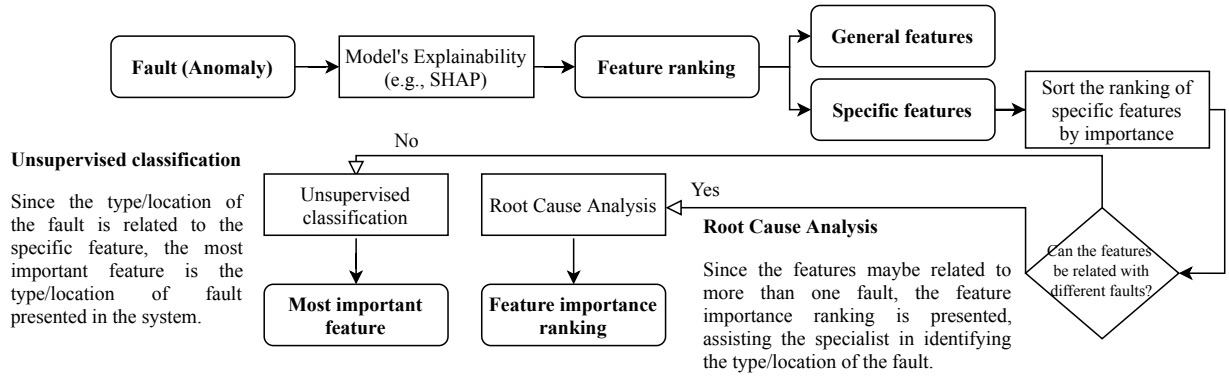


Figure 4.2: Specific framework of the proposed methodology for Unsupervised Classification / Root Cause Analysis.

classification, the specific features are analyzed, and the fault is classified based on the feature most relevant to the result. In other cases, where the features are related to more than one fault, root cause analysis is performed and the feature importance ranking is presented. The methodology is applicable for the other types of rotating machinery, where the only modification is the type of feature to be analyzed. In the case of the gearbox, for example, instead of BPFO, BPFI and BSF, the specific features would be the gear mesh frequency and its harmonics for each stage, 1xGMF (Gear Mesh Frequency), 2xGMF, 3xGMF, 4xGMF (1^{st}Stage), 1xGMF, 2xGMF (2^{nd} Stage) and for the case of mechanical faults (rotor), the features would be: 1xfr (rotating frequency), 2xfr, 3xfr and 4xfr. It is important to highlight that other features can be used according to the knowledge related to the problem under study.

Understanding which features the model uses to identify the anomaly is essential to perform root cause analysis/classification. In other words, through explainability it is possible to mimic human knowledge. Without the use of an explainability algorithm such as SHAP and Local-DIFFI, it is not possible to carry out the analysis, since the models used do not present explanations of how the final results were obtained. Thus, the association of state-of-the-art models to identify anomalies in the signals with algorithms that perform the explanability, allows the proposition of the new methodology.

Even though it is the state-of-the-art in explainability and model-agnostic, SHAP presents a high computational cost in relation to model-specific solutions. Therefore, a comparison was made using the recent proposed explainability algorithmic, Local-DIFFI

for the Isolation Forest model. As stated above, the choice of model-specific Local-DIFFI is due to the fact that Isolation Forest presents excellent results in the literature and good robustness in relation to the variation of hyperparameters. Moreover, in general, Local-DIFFI presents very similar results to SHAP, as shown in Carletti et al. (2020). The similarity of the models was verified through Kendall-Tau rank distance, a metric commonly used for evaluation between two ranking lists.

Algorithm 1 Pseudo-Code

```
procedure UNSUPERVISED CLASSIFICATION
    Type: specific analysis / specific feature related to a single fault
    Input: new sample
    Output: fault classification (most important specific feature)

    if new sample = anomaly then
        feature importance ranking ← shap or local-diffi(new sample)
    feature importance ranking.drop(general features)
    feature importance ranking ← sort(feature importance ranking)
    most important feature ← feature importance ranking[0]
    print('The fault is located in: ', most important feature)

procedure ROOT CAUSE ANALYSIS
    Type: general analysis / specific feature related to different faults
    Input: new sample
    Output: root causes (most important specific features)

    if new sample = anomaly then
        feature importance ranking ← shap or local-diffi(new sample)
    feature importance ranking.drop(general features)
    feature importance ranking. ← sort(feature importance ranking)
    print('The root causes are related to: ', feature importance ranking)
```

4.4 Experimental procedure

4.4.1 Data description

Three datasets were used to address different faults found in rotating machinery. As showed in Lei (2017) even though the rotating machinery is diversified, some common essential rotating parts are: rotors, rolling element bearings, and gears. Due to the importance of each component and its respective fault for mechanical systems and consequently for industrial production, the proposed datasets were chosen. The faults

analyzed were: defects in bearing and gearbox, misalignment, unbalance, mechanical looseness and combined faults. The use of different datasets, with different monitoring approaches, aims to validate the proposed methodology in different scenarios.

Case 1: Bearing Dataset

The first dataset considered (publicly available Qiu et al. (2006)), namely *Bearing Dataset*, is composed by three run-to-failure tests with four bearings in each test. The rotation speed was kept constant at 2,000 rpm by an AC motor coupled to the shaft via rub belts. A radial load of 6,000 lb was applied to the shaft and bearing by a spring mechanism. Rexnord ZA-2115 double row bearings were installed on the shaft. PCB 353B33 accelerometers were installed on the bearings housing. All failures occurred after exceeding the projected bearing life, which is more than 100 million revolutions Qiu et al. (2006). For the study, bearing 01 of test 02 was used. Each test consists of individual files of vibration signals recorded at specific intervals. Each file consists of 20,480 points with the sampling rate set at 20 kHz. NI DAQ Card 6062E was used for collection.

The dataset consists of run-to-failure tests, therefore no labels are available indicating the fault start: the only information provided is the type of fault present at the end of each test. To assess the efficiency of the AD model, the data was manually labeled. In the analysis, it was considered that after starting the defect, all subsequent observations correspond to a faulty bearing. It is worth mentioning that the labels were used only to evaluate the efficiency of the methodology and they were not used by the AD model.

The test has 984 observations, with the first 531 observations labeled as normal and the last 453 as anomalies (fault). The fault was identified in the outer race. The features used were: kurtosis, rms, BPFI, BPFO and BSF, which are widely used in bearing fault detection Bolón Canedo et al. (2013); Zhang et al. (2011, 2018); Lei and Zuo (2009); Li, Yang, Li, Xu and Huang (2017); Singh and Shaik (2019). Specific features are those that indicate the type of fault (BPFI, BPFO and BSF) and general features are those that indicate the presence of a defect (kurtosis and rms). The bearing fault frequencies are important to assess the type of defect and confirm its existence, which is not always noticed by other features. It is also important mentioning that there are cases where the fault does not present the classic

defect behavior with the deterministic bearing frequencies in evidence Smith and Randall (2015), which makes it important to use other features. Knowing that bearing faults are generally associated with impacts, kurtosis is a relevant feature for the study. Finally, the rms value represents the global behavior of the system, indicating a general degradation and accentuation of the defect. The purpose of using this dataset, in addition to identifying the presence of the fault in a real monitoring situation, is to classify the type of fault using the proposed methodology. It is important to highlight that other datasets (as in Wang et al. (2020)) were tested to validate the methodology, but due to the similarity of the results, it was decided to present only this one.

Case 2: Gearbox Dataset

The second dataset considered, the *Gearbox Dataset*, was presented in Cao et al. (2018) and it is used to evaluate faults in gearbox. A 32-tooth pinion and an 80-tooth gear were installed on the first stage input shaft. The second stage consists of a 48-tooth pinion and 64-tooth gear. The data were recorded using an accelerometer through a dSPACE DS1006 system, with sampling frequency of 20 KHz. Nine different gear conditions were introduced to the pinion on the input shaft, including healthy condition, missing tooth, root crack, spalling, and chipping tip with five different levels of severity. For each gear condition, 104 observations were collected resulting in a total of 936 observations.

It is common knowledge that general gear problems tend to increase the energy of the sidebands spaced from the rotation frequency around the Gear Mesh Frequency (GMF) and their respective harmonics. Thus, simulating a real condition, the features used were: kurtosis, rms, 1xGMF, 2xGMF, 3xGMF, 4xGMF (1^{st}Stage), 1xGMF, 2xGMF (2^{nd} Stage). Due to non-stationary issues and the uncertainty caused by speed varying, instead of using the energy value in each GMF and respective side bands, the energy in the GMF band +/- 4*(nominal rotation frequency) was calculated. In addition to being able to detect the fault (AD), the use of this fault dataset aims to identify the location of the fault in the gearbox (first or second stage) and not to classify the type of fault (missing tooth, root crack, spalling and chipping tip).

Case 3: Mechanical Fault Dataset

The last dataset, *Mechanical Fault Dataset*, was developed by one of the authors Brito and Pederiva (2002, 2003): the dataset contains different electrical and mechanical faults which were inserted in a experimental test rig, where one of the components is a rotor; in this work we will consider the following faults: unbalance, misalignment, looseness and combined faults (being the combination of the previous ones). Six accelerometers were used to acquire the vibration signals, in the horizontal, vertical and axial positions, three in the fan-end side and three in the drive-end side.

The rotation speed was kept constant at 1717.5 rpm. The observations were labeled according to the fault introduced in the test rig, and later analysis of the vibration spectrum. Each file consists of 3,200 points with $df = 0.125$ Hz. The dataset contains 5 conditions with a total of 1418 observations (532 normal, 557 unbalance, 283 misalignment, 28 mechanical looseness and 18 combined fault).

In general, the unbalance is commonly identified in the vibration signal by increasing the energy in 1 x fr (rotating frequency). It is noteworthy that other faults can also appear in 1 x fr as structural problems and even mechanical looseness. The most common types of misalignment and mechanical looseness show an increase in energy level in 2 x and 3 x fr, and therefore may have similar characteristics. The mechanical looseness can still have multiple and sub-harmonics of fr. Considering the types of faults and the respective behaviors, the following features were used: rms, energy level in 1 x fr, 2 x fr, 3 x fr and 4 x fr.

In addition to the basic objective of identifying the fault, the use of this dataset aims to evaluate the classification methodology with a focus on root cause analysis, when the features are correlated with more than one type of fault. The dataset also provides the possibility to study isolated and combined faults that, although known, have been little used in studies involving fault detection and new techniques of artificial intelligence compared to bearings and gearboxes.

4.4.2 Analysis approaches

Two approaches were used to define 3 different scenarios [Case 1, 2, 3] that can be found in real-world monitoring applications.

In the first approach, a dynamic condition was considered with the data collected in sequence, where a temporal relationship and fault evolution is presented [Case 1]. For the study, a sliding window was used, where the training group was updated with each new sample, in case it was considered normal. 100 samples were initially used for the training group in order to ensure stability in the models. For this situation, as the model was started together with the machine under normal conditions (e.g.: after maintenance or a new machine), there are no anomalies in the training group. It is worth noting that this approach can also be used if there are anomalies in the training group (e.g.: cases of continuous monitoring where the machine was repaired after a fault, and it is desired to use all the signals to increase the amount of data in the model).

In the second approach, a static condition was considered, where the signals do not have a temporal correlation with each other [Case 2 and 3]. This approach simulates when historical data are available for the machine without labels. They also refer to different types of faults and normal conditions, however not necessarily collected in sequence. It is important to highlight that although Cases 2 and 3 represent the same condition called static condition, the types of faults studied are different in each case. The data were divided into training and test groups. Due to the number of observations available, the size of the training group is limited by the number of normal samples and the rest designated as a test. The training group consisted of 80% of samples of normal condition and 20% of anomalies selected at random. The proportion has been defined as a machine in operation is mostly in normal condition, and few situations with faults. Such an approach also shows that it is possible to implement the proposed methodology even with the presence of anomalies in training set.

4.4.3 Hyperparameter tuning

The hyperparameters for each model were adjusted based on the training group to obtain the best performance and are shown in Table 4.1. A cross-validation procedure was applied, using a Leave P Out (LPO) approach for the dynamic condition, where 5% of the training samples were removed in each new update of the training group. For the static condition, a cross-validation method was also applied but, in this case, using random permutations (shuffle and split), where the signals were randomly chosen for the test and training group at each iteration of the model. The hyperparameters are presented in relation to the library used Zhao, Nasrullah and Li (2019). As the models did not show significant differences in the final result in relation to the hyperparameters for each case, the hyperparameters were kept the same for all analysis.

Table 4.1 – Hyperparameter selected using cross-validation for each model

Model and Hyperparameter	
kNN	n_neighbors=5, method=largest, metric='minkowski'
MCD	assume_centered=False
LOF	n_neighbors=16
CBLOF	n_clusters=6, alpha=0.8, beta=4
OCSVM	kernel='rbf', gamma=0.2, nu=0.7
FB	base_estimator=LOF, n_estimators=10, max_features=1.0, combination='average'
FastABOD	n_neighbors=5
IF	n_estimators=100, max_samples=128
HBOS	n_bins=5, alpha=0.1, tol=0.5
LODA	n_bins=5, n_random_cuts=50

4.4.4 Evaluation metrics

For the fault detection part, as an unsupervised methodology, at the end of the test the anomaly score is calculated, where samples with high anomaly score values are usually anomalies. To verify the performance of the proposed methodology, threshold values were defined based on the training group. For the bearing dataset (Case 1) the threshold was defined based on the assumption that the training group is composed of only signals in normal condition (considering that the initial signals correspond to the start of operation of the bearing). As the gearbox dataset (Case 2) and mechanical fault (Case 3), the contamination ratio is known, its value was used to define the threshold. It is also worth mentioning that, due to the knowledge about the fault characteristics and respective behavior, the user

can adjust the contamination rate of the methodology during the application, based on a preliminary analysis of the training data.

For the static condition, each test was performed 100 times to show the stability of the model. The signals were randomly chosen for the test and training group at each iteration of the model. For the dynamic condition, in each new update of the training group, 5% of the samples were randomly excluded to also assess the stability of the model. As the update occurred more than 400 times in the tested dataset, the complete test was performed 10 times. In addition to the variation of the dataset, each iteration of the model was performed with different random seeds.

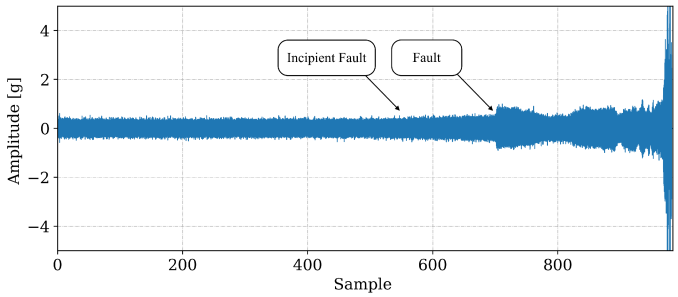
The results are presented using the F1-Score, PR-AUC (Precision-Recall Area Under the Curve) and average confusion matrix of the iterations with respective standard deviations. The metrics were chosen due to the greater interest in correctly identifying samples referring to faults (anomalies). Although it is a problem to have false positives in the final result, failing to acknowledge a fault is even worse as it can result in the machine breakdown. Moreover, these metrics are also important when dealing with unbalanced dataset (common situation in the real scenario).

For the fault diagnosis using the unsupervised classification approach (Case 1 and 2), each sample identified as anomaly is classified in relation to the type / location of the fault. As the classification was performed only for the anomalies identified, accuracy was used as an evaluation metric. For the root cause analysis (Case 3), the feature importance ranking is presented. Kendall Tau distance was used to compare SHAP and Local-DIFFI. The tests were performed using 2.2 GHz Intel Core i7 Dual-Core, 8 GB 1600 MHz DDR3, Intel HD Graphics 6000 1536 MB.

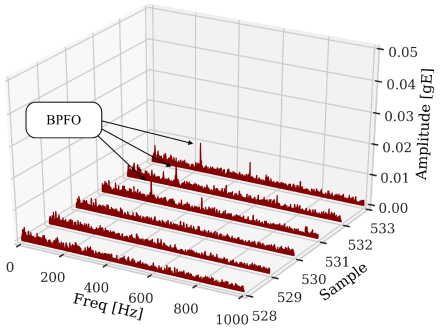
4.5 Results and discussion

4.5.1 Data Exploration

In this subsection the data used in this work for Case 1-3 are analyzed and discussed.



(a) Complete signal for the test in time domain



(b) Waterfall envelope spectrum

Figure 4.3: Bearing dataset signal.

Case 1: Bearing Dataset

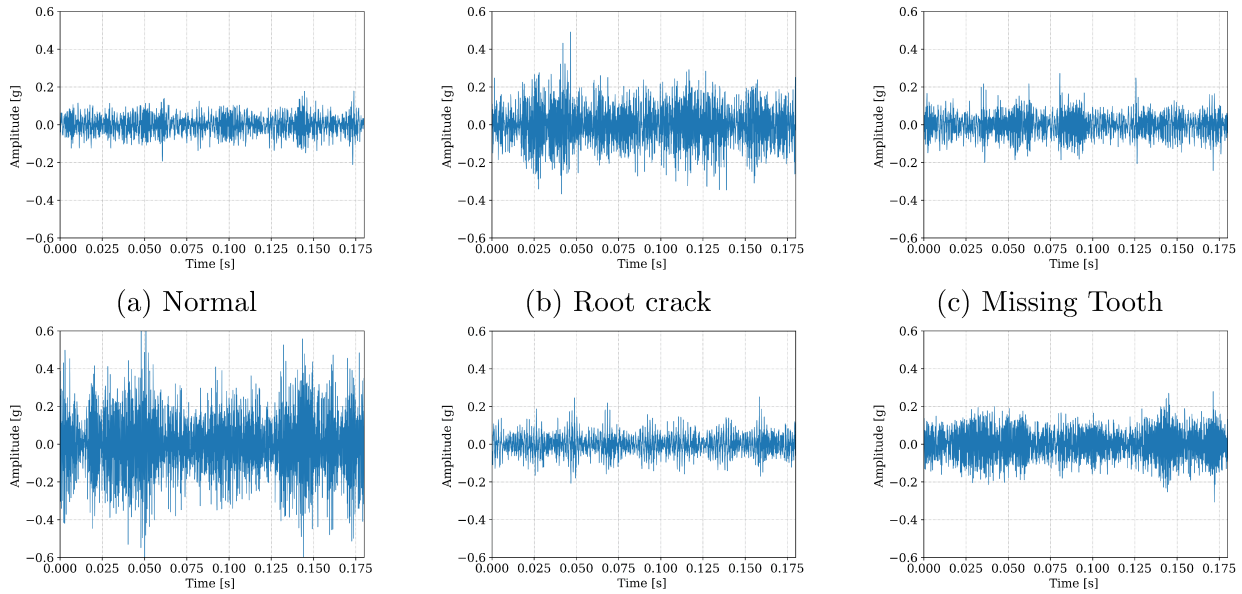
Fig. 4.3a shows the complete signal for the test in time domain. As the signal was not collected continuously (24/7), it was decided to present it according to the sample (x-axis). The point at the incipient fault starts, as well as the fault are identified. In Fig. 4.3a it can be seen that although the fault is easily identified by the signal trend in the time domain, the incipient fault is not easily identified by visual analysis. Making it important to use the ML model with the appropriate features to provide the maintenance team adequate time to schedule an intervention. Fig. 4.3b shows the moment of beginning of the incipient fault presented in the envelope spectrum and used to define the labels of the signals. It is noted that from the sample 531 there is evidence of BPFO, being defined as indicative of incipient fault and, therefore, anomaly. Based on the adopted methodology, all samples after this signal are considered faults (anomalies).

Case 2: Gearbox Dataset

The signals for the different types of faults present in the dataset are shown in Fig. 4.4. Due to the possibility of non-stationarity caused by the variation of the load, it was

decided to show the signals in the time domain. In addition, some defects, such as broken / cracked tooth, can also be better viewed.

It is possible to notice an increase in the energy level in the signal for defects such as root crack, spalling and chipping tip (most severe). On the other hand, differentiating a normal signal from one with a missing tooth or chipped tip in the initial stage is not so simple. Therefore, the feature extraction and the use of artificial intelligence techniques become essential for more assertive monitoring.



(a) Normal (b) Root crack (c) Missing Tooth

(d) Spalling (e) Chipping tip (least severe) (f) Chipping tip (most severe)

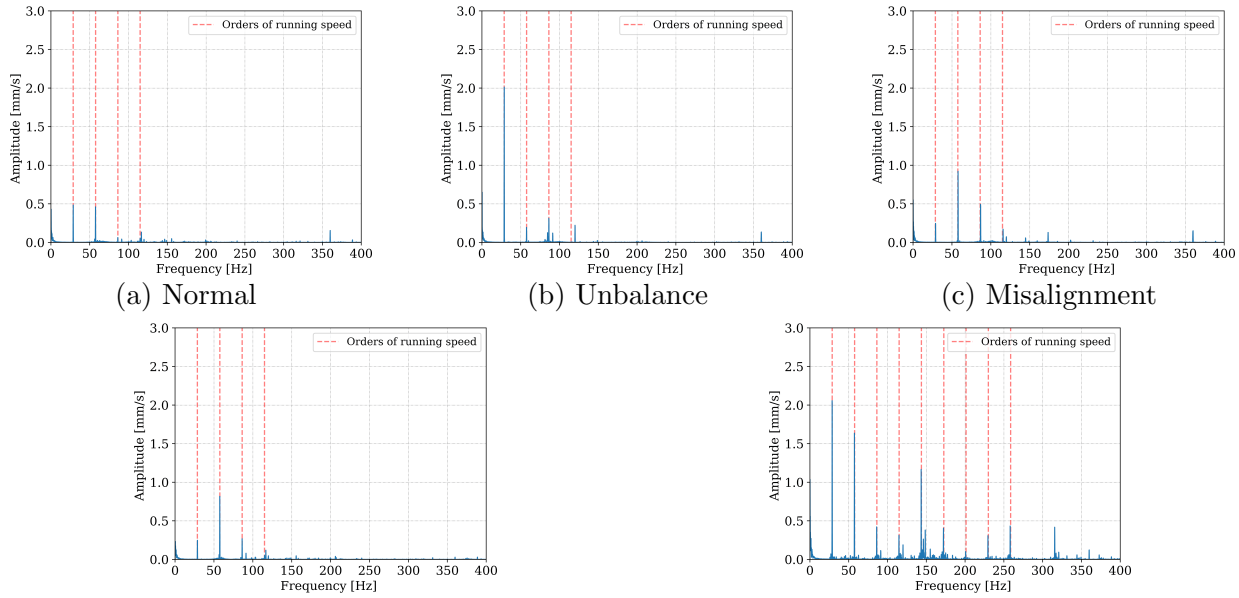
Figure 4.4: Vibration signal examples under different gear health conditions.

Case 3: Mechanical Faults

In Fig. 4.5 some examples of faults are shown. The frequency domain was used to better characterize the faults, knowing that the speed rotation was kept constant.

For the normal situation, there is no predominance of any characteristic frequency, in addition to presenting a low level of vibration in relation to other situations. In the unbalance case, it is evident the increase in energy in 1 x fr, characteristic of the fault. Misalignment and mechanical looseness exhibit very similar behavior in the signal with 2 x fr greater than the other harmonics. The differentiation was performed based on the type of fault inserted in the test rig. For the situation of combined failures (unbalance, misalignment and mechanical looseness) the characteristics of all faults are noted.

For the reasons mentioned above, the classification of such faults includes the analysis of signals in other positions and complementary techniques. Therefore, for this case, the proposed classification methodology will provide only the most relevant features for the identification of the fault, assisting the specialists in the search for the root cause of the problem.



(a) Normal (b) Unbalance (c) Misalignment

(d) Looseness (e) Combined Faults

Figure 4.5: Examples of vibration signals for different faults present in the dataset.

4.5.2 Fault detection: Anomaly Detection

Using the proposed methodology, the results obtained for the fault detection are presented in Table 4.2. Table 4.2 also shows the average time spent for training and testing (a new sample). The top three results for each metric are shown in bold. It can be seen in Table 4.2 for Case 1 and 3, that the models that had better identification of the faults through the F1-Score were: MCD, HBOS and IF. For Case 2, although HBOS had a very close performance, the models that showed the best results were: MCD, kNN and IF.

In order to evaluate the general efficiency of the model, regardless of the defined threshold, the PR-AUC value was calculated. The results show that by modifying the threshold, the models can present even better results. It is noteworthy that the threshold used for comparison and calculation of the F1-Score was defined based on the previous analysis of the training group, simulating a real condition where the data for testing are

Table 4.2 – Fault detection results

Metric	kNN	MCD	LOF	CBLOF	OCSVM	FB	FastABOD	IF	HBOS	LODA	Ensemble
Case 1											
F1-Score	63.21	**99.45**	57.94	60.87	60.78	61.99	88.17	**97.19**	**98.43**	39.12	70.20
	(1.04)	(0.03)	(0.71)	(0.48)	(0.17)	(2.93)	(1.34)	(1.06)	(0.79)	(10.82)	(1.40)
PR AUC	96.45	**99.91**	92.02	93.55	93.28	94.35	98.80	**99.92**	**99.91**	94.23	98.85
	(0.04)	(0.00)	(0.12)	(0.21)	(0.02)	(0.19)	(0.08)	(0.01)	(0.01)	(0.81)	(0.08)
Time [s]	**0.0041**	0.3066	**0.0058**	0.0835	0.0101	0.0555	0.1081	0.3671	**0.0049**	0.0307	1.3435
Case 2											
F1-Score	**99.82**	**99.84**	99.26	99.06	89.68	99.64	97.16	**99.71**	99.49	90.14	99.70
	(0.08)	(0.08)	(0.39)	(0.69)	(2.21)	(0.20)	(1.80)	(0.21)	(0.18)	(6.42)	(0.05)
PR-AUC	99.97	**99.99**	99.98	99.95	99.74	**99.99**	99.88	**99.99**	99.95	99.83	**99.99**
	(0.02)	(0.00)	(0.01)	(0.04)	(0.71)	(0.01)	(0.12)	(0.01)	(0.04)	(0.11)	0.01
Time [s]	0.1706	0.0807	**0.0165**	0.0575	**0.0073**	0.1119	1.0303	0.4121	**0.0101**	0.0434	1.9404
Case 3											
F1-Score	96.27	**98.15**	92.01	94.83	95.62	92.35	95.76	**97.20**	**99.22**	97.16	97.10
	(0.00)	(0.01)	(0.00)	(0.15)	(0.00)	(0.33)	(0.00)	(0.27)	(0.00)	(0.39)	(0.12)
PR-AUC	99.60	**99.91**	98.76	99.38	99.50	98.83	99.51	**99.74**	**99.96**	99.69	99.53
	(0.00)	(0.00)	(0.00)	(0.03)	(0.00)	(0.04)	(0.00)	(0.05)	(0.00)	(0.20)	(0.00)
Time [s]	0.1721	0.6172	**0.0251**	0.1030	**0.0209**	0.1808	0.6227	0.4195	**0.0361**	0.0411	2.2384

not yet available. For this reason, it was decided to present the F1-Score value based on the defined threshold instead of the optimum value that could be obtained by adjusting the threshold in the complete dataset. Nevertheless, it can also be analyzed that, despite the improvement in the results, in general, the models that showed better performance in relation to PR-AUC were the same ones with highest F1-Score value: IF, HBOS and MCD.

Although in general HBOS, MCD and IF presented good results for the three cases, it can be seen that depending on the dataset, other models can obtain better performance, such as kNN in Case 1 and 2. The good results obtained in Table 4.2 for the three cases show that it is possible to detect faults in rotating machinery through the models studied in an unsupervised way.

Among the models with the best results, HBOS presented the lowest computational time. In general, LOF and OCSVM also presented low values. On the other hand, FastABOD, MCD and IF demanded more computational time in relation to the other models (in a general analysis, excluding Ensemble). The low average time for most models in training and testing a sample allow implementation in an industrial environment focused on predictive maintenance.

For the proposed comparison between SHAP and Local-DIFFI, and due to the good overall performance of Isolation Forest, the details of the methodology results are presented for the model. The average values for the confusion matrix are presented in Table 4.3 (the

sample quantities were rounded up because they are integer values). The confusion matrix allows a better visualization of the results in relation to the distribution of the signals in the respective classes. The results are presented both in percentage and in quantity of signals.

The results present in Table 4.3 for Case 1, show that the samples of the normal group

Table 4.3 – Confusion Matrix

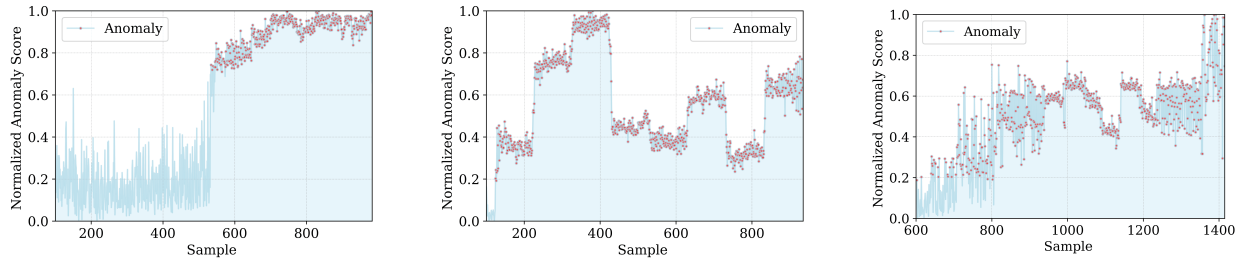
Case 1	Normal[2]	Fault[2]	**Case 2**	Normal[2]	Fault[2]	**Case 3**	Normal[2]	Fault[2]
Normal[1]	48.75 % (0%) 431 (0)	0 % (0%) 0 (0)	Normal[1]	2.63 % (0.23%) 22 (2)	0.24 % (0.11%) 2 (1)	Normal[1]	10.05 % (0.36%) 82 (3)	3.06 % (0.36%) 25 (3)
Fault[1]	2.82 % (1.13%) 25 (10)	48.41 % (1.01%) 428 (9)	Fault[1]	0.24 % (0.23%) 2 (2)	96.89 % (0.35%) 810 (3)	Fault[1]	1.84 % (0.24%) 15 (2)	85.05 % (0.24%) 694 (2)

[1] True Label, [2] Predicted Label

were all correctly classified. The anomalies had an average classification error of 25 samples in a total of 453 anomalies, confirming the good performance of the model. For Case 2, on average, 2 anomalies of 812 were classified incorrectly, and 2 normal samples of 24 were classified as anomalies. For Case 3, 694 of 709 anomalies were classified correctly and 82 of 107 normal samples were also classified correctly. As in Case 1 and 2, the results show the good performance of the model.

Such performances in a real application, will allow not to intervene in the machine unnecessarily (which is also a big problem, considering the need to stop the production and high cost of some components that could be replaced without need). Moreover, the model was able to correctly identify most anomalies, including those at an early stage of fault, allowing the maintenance team to schedule the machine shutdown without directly interfering in the production process.

Keeping in mind that the essence of anomaly detection methods is unsupervised, that is, without defining even the threshold value (in addition to not having labelled data in training), the normalized anomaly scores are presented for the entire test, Fig. 4.6. The anomalies identified in the Fig. 4.6 are presented based on the defined threshold. The x-axis values refer to the test samples only. Fig. 4.6 shows the evolution of the anomaly score with the development of the fault. It is possible to notice the gradual increase near the region identified as the beginning of an incipient defect, sample 531, as shown in Fig. 4.3b. Subsequently, there is an increase in the anomaly score in relation to the normal condition, indicating a permanent change in the behavior of the equipment, and consequently, a fault.



(a) Case 1 - Bearing Dataset. (b) Case 2 - Gearbox Dataset. (c) Case 3 - Mechanical Fault Dataset.

Figure 4.6: Anomaly scores for Isolation Forest.

For Case 2 and 3, the anomaly score value does not show the evolution of the fault, since the analysis was performed in a static way. Analyzing Case 2, it is noted that the scores for samples considered anomalies are higher than the normal group (first 24 samples), enabling identification. The samples were grouped in sequence with respect to the equipment condition to allow comparison between the faults (healthy condition, missing tooth, root crack, spalling, chipping tip 1, chipping tip 2, chipping tip 3, chipping tip 4, chipping tip 5). It can be seen that similar faults with similar condition have similar anomaly scores. Thus, in a real monitoring situation, if the anomaly score changes suddenly, it can be concluded that a possible new fault is occurring or that the severity of the current fault has been accentuated.

For Case 3, due to the presence of different fault conditions in each group, it is not possible to distinguish the type of fault through visual analysis of the anomaly score. However, the variation between normal samples (first 107 samples) and those considered to be fault still differ, even if less than the other cases.

Despite the defined threshold value for comparison of the models, it is possible to notice the difference between normal and faulty samples, even for where the fault is incipient. This difference allows the user to correctly identify anomalies present in the monitoring. The importance of anomalies detected in the incipient period is emphasized, as it is a stage where it is not easily identified by visual analysis or variation of the time signal energy trend (often used as a metric to define maintenance alarm levels based on standards).

The difference in the anomaly score values also prove that the faults studied for rotating machinery behave as anomalies / outliers. In other words, they are samples that have values so different from other observations that they are capable of raising suspicions

about the mechanism from which they were generated Hawkins (1980). As the samples share this basic principle, it is concluded that the use of AI models based on anomaly detection allows to identify faults in rotating machinery in a satisfactory and unsupervised way.

4.5.3 Fault Diagnosis: Unsupervised Classification / Root Cause Analysis

The results obtained using the proposed methodology for the unsupervised classification are presented in the Table 4.4. The definition of the type of fault diagnosis to be used is performed during the extraction of features based on the premise of the features being related to different types of fault or not.

Case 1 and 2 present specific features related to a single type of fault / location, which allow the identification of the type of fault and the location, respectively. Therefore, the proposed Unsupervised Classification can be performed. For Case 1, it can be noted

Table 4.4 – Fault Diagnosis: Unsupervised Classification

Case	kNN	MCD	LOF	CBLOF	OCSVM	FB	FastABOD	IF	HBOS	LODA	Ensemble
Case 1	BSF	BPFO	BPFO	BSF	BSF	**BPFO**	BPFO	**BPFO**	**BPFO**	BPFO	BPFO
Accuracy	39.46	82.23	85.44	4.95	7.74	**95.43**	75.22	**99.57**	**99.38**	88.82	83.80
Std	(7.60)	(1.08)	(1.71)	(1.06)	(0.69)	(0.99)	(1.84)	(0.58)	(0.19)	(3.05)	(3.38)
Time [s]	2.6094	**0.0968**	0.2912	**0.1913**	0.2571	0.9428	4.3812	0.2890	**0.1564**	0.2190	9.7058
Case 2	**1stStage**	1stStage	1stStage	**1stStage**	1stStage	1stStage	1stStage	1stStage	1stStage	1stStage	**1stStage**
Accuracy	**96.47**	58.13	84.90	**93.42**	69.18	91.09	88.91	86.12	86.84	89.83	**96.72**
Std	(0.77)	(12.64)	(2.88)	(2.19)	(3.61)	(3.27)	(3.83)	(3.06)	(3.88)	(8.31)	(1.11)
Time [s]	2.3111	**0.1774**	0.3556	**0.1998**	0.2033	0.9097	7.2525	0.2538	**0.1939**	0.2644	12.6619

that IF, HBOS and FB models had better results. Using IF as an example, in 99.57% of the samples analyzed, the specific feature BPFO was considered the most relevant, and consequently, correctly classifying the type of fault. Analyzing the results obtained in the previous stage of fault detection, IF and HBOS are good models for the methodology, since they showed good ability to detect and diagnose the fault. FB on the other hand, notwithstanding a good result in diagnosis, presented a low fault detection rate, which in this case would fail to identify some anomalies in the equipment. Some models such as CBLOF, kNN and OCSVM classified the fault as BSF instead of BPFO, being considered an error. The other models, despite having correctly classified the type of bearing fault, had a lower hit rate than those mentioned above, both in the fault detection stage and in the unsupervised classification.

For Case 2, the fault was classified in relation to the location in the gearbox. The

most relevant features were associated according to their stage. In other words, using the Ensemble model as an example, in 96.72% of the samples analyzed, the most relevant feature was related to fault in the first stage. The models with the highest hit rate were CBLOF, kNN and Ensemble. The models showed good results for both fault detection and diagnosis. It is worth mentioning that for the dataset under analysis, most models showed good results in detecting faults, possibly because they have well-characterized behaviors. IF and HBOS which presented good results for the fault detection in all cases, showed inferior performance, erroneously classifying approximately 15% of the fault as present in the second stage. In general, the MCD that showed good results for fault detection, was not as effective in the fault diagnosis part.

For Case 3, the features may be related to more than one fault, therefore, it is not possible to perform the unsupervised classification directly. In this case, the general analysis, using the Root Cause Analysis procedure is applied, Table 4.5.

Table 4.5 – Fault Diagnosis: Root Cause Analysis results

Case	kNN	MCD	LOF	CBLOF	OCSVM	FB	FastABOD	IF	HBOS	LODA	Ensemble
Case 3.1[1]	3xfr	2xfr	3xfr	3xfr	1xfr	4xfr	4xfr	1xfr	3xfr	2xfr	1xfr
	33.17	36.29	46.90	36.00	62.60	42.61	49.64	55.83	50.85	41.29	27.22
	(0.00)	(5.10)	(0.00)	(6.07)	(0.00)	(4.47)	(0.00)	(6.49)	(0.00)	(14.92)	(2.54)
Time [s]	1.6272	**0.1075**	0.2681	**0.1389**	0.1981	0.8740	4.3299	0.3905	**0.1301**	0.2256	8.3899
Case 3.2[2]	3xfr	3xfr	3xfr	2xfr	3xfr	2xfr	3xfr	2xfr	3xfr	3xfr	3xfr
	44.09	35.28	49.74	49.84	68.25	34.13	39.80	45.24	61.08	53.00	43.13
	(0.00)	(4.93)	(0.00)	(3.27)	(0.00)	(10.21)	(0.00)	(7.94)	(0.00)	(11.11)	(3.14)
Time [s]	1.3873	**0.1207**	0.2199	**0.1373**	0.2021	0.8537	4.3308	0.3436	**0.1136**	0.1607	8.0897
Case 3.3[3]	2xfr	1xfr	4xfr	2xfr	1xfr	2xfr	2xfr	2xfr	2xfr	1xfr	2xfr
	56.52	47.36	39.78	42.55	66.67	51.26	60.86	45.53	47.82	49.70	25.73
	(0.00)	(0.00)	(0.00)	(13.83)	(0.00)	(27.57)	(0.00)	(19.51)	(0.00)	(19.21)	(7.02)
Time [s]	1.3114	**0.1210**	0.1941	**0.1212**	0.1799	0.8174	3.6984	0.2489	**0.1136**	0.1557	7.1616
Case 3.4[4]	2xfr	3xfr	1xfr	2xfr	2xfr	1xfr	4xfr	2xfr	4xfr	4xfr	1xfr
	66.66	64.28	46.66	38.53	73.33	41.20	66.66	42.93	86.66	43.94	50.00
	(0.00)	(0.00)	(0.00)	(10.77)	(0.00)	(15.18)	(0.00)	(28.14)	(0.00)	(20.13)	(10.34)
Time [s]	1.5150	**0.1303**	0.2225	**0.1144**	0.1740	0.8191	3.8296	0.2812	**0.1198**	0.1586	7.5644

[1] Unbalance, [2] Misalignment, [3] Mechanical Looseness, [4] Combined Faults

For better visualization, the results are presented based on the most relevant feature obtained by the methodology. A sub-division for each type of fault was carried out in order to provide more details on the method. An example of the complete results is presented for the IF and Case 3.1, Table 4.6.

The unbalance fault is presented in Case 3.1 and the results are shown in Table 4.5 and Table 4.6. Due to the unbalance behavior predominantly manifesting in 1xfr, it is expected that this features will show greater relevance for the analysis, as presented in the IF

and OCSVM models. On the other hand, as the features are directly or indirectly related to more than one fault, the model can use the relationship with another feature, instead of what is expected. For example: it is known that unbalance manifests itself in 1xfr, however, if the energy in 2xfr is greater than 1xfr, possibly the sample presents a misalignment (excluding other fault possibilities just for example). Thus, assuming an unbalanced sample, the model can use 2xfr, as a basis to know if it is less or greater than 1xfr and thus 2xfr becomes the most relevant feature, even if the fault is an unbalance. In addition to the aforementioned justification, the type of fault introduced was considered to label the samples. Thus, in some cases the fault behavior was not evident in the signal, which justifies the model to identify other features as more relevant. For example: for a small unbalance, the acquired signal is considered to be unbalanced, even if it does not significantly increase the amplitude in 1xfr.

Table 4.6 shows that in 55.83 % of the samples, 1xfr was classified as the most relevant feature. Subsequently, the features 2x and 3xfr are the most important. Such features are related to the way of identifying an unbalance in a vibration signal, and therefore they can be used by the specialist to analyze the root cause of the fault. It is also noted that the 4xfr feature in most cases was classified as less relevant, since the feature (for the case under study) is not so important for identifying or distinguishing this fault. In Case

Table 4.6 – Fault Diagnosis: Root Cause Analysis full ranking

Feature/Position	1^{st}	2^{nd}	3^{rd}	4^{th}
1xfr	55.83 (6.50)	20.16 (1.86)	16.00 (4.92)	8.01 (3.66)
2xfr	15.97 (3.20)	45.89 (10.14)	27.83 (9.09)	10.31 (2.82)
3xfr	19.50 (5.40)	26.87 (9.57)	35.88 (9.54)	17.75 (4.78)
4xfr	8.70 (2.32)	7.08 (1.86)	20.29 (4.07)	63.93 (5.08)

3.2 the misalignment is presented. Usually this type of fault is identified by the analysis of 2xfr and 3xfr. All models presented, as the most important feature, the same one used by the human specialist.

The mechanical looseness, Case 3.3, as well as the combination of faults, Case 3.4, can present energy in all extracted features. Thus, the variation of the features selected by each model is acceptable, since all features are relevant. It is noteworthy that the variation between the models is due to the different approaches present in each algorithm.

The importance of root cause analysis is to eliminate features that are not relevant

to the analysis, helping the specialist to identify the problem. Thus, in an application where different features and faults are present, the methodology provides a better direction to the specialist about the current fault.

The standard deviation presented in some analysis in Case 3, can be justified by the random selection of samples at each iteration. As mentioned earlier, in addition to the different possibilities of the model in relating the features to the faults, the samples can also present different behaviors, even within the same type of faults, resulting in different selected features. As some models have stochastic behavior, they are more sensitive to variation. Note that for Cases 1 and 2, where there are no major variations in relation to the type of fault, the models have low standard deviations.

The models with the lowest computational costs for the fault diagnosis methodology were: CBLOF, HBOS and MCD, with HBOS being one of the fastest also for fault detection.

Through the explainability of the artificial intelligence models used, it can be concluded that the proposed methodology is able to assist the specialist in identifying the root cause of the problem or even to classify the type of fault present in the equipment in an unsupervised way.

4.5.4 XAI: SHAP and Local-DIFFI

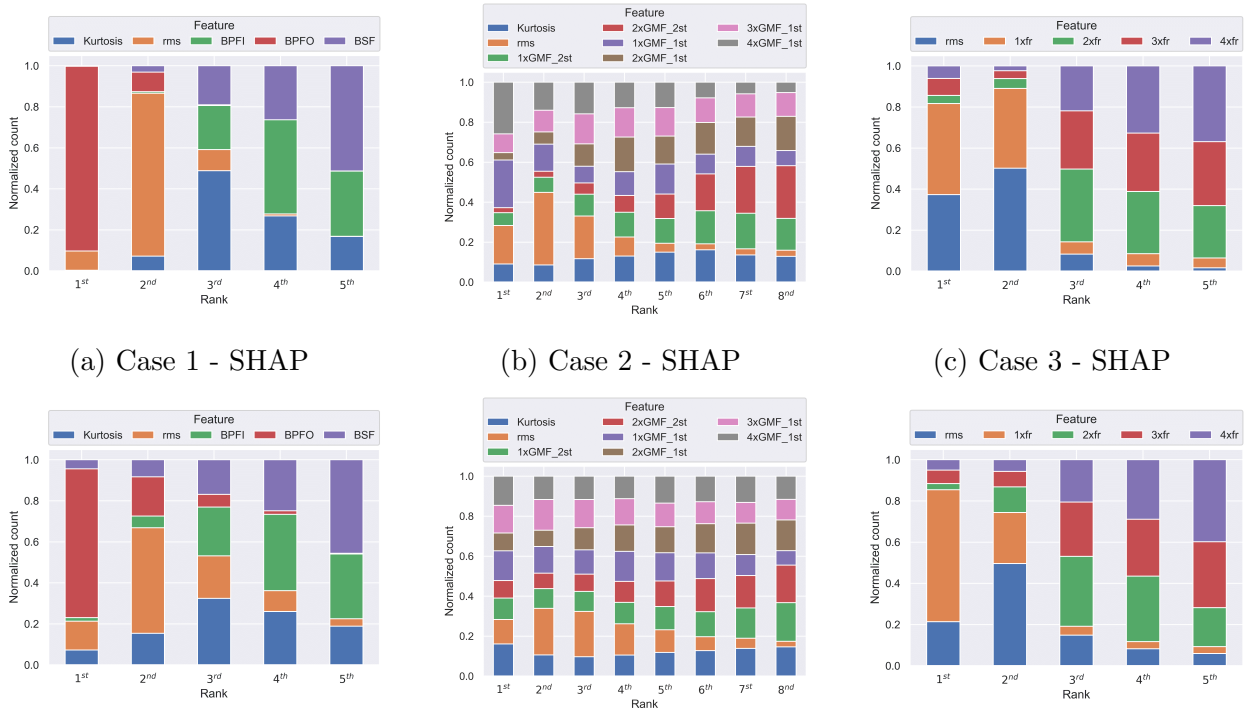
As presented in the methodology, the unsupervised classification/root cause analysis is performed through the ranking of importance of the specific features obtained by the model's explainability. To study the possibility of the methodology in working with different explainable models and the feasibility of implementing a computationally faster model, in Table 4.7 is shown a comparison for the complete relevance rankings obtained by SHAP and Local-DIFFI. As the main goal is to compare the two methods, the values for Case 3 were calculated for all faults. From Table 4.7, the time taken to perform the explainability was higher using SHAP than Local-DIFFI. As a model-specific, Local-DIFFI presents a superior performance of approximately 6.5-8.0x in relation to SHAP, being extremely relevant in applications where the execution time is essential.

The comparison made through Kendall-Tau distance shows that the models have

Table 4.7 – XAI: SHAP vs. Local-DIFFI

Metric/Case	Case 1	Case 2	Case 3
Kendall-Tau Distance	0.348	0.127	0.455
SHAP: Time [s]	0.2890	0.2538	0.3012
Local-DIFFI: Time [s]	0.0361	0.0365	0.0453

similarities in the rankings of relevance, visually presented in Fig. 4.7. Since the main objective is to compare the two models, all the features used by the models for fault detection are considered, without excluding the general features proposed in the application of the fault diagnosis part. It can be seen in Fig. 4.7, for Case 1, that the most relevant feature for both



(a) Case 1 - SHAP (b) Case 2 - SHAP (c) Case 3 - SHAP

(d) Case 1 - Local-DIFFI (e) Case 2 - Local-DIFFI (f) Case 3 - Local-DIFFI

Figure 4.7: SHAP and Local-DIFFI feature importance ranking.

models is precisely BPFO, allowing the unsupervised classification to achieve good results. For Local-DIFFI, some samples presented BPFI and BSF as the most important specific feature, leading the methodology to misclassify the type of fault. Case 2 presented the lowest relationship between the rankings. Among the most relevant features, SHAP showed less occurrence of the features related to the second stage than Local-DIFFI, leading the model to make fewer errors during the application of the proposed methodology. Despite the minor similarity, the main feature (1xGMF_1st) was also the same in both models. For Case

3, the most relevant feature for both models is 1xfr with a good similarity for positions 3, 4 and 5. Thus, through the analysis of the Kendall-Tau distance, it is possible to verify that the rankings show similar behaviors. Because it is a model-specific method, Local-DIFFI is subject to noise due to the stochasticity of the IF, which can reduce its result. Finally, the choice of the explainability model to be used is based on a trade-off between response time and precision.

4.6 Conclusions

This chapter proposes a new approach for fault detection and diagnosis in rotating machinery. A three-stage scheme is adopted 1) Feature extraction; 2) Fault detection: Anomaly Detection; 3) Fault diagnosis: Unsupervised classification / Root cause analysis. The vibration features in the time and frequency domains were extracted based on human knowledge already available. In the fault detection, the presence of fault was verified in an unsupervised manner based on anomaly detection algorithms. Finally, in fault diagnosis, through the feature importance ranking obtained by the model's explainability, the fault diagnosis was performed, being: unsupervised classification or root cause analysis.

The results show that the proposed methodology allows the unsupervised fault detection in rotating machinery. And, in addition to providing explainability about the models used, the methodology provides relevant information for root cause analysis, or even unsupervised fault classification.

In general, the results obtained can be summarized as follows: i) for anomaly detection, in Case 1, a maximum F1-Score value of 99.45% was obtained, with the three main models being MCD, HBOS and IF. In Case 2, a maximum value of 99.84% was obtained, and the best models were: MCD, kNN and IF. In Case 3, the maximum value obtained was 99.22%, and the main models being: HBOS, IF and MCD. ii) For the unsupervised classification, in Case 1 a maximum accuracy of 99.57% was obtained using IF, and having HBOS and FB as main models. For Case 2, a maximum value of 96.72% was reached, with the Ensemble, kNN and CBLOF models having the best results. For the root cause analysis, the features that most correlate with the fault under study are presented, as for example in the case of unbalance, where 55.83% of the samples indicated 1xfr as the most important

features, in line with the fault behavior, which helps the specialist to eliminate features that are not related to the fault. iii) The comparison between Local-DIFFI and SHAP showed greater similarity for all cases, especially for Case 3, with a Kendall-Tau Distance value equal to 0.455. In terms of computational performance, Local-DIFFI presented a superiority of approximately 6.5-8.0x over SHAP.

Different state-of-the-art ML algorithms in anomaly detection were studied showing the possibility to change models according to the dataset. The new approach can be applied to different types of faults just by modifying the extracted features associated with a potential fault as shown for the 3 datasets studied. Since the approach does not require previously labeled data, and only knowledge currently available on fault detection through vibration analysis, the methodology has many possible industrial applications. Future work will focus on domain adaptation and transfer learning associated with methods for model interpretability to improve the applicability of the proposed approach in different industrial scenarios.

CHAPTER V

AN EXPLAINABLE ARTIFICIAL INTELLIGENCE APPROACH FOR FAULT DIAGNOSIS BASED ON TRANSFER LEARNING FROM AUGMENTED SYNTHETIC DATA TO REAL ROTATING MACHINERY

Due to the growing interest for increasing productivity and cost reduction in industrial environment, new techniques for monitoring rotating machinery are emerging. Artificial Intelligence (AI) is one of the approaches that has been proposed to analyze the collected data (e.g., vibration signals) providing a diagnosis of the asset's operating condition. It is known that models trained with labeled data (supervised) achieve excellent results, but two main problems make their application in production processes difficult: (i) impossibility or long time to obtain a sample of all operational conditions (since faults seldom happen) and (ii) high cost of experts to label all acquired data. Another limitating factor for the applicability of AI approaches in this context is the lack of interpretability of the models (black-boxes), which reduces the confidence of the diagnosis and trust/adoption from users. To overcome these problems, a new generic and interpretable approach for classifying faults in rotating machinery based on transfer learning from augmented synthetic data to real rotating machinery is here proposed, namelly FaultD-XAI (Fault Diagnosis using eXplainable AI). Synthetic vibration signals are created by mimicking the characteristic behavior of faults in operation. To enable transfer learning, the real machine signal is used as a reference. The data are augmented by increasing the number of samples for training the model used, 1D Convolutional Neural Network (1D CNN). With the signals in the frequency

domain, the model is trained, and the classification performed. Subsequently, the application of Gradient-weighted Class Activation Mapping (Grad-CAM) allows the interpretation of results, supporting the user in decision making. The proposed approach not only obtained promising diagnostic performance, but was also able to learn characteristics used by experts to identify conditions in a source domain and apply them in another target domain. The experimental results obtained on three datasets containing different mechanical faults suggest the method offers a promising approach on exploiting transfer learning, synthetic data and explainable artificial intelligence for fault diagnosis. Lastly, the dataset developed by the authors will be available.

5.1 Introduction

Currently, with the great need to increase the amount of final product manufactured, the industry has been looking for ways to monitor its assets in order to avoid unexpected breaks that can directly impact production Brito et al. (2022). Due to the growing demand, searches for alternatives in the monitoring of rotating machinery have been commonplace, leading to a large amount of information and research being generated: from the definition of informative signals to the development of smart data processing techniques, from new sensors to new best practice in data acquisition and monitoring.

The increase in information generates a need for experts to achieve quick analysis and effective diagnosis. Thus, artificial intelligence models have been proposed to assist in the diagnosis and monitoring of assets.

Kumar and Hati (2020) present a review of the main Machine Learning (ML) and Deep Learning (DL) techniques applied in the monitoring of induction motors, aiming to detect faults such as: broken bars, bearings failures, stator failures and eccentricity. Liu et al. (2018) also present a review on artificial intelligence for fault detection in rotating machinery, in which more than 100 cited references refer mostly to fault classification (fault diagnosis). More recently, Lei et al. (2020) presented a comprehensive review with more than 400 citations, focused on Artificial Intelligence (AI) applications for fault diagnosis, providing a historical overview, in addition to current developments and future prospects. Further details on the state-of-the-art in Artificial Intelligence (AI) for fault diagnosis in

rotating machinery can be found in Hoang and Kang (2019); Duan et al. (2018); Khan and Yairi (2018); Zhao, Yan, Chen, Mao, Wang and Gao (2019); Saufi et al. (2019).

By analysing the aforementioned studies, it can be noted that, for the task of monitoring rotating machinery, AI models have high accuracy rates when trained in supervised condition, ie. in the presence of labeled data Lei et al. (2020) representing both normal and faulty conditions. Unfortunately, the supervised scenario is difficult to be applied in real world industrial settings, for two main reasons Yang et al. (2019):

i) *Faults rarely happen in real world*: by design, equipment are made to be ideally operating all the time in a normal condition, therefore not providing sample labels for faulty conditions. When a fault manifests itself, the MCP (Maintenance Planning and Control) team schedules as soon as possible the corrective maintenance for repair or replacement, which results in a low number of acquired signals of fault conditions. In most conventional AI models, unbalanced datasets are associated with many issues, like for example low capability of prediction the least rapresented class, as the model tends to learn more about the condition that is most presented to it;

ii) *High cost to obtain and label data*: with the development of AI approaches in the industry, more complex and sophisticated models are being used, like DL approaches; this is due to the fact that DL techniques can, among other things, automatically extract features from raw signals and can achieve high performance. This reduces the expert's intervention in feature engineering, a fundamental process for using conventional machine learning models. However, for successful training, such models require great amount of data, which is time consuming to obtain or not always feasible. To label the acquired data, experts are typically required: although the vast majority of faults have their characteristic behaviors known, the analysis requires specific knowledge that is not always available within the team.

Fortunately, *transfer learning* Li, Zhang, Qin and Estupinan (2020) approaches can overcome such weaknesses by applying the knowledge learned from one task (or multiple tasks) to new, related, ones. Pan and Yang (2010); Lei et al. (2020); Zhuang et al. (2021) present some approaches developed with a focus on transfer learning, showing that transfer-learning theories can alleviated the lacking labeled samples and enhancing the applications of Intelligent Fault Diagnosis (IFD) in industry, in such works several approaches are considered,

namely: feature-based approaches Yang et al. (2019), generative adversarial network (GAN) based approaches Guo et al. (2019), instance-based approaches Shen et al. (2015), and parameter-based approaches Hasan and Kim (2018). Despite the promising studies, the vast majority of approaches are based on obtaining at least the Source Domain (samples on which the models will be trained) with real and labeled signals, which does not fully solve the aforementioned problems.

In order to reduce the need for labeled real data, a few recent studies propose the use of synthetically generated data from numerical and physical models Chen et al. (2013); Khan et al. (2021); Handikherkar and Phalle (2021); Gecgel et al. (2021); Sobie et al. (2018). Among the studies, stands out the methodology presented in Sobie et al. (2018). In the study, a simlutation-driven machine learning for bearing classification is proposed where training signals were generated from high resolution simulations in Siemens LMS Imagine. Lab Amesim simulation software using the one-dimensional 3-DOF model Sassi et al. (2007). As with the other works, despite the satisfactory results obtained, the approaches are limited in terms of adjustment of boundary conditions, element properties, definition of solvers, model complexity, and application to only one type of failure (in this case, bearing fault), which can limit large-scale industrial applications.

Similarly to the problem of obtaining a complete and labeled training set, the adoption of AI models in the industry comes up against its interpretability Li, Zhang and Ding (2019). Called Explainable Artificial Intelligence (XAI), the area has been showing great interest by researchers, and is identified as one of the solutions to bridge the gap between AI researches and engineering applications Lei et al. (2020). Many models are considered as black boxes, that is, the user responsible for receiving the diagnosis does not know how the model reached the final conclusion. By not knowing exactly how or which features of the signal the model was based on for decision making, the reliability of the diagnosis is compromised, implying even the non-use of the model. To solve this problem, studies have been developed to explain why the model obtained the final classification. Among the techniques that provide visual explanation, the state-of-the-art technique in post-hoc methods is Gradient-weighted Class Activation Mapping (Grad-CAM). As faults in IFD using vibration are generally identified through a visual analysis of the signal in the frequency

domain, it is interesting to provide a heatmap overlaid on the input signal, identifying the most relevant frequencies for the classification. Despite the wide application in other areas, at present, Grad-CAM algorithm is seldom used in AI models for fault diagnosis Feng et al. (2020); Lin and Jhang (2021); Yu et al. (2022); Saeki et al. (2019).

To overcome the problems presented, a new interpretable approach for classifying faults in rotating machinery based on transfer learning from augmented synthetic data to real rotating machinery is proposed. The approach starts from the concept where, transfer learning models focus on storing knowledge gained while solving one problem and applying it to a different but related problem. However, instead of working on fine-tuning, or retraining the model, the approach focuses on synthetic creation of the source domain: in this way the trained model learns features, which allow the transfer knowledge through the shared features of the source and target domain.

In the proposed approach, called FaultD-XAI (Fault Diagnosis using eXplainable AI), we propose to train the AI model with synthetic signals. The signals are generated from the knowledge about different faults in rotating machinery combined with the original signal of the machine under analysis, avoiding the use of complex mechanical models. The original signal is used only as a reference to create the synthetic signals, not requiring samples of all possible failures, which solves the problem of lack of labeled data in training set and enables engineering applications. To increase the variability of the training dataset, and consequently the robustness of the AI model, signals are generated using data augmentation techniques. To avoid the need to perform featuring engineering, reduce the computational cost, and enable the model's interpretability through the use of Grad-CAM, 1D Convolutional Neural Network (1D CNN) is used. Vibration signals in frequency domain are used, bringing the analysis as close as possible to that performed by experts. Due to the importance of rotors, bearings and gears for rotating machinery and especially their respective faults, three datasets are used for validation of the methodology, each one being related to a respective component and its possible faults.

The use of synthetic data makes it possible to apply the AI model in different rotating machinery even without having real signals in all operating conditions. Such ability can be seen as a transfer learning approach, since the proposal can be used in any type

of machine, just by modifying the reference signal. In addition, the model is trained with a dataset belonging to the source domain, which, in turn, presents a difference from the signals present in the target domain. The data augmentation of the data allows the model to be implemented quickly, since few collections are enough to generate the training data. Finally, the interpretability of the approach allows the user to have reliability and confirm the diagnosis, enabling its implementation, and even, for the person responsible for the development of the model, to verify the coherence of the learning during the training phase, allowing adjustments if necessary.

The main contributions of FaultD-XAI are: i) a new transfer learning classification approach based on a synthetic dataset, without the need to have signals of real fault conditions; ii) possibility of interpreting the way in which the final result was obtained by the model, supporting decision making (a new contribution to the study of XAI in fault diagnosis); iii) a generic and simple way to generate synthetic fault data for training, based on the knowledge available in vibration analysis, without the need for complex models; iv) possibility of generating varied training datasets of different sizes (data augmentation - targeting deep learning applications); v) new dataset, publicly available, to study failures such as: unbalance, misalignment and looseness; vi) possibility of application in different types of faults; vii) faster deployment of the model in production; viii) industrial application on real world datasets.

The remainder of this chapter starts with a brief explanation about the 1D CNN and Grad-CAM. The proposed approach is presented in Section 3. Experimental procedure is shown in Section 4. Results and discussion are given in Section 5. Finally, Section 6 concludes this chapter by drawing conclusions and discussing potential future works.

5.2 Background

5.2.1 1D Convolutional Neural Network (1D CNN)

Convolutional Neural Networks (CNNs) have become the de facto standard for various Computer Vision and Machine Learning operations. CNNs are feed-forward Artificial Neural Networks (ANNs) with alternating convolutional and subsampling layers Kiranyaz

et al. (2021). The high efficiency of CNNs and the ability to extract features from the raw signal without the need for feature extraction or engineering make them basic tools in many 2D applications such as images and videos. The fact of being able to extract features automatically from the raw signal, avoids a lot of analysis and signal processing that need to be done when working with other techniques, such as those shown in Brito et al. (2021).

Recently, several research works involving vibration signals and deep 2D CNNs were carried out, where researchers adopted signal processing strategies to convert the original 1D signal to 2D, and thus be able to use the architecture Lu et al. (2017); Ding and He (2017); Appana et al. (2017); Li, Liu, Tang, Lu and Hu (2017). Despite the good results obtained, such techniques need to perform a dimension modification and can considerably increase the computational cost, in addition to the one already present in 2D CNN, making the applications sometimes unfeasible.

To solve this problem, 1D CNNs were proposed Kiranyaz et al. (2016); Avci et al. (2018); Abdeljaber et al. (2017); Ince et al. (2016); Kiranyaz et al. (2018), and quickly became state-of-the-art in applications such as: biomedical data classification and early diagnosis, structural health monitoring, anomaly detection and identification in power electronics and electrical motor fault detection. In the area of fault detection in rotating machines, some studies involving 1D CNN are Zhang et al. (2017); H.Chen et al. (2019); Yibing et al. (2019); Wang, Mao and Li (2021); Chuya-Sumba et al. (2022); Eren et al. (2019).

Among the main advantages of using 1D CNNs, the following stand out: i) significant computational cost reduction; ii) more compact networks (with 1-2 hidden CNN layers and configurations with networks having less than 10K parameters Kiranyaz et al. (2021)); iii) possibility of training on any CPU; iv) possibility of being used for real-time and low-cost applications due to its low computational cost.

Basically, the configuration of a 1D CNN consists of: i) Hidden CNN and Multilayer perceptron (MLP) layers/neurons; ii) Filter (kernel); iii) Subsampling factor; iv) Pooling and activation functions.

The input layer receives the raw 1D signal, while the output layer is an MLP layer with a number of neurons equal to the number of classes. A kernel function moves in one direction: it first performs a sequence of convolutions, the sum of which is passed through

the activation function, followed by the sub-sampling operation. The features are extracted, and used by the MLP layer to perform the classification. For more details on how 1D CNN works, please refer to Kiranyaz et al. (2021).

5.2.2 Gradient-weighted Class Activation Mapping (Grad-CAM)

Grad-CAM Selvaraju et al. (2017) is an improvement on traditional CAM Zhou et al. (2016). Studies show the use of the method in the monitoring of rotating machinery, as in Lin and Jhang (2021) that used 1D CNN and 1D LeNet-5[1] (I1DLeNet) to prove the efficiency of the proposed method SBDS (smart bearing diagnosis system), which uses a gradient diagnosis -weighted class activation mapping (Grad-CAM) - based convolutional neuro-fuzzy network (GC-CNFN). Proving that the method is not only capable of correctly classifying bearing faults, but also helping the user to understand the result.

Kim and Kim (2020) used Grad-CAM and Acoustic Emission Signals to detect bearing faults, again providing the user interpretability regarding the frequencies used by the model to obtain the result. Feng et al. (2020) applied the methodology in the monitoring of bearings and validated it in a gear fault dataset, showing that it is possible to use the methodology to identify the most attentive part of the model in relation to each type of fault. Yu et al. (2022) used Grad-CAM and eigenvector-based class activation map (Eigen-CAM) to interpret the ResNet06 (a popular CNN architecture He et al. (2016)) in 4 databases, 3 bearing and 1 gearbox dataset. In order to interpret the effectiveness of the method, Grad-CAM is applied to localize the regions in the input that contribute the most to the network's prediction. The proposed method is validated by a motor bearing dataset and an industrial hydro turbine dataset Wu et al. (2021).

Grad-CAM uses the gradients of any target concept flowing into the final convolutional layer to produce a coarse localization map highlighting the important regions in the image for predicting the concept Selvaraju et al. (2017). The signal is forward propagated by the CNN part of the model and then processed to obtain the raw score for the category. Gradients are reset for all classes except the class under analysis, set to 1. The signal is then backpropagated to the rectified convolutional feature maps of interest, which is combined

[1]LeNet is classic CNN architecture Lecun et al. (1998)

to generate the heatmap. The highlighted points (i.e., the highest gradient value) are the most relevant regions that the model uses to make the decision. For more details about Grad-CAM please refer to Selvaraju et al. (2017).

5.3 Proposed approach

The proposed approach is presented in Fig. 5.1, and consists of six parts: i) Data Acquisition; ii) Signal Generation; iii) Data Augmentation; iv) Signal Processing; v) Fault Diagnosis; vi) Explainable Artificial Intelligence (XAI). Initially the real signals (x_r) are collected from the machine. To create different operating and fault conditions, synthetic signals (x_s) imitating fault characteristics are generated and then combined with real signals (x_r). To reduce the amount of original data needed for training the model, the signals are augmented ($x_r x_s + x_a$). The fourth step consists of the last phase of data preparation, where the signals are converted from the time domain to frequency domain. After pre-processing the data, the model is created and trained, enabling the diagnosis of the machine's condition (fault diagnosis). Finally, the model is explained, identifying the most relevant frequencies used for classification, supporting the user's decision making.

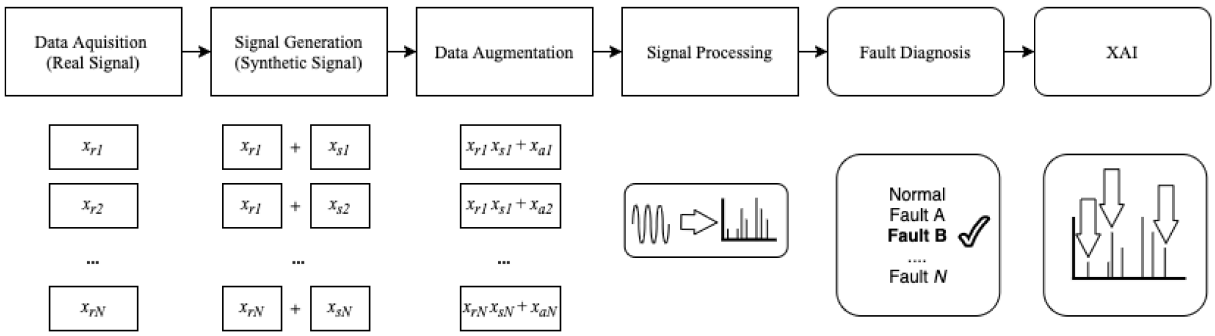


Figure 5.1: General framework of the proposed methodology.

5.3.1 Data Acquisition

Among the sensors used for monitoring rotating machinery, the vibration-based diagnostic method is the most popular and researched. The interest is justified by the fact that the vibration signals directly represent the dynamic behavior of the equipment Ciabattoni et al. (2018); Samuel and Pines (2005); Dalvand et al. (2017); Wei et al. (2019) and are a non-invasive technique Brito et al. (2022). Therefore, vibration signals were used.

The real signal is fundamental to generate the synthetic dataset, enabling transfer learning between the source domain (synthetic signals) and target domain (real fault signals). Through the real signal of the machine under analysis, it is possible to generate a dataset that presents similar characteristics, favoring the learning of the model. It is worth mentioning that the real signal used is the current operating condition of the machine, not requiring a sample of each condition.

5.3.2 Signal Generation

Engineering applications involve a major limitation regarding the data available for training the AI model. In the vast majority of cases, we do not have real and labeled samples of all possible operating conditions, since faults seldom happen. In applications where data is available, it takes a lot of time and cost with experts to label and analyze all signals. A strategy to overcome this problem is to create the training dataset synthetically, since the vast majority of faults in rotating machines have known characteristic patterns.

Different approaches can be used to model faults in rotating machinery, such as: Digital Twin, where through equations of motion, numerical models, optimization and support software, the machine is created virtually, and through modifications in its conditions it is possible to verify the signal response, identifying a fault or a new operational condition.

Despite the available approaches, modeling a rotating machinery is not always an easy task or a low computational cost, which can make some industrial applications unfeasible. Therefore, we chose to model the faults, in a simple way, through the equation of a waveform, with a sine function, considering that the faults are mostly related to deterministic frequencies or characteristic behaviors in the signal.

In general, the most common faults found in the industry are related to bearing defects, classic problems such as unbalance, misalignment, looseness, as well as problems in gears, and for this reason they were chosen for analysis of the proposed approach. It is worth mentioning that the way the approach is proposed, other faults can be modeled and added, without changing its functioning (e.g., cavitation, electrical defects, oil whirl and oil whip etc). It is up to the specialist to determine the synthetic faults that will be created referring to the machine under analysis. It is also important to highlight that the methodology applies

to stationary operating conditions, considering that the faults are created as a function of certain frequencies, which are mostly associated with the rotation frequency of the machine.

The representation of the oscillatory signal *x(t)* used to generate the faults (unbalance, misalignment, looseness and gear fault) is defined Equation 5.1.

$$x(t) = A\,sin(2\pi\,ft + \theta\,) \tag{5.1}$$

Where A indicates the peak amplitude of the signal, f frequency [Hz], t time vector [s] and θ phase [rad]. As accelerometers and no tachometer are used, the phase reference was not applied, although it can help in the identification of some faults.

The amplitudes were randomly generated so that the synthetic fault signal presented an approximate variation of at least 3 dB in the characteristic frequencies of the fault in relation to the normal signal. Since, according to ISO (2009) and the experience of experts in vibration analysis, a variation of less than 3 dB does not represent a significant change in the behavior of the machine in terms of predictive maintenance or failure. The amplitudes are randomly generated to represent different stages of fault, and improve the robustness of the AI model (greater variability in the training group). It is worth mentioning that, depending on the fault, minor variations can be significant, and it is up to the specialist to determine when assembling the model.

The unbalance occurs predominantly at 1 x fr (Rotation Frequency), so a signal was generated with a frequency equal to the rotation frequency. It is worth mentioning again that the phase is not being taken into account to differentiate the faults, only the characteristic frequencies of the signal.

Misalignment presents high vibration values at 1x, 2x and 3x rpm, and can be angular, parallel or mixed (when the two are combined). In that case, Eq. (5.1) was used three times, being the final signal the sum of each equation with a respective frequency (1x, 2x and 3x rpm).

The looseness caused by loose pillow block bolts, cracks in frame structure or in bearing pedestal, or by improper fit between components parts, causes harmonics and

sub-harmonics of the rotation frequency. The signal will present multiples of the rotation frequency as: 1x, 2x, 3x, ... Nx and sub-harmonics: 0.5x, 1.5x, 2.5x, ... $N/2$x, where N is the multiple of the rotation. As well as the misalignment, the final signal was composed of the sum of harmonics and sub-harmonics.

Gear faults are usually related to Gear Mesh Frequency (GMF) and its harmonics. GMF is the product of the number of teeth on the gear (n_{teeth}) multiplied by the running speed of the gear (g_{fr}), described in Equation 5.2.

$$GMF = n_{\text{teeth}} * g_{\text{fr}} \tag{5.2}$$

Defects such as wear, misalignment or eccentricity and backlash in gears may not only cause a significant increase in 1x GMF. In many cases, there is an increase in the amplitude of its harmonics as 2x GMF and 3x GMF. Such defects not only cause variation in the GMF and its harmonics, but in general, also present sidebands of the rotating speed of the defective gear. Therefore, the faults were simulated using Eq.(5.2) where, the frequency for each signal is given by GMF and its harmonics (1x, 2x, 3x GMF) and by the sidebands (GMF +/- rotating speed of the defective gear). The final signal was composed by the sum of each signal generated.

The faults in rolling element bearings are related to their components: inner race, balls or rollers, cage and outer race. Deterioration will cause characteristic frequencies in the signal allowing fault identification. The four frequencies are: i) BPFO (Ball Pass Frequency Outer); ii) BPFI (Ball Pass Frequency Inner); BSF (Ball Spin Frequency); iv) FTF (Fundamental Train Frequency). The calculation basically takes into account: the number of elements, internal and external diameter, contact angle, and its equation can be easily found in vibration analysis software, supplier catalogs and existing literature.

High frequency resonances between the bearing and the response transducer are excited when the rolling elements strike a local fault on the outer or inner race, or a fault on a rolling element strikes the outer or inner race Randall and Antoni (2011).

The impact frequency depends directly on the machine components and their

natural frequencies, however, as the interest is to evaluate the characteristic fault frequencies, this value can be arbitrarily defined in the signal modeling. The impact signal can be modelled using Eq.(5.1) where A indicates the peak amplitude of the signal, f impact frequency [Hz], t time vector of impact [s]. To make the impact periodic, it is convoluted with a Comb function, where the non-null values (periodic peaks) correspond to the fault frequencies of the elements. The Comb is a sum of time shifted Dirac Delta, that is, it is defined to be zero at alternate time points.

After creating the synthetic faults, the original signal is added to represent the dynamic behavior of the machine, allowing the transfer knowledge. Due to the types of faults under study, at the end of the process, each original signal will result in 7 signals, being: normal condition, outer race fault (BPFO), inner race fault (BPFI), unbalanced, misaligned, mechanical looseness and gear fault (due to their common usage, the terms BPFO and BPFI will also be used to identify the type of fault under analysis, that is, fault in the outer and inner races, respectively.). Although the inner race fault (BPFI) was not present in any of the cases studied, it was introduced to increase the complexity of the training and also validating the approach.

5.3.3 Data Augmentation

Data augmentation is a widely used practice in ML when it is desired to increase the relevance of the dataset under study, mainly for image classification, natural language understanding, semantic segmentation and also in fault diagnosis Khan et al. (2021); Li, Zhang, Ding and Sun (2020); D.S.Alves et al. (2020); Tang et al. (2020). In addition to increasing the amount of data for training, it is possible to insert small variations that, do not change the general context of the sample (although they exist). Some popular techniques are: flip, rotation, scale, crop, translation, gaussian noise. This makes the model able to correctly classify the sample, even if it presents noise, translations, different size, lighting etc. Performing augmentation can prevent the model from learning irrelevant patterns and consequently improve overall performance.

Just as the image of an object can be obtained from different perspectives by varying the context, but not the object under analysis, the same can be observed in the

vibration signal of a rotating machinery. In an industrial plant, machines are subject to process variations, human interference, disturbances from nearby machines that can introduce extra information into the signal. Therefore, a robust classification model should be irrelevant to such variations, focusing only on identifying the operating condition of the asset under analysis.

In addition to improving data variability, data augmentation contributes to increasing data in the training set, enabling the implementation of more complex and deep learning models.

For the proposed approach, each synthetic signal is augmented, randomly varying the method variables. The methods presented in Li, Zhang, Ding and Sun (2020) were used. It is worth mentioning that other methodologies can also be used. The proposed methods are:

i) Gaussian Noise: apply random noise to the raw signal, Equation 5.3.

$$\bar{x} = x + \alpha_{\text{gauss}} G,\ G \sim N(0,1) \tag{5.3}$$

where $\alpha_{\text{gauss}} > 0$ is the Gaussian noise coefficient, $\bar{x}$ is the augmented data, x is the raw signal and G the gaussian noise.

ii) Masking Noise: an alternative way to apply random noise, where a fraction α_{mask} of the elements in x are randomly selected and set to 0.

iii) Signal Translation: the signal is randomly shifted forward and backward, with the data gap resulting from the translation complete with zeros, maintaining the original size of the signal.

iv) Amplitude Shifting: Moderate variations in signal amplitude are common, and do not significantly alter the machine's operating state. Therefore, performing modifications of this nature tends to enrich the dataset. Amplitude shifting is implemented through a scaling factor (α_{scal}), Equation 5.4.

$$\bar{x} = \alpha_{\text{scal}} * x \tag{5.4}$$

v) Time Stretching: The signal is stretched along the time axis by a stretching factor ($\alpha_{\text{stre}} > 0$). A sample with N_{aug} values in the range of [1, α_{stre}] if $\alpha_{\text{stre}} \geq 1$ or [α_{stre},1] is obtained. The sample is centered and stretching is applied. Points outside the dimensions are clipped out and gaps filled with zeros.

At the end of the data augmentation process, each signal will generate five new signals, and the process can be repeated n times. In summary, the entire process of creating synthetic signals and augmentation from one original signal will result in 35 new signals (7 synthetic signal * 5 augmented signal). The number of times (q_{aug}) the procedure needs to be executed to obtain a total of n signals (n_{total}) in the training group can be calculated in Equation 5.5.

$$q_{\text{aug}} = \frac{n_{\text{total}}}{7 * 5 * n_{\text{r}}} \tag{5.5}$$

where, n_{total} is the desired amount of samples in the training set, 7 is the amount of synthetic signals, 5 is the amount of augmented signals and n_{r} the amount of real signals used. When the relationship does not result in an integer, the value is rounded up to the nearest integer, and the excess is randomly selected and discarded.

5.3.4 Signal Processing

DL models are able to extract features from raw signals and learn characteristic behaviors without the need for feature engineering as in classic ML models.

Although the raw signal (time domain) presents relevant information, due to the amount of frequencies present, the analysis often becomes complex. Therefore, it is common to perform analysis in the frequency domain. As one of the goals of the work is to provide interpretability of the result, the input signal was used in the frequency domain instead of the time domain.

The signal was normalized using Z-score normalization, and finally, cut to reduce the number of points, and consequently the computational cost of the model. For the cut, the maximum frequencies at which the fault could manifest and their harmonics were analyzed,

so according to the monitored asset, the size of the signal used may vary. As the dataset under analysis present constant speed, it is not necessary to perform order tracking or other method as a pre-processing step to remove the effect of shaft speed variations.

5.3.5 Fault Diagnosis

CNN are present in several applications due to their main characteristics: i) possibility to learn features directly from raw signal in training; ii) are immune to small transformations/disturbances in the input data; iii) can adapt to different sizes; iv) excellent computational cost compared to conventional fully-connected Multi-Layer Perceptrons (MLP) networks (CNN neurons are sparsely-connected with tied weights Kiranyaz et al. (2021)).

Since AlexNet's proposal Krizhevsky et al. (2012), when it achieved 16.4% error rate in the ImageNet benchmark dataset (this was about 10% lower than the second top method), several deep 2D CNNs have been used. However, there are certain drawbacks and limitations of using such deep CNNs: i) high computational cost and special hardware for training; ii) requires massive size dataset for training to achieve reasonable generalization capability Kiranyaz et al. (2021). Factors that are not always available when it comes to industrial applications involving vibration signals. In addition, vibration signals are in 1D, which would make it necessary to modify their dimension, further increasing the computational cost and need for pre-processing. On the other hand, 1D CNN makes it possible to use the vibration signal in its original dimension. It has low computational cost compared to 2D CNN and is well-suited for real-time and low-cost applications especially on mobile or hand-held devices Kiranyaz et al. (2021). For these reasons, it was used in the study.

In this part, the synthetically generated and augmented signals are used to train the model (1D CNN) and adjust the hyperparameters. It is worth mentioning that no real fault signals are used in the training. The test group consists of other real machine signals under different operating conditions. After training, the model is ready to perform the classification.

5.3.6 Explainable Artificial Intelligence (XAI)

Explaining the results obtained by the AI model is essential to ensure its use in the industry, since it is associated with increased trust. Vibration signals are preferably analyzed in the frequency domain in order to facilitate the correlation between the highlighted frequency/behavior and the fault pattern. In this way, applying a method that evaluates the points of relevance in the signal used for the classification, allows the end user to understand and validate the result obtained. For this purpose, the Gradient-weighted Class Activation Mapping (Grad-CAM) approach was used.

After classifying the sample by CNN 1D, the method is applied, returning a heatmap with the most relevant frequencies for the model. The heatmap obtained helps the specialist in decision making, allowing the validation of the classification determined by the model.

5.4 Experimental procedure

5.4.1 Data description

Three datasets were used to address different faults found in rotating machinery, two of which were publicly available Qiu et al. (2006); Cao et al. (2018) and one developed by the author for the study. As shown in Lei (2017) even though the rotating machinery is diversified, some common essential rotating parts are: rotors, rolling element bearings, and gears Brito et al. (2022). Therefore, the datasets were chosen and developed to address the main faults present in rotating machinery components, namely: defects in bearing and gearbox, misalignment, unbalance and mechanical looseness. The use of different datasets aims to validate the proposed methodology in different scenarios. As they are publicly available datasets and already explored in Brito et al. (2022), Case 1: Bearing Dataset and Case 2: Gearbox Dataset will be briefly discussed. For more details, please refer to Qiu et al. (2006); Cao et al. (2018); Brito et al. (2022). All datasets are vibration signals collected using accelerometers.

Case 1: Bearing dataset

The dataset Qiu et al. (2006) is composed of remaining useful life (RUL) bearing tests, each file consisting of 20,480 points with the sampling rate set at 20 kHz. As this is a run-to-failure test, no labels are available (the only information provided is the type of fault present at the end of each test). Therefore, the data were analyzed and labeled manually. For the study the bearing 01 of test 02 was used. The test presents 984 observations, with the first 531 observations labeled as normal and the last 453 as fault in the outer race (BPFO).

Case 2: Gearbox dataset

To evaluate faults in gears, the dataset Cao et al. (2018) was used. Nine different gear conditions were introduced to the pinion on the input shaft, including healthy condition, missing tooth, root crack, spalling, and chipping tip with five different levels of severity. For each gear condition, 104 observations were collected resulting in a total of 936 observations with sampling frequency of 20 kHz. For the test, the samples were considered as: 104 normal and 832 gear fault.

Case 3: Mechanical faults dataset

The last dataset was developed by the authors, aiming to address classic faults in rotating machines such as: unbalance, misalignment and mechanical looseness, in addition to the normal operating condition.

The faults were introduced on a test bench, Fig. 5.1, being: motor, frequency inverter, bearing house, two bearings, two pulleys, belt and rotor (disc).

Figure 5.2: Bench test.

- Motor: Three-phase induction motor, B56 B4, Manufacturer Eberle, nominal speed 1650 rpm, power 0.09 kW, voltage 220 V, current 0.70 A, bearings 6200 ZZ.

- Frequency Inverter: Vector Inverter, model CFW300, manufacturer WEG.

- Driven Pulley (Disc Side): model A80, external diameter 80 mm, internal diameter 53 mm.

- Motor Pulley (Motor Side): model A60, external diameter 60 mm, internal diameter 37 mm.

- Belt: model V - A23, top width 12.7 mm, outer perimeter 630 mm, inner perimeter 580 mm.

- Rotor (Disc): diameter 149 mm, 36 holes, thickness 6.5 mm.

- Bearings 1205, bearing house SN505, bushing H305.

For the acquisition of the signals were used: 04 accelerometers PCB 352C33 (2 in each bearing), mounted in the vertical and horizontal positions (y- and x-axes), 01 Power Supply PCB 482A20, 01 acquisition board Hi-Speed USB Carrier, NI USB-9162. A Python language program was developed to collect the data. All accelerometer sensitivities were adjusted according to the calibration chart. The rotation was kept constant with a value measured on the shaft of approximately 1238 rpm.

20 tests were performed, 5 for each condition (normal, unbalance, misaligment and looseness). Each test consists of 4 sets of 420 signals collected continuously, each file consisting of 25,000 points with the sampling rate set at 25 kHz (420 signals per

accelerometer). Resulting in the end of all tests, in a total of 8400 signals per accelerometer.

The sequence of tests was randomly defined. Before starting any test, the bench was dismantled and returned to normal operating condition, to later introduce the fault. The experimental procedure allows variations to occur, making the tests closer to industrial reality. All data used in this study will be made available.

To increase the randomness of the tests in the fault classification (AI), and to validate the robustness of the model, only 3 tests per condition were used in the analyses, resulting in 12 tests in total (5040 signals). Since the fault classification was performed 10 times, at each new run of the model, 3 tests were randomly selected for each class of fault.

As it is a small bench test, the measurement points are close, and in order to reduce the computational cost, only the signals from the horizontal position of the accelerometer present in the bearing (near the pulley) were used in the analyses.

5.4.2 Hyperparameter tuning and evaluation metrics

The hyperparameters were adjusted based on the training set to obtain the best performance and stability in the model. The cross-validation method was used, where the signals were randomly chosen between the training and validation group, being 90% for training and 10% for validation. It is worth mentioning that the training group is composed only of synthetic signals, and therefore does not contain any real fault signals. The test set is composed of all signals excluding the normal signals that were used for generating the synthetic signals.

The model was composed of a 1D CNN layer, followed by a dropout layer for regularization, then a pooling layer. The objective of using only 1 layer is to reduce the complexity and computational cost of the model. As CNN has high learning capacity, the dropout layer was used to slow down the learning process and hopefully avoid overfitting. Pooling learning reduces the amount of learned features by consolidating them and keeping only the essential ones, which prevents the network from learning features uncorrelated with the condition.

The learned features are flattended into a single long vector and passed to the fully connected layer before the output layer used to make a prediction. The fully connected layer

provides a buffer between learned features and the output with the intention of interpreting learned features before making a prediction.

For the model, 32 parallel feature maps and kernel size of 5 were used. ReLu activation function and Adam optimizer were also used. To automate the training, the model was trained with a variable number of epochs, based on the Early Stopping technique, with a patience of 8 and a variation of 0.001 in relation to the validation accuracy. Batch size of 32 samples was used. Categorical cross entropy loss function and Softmax (activation function) were used as it is a multi-class classification problem.

The model was trained to classify 7 conditions, regardless of whether the case under analysis presents the fault or not, being: normal condition, BPFO, BPFI, unbalance, misalignment, looseness and gear fault. This simulates a condition where different faults may appear on rotating machinery, and the model will be prepared to identify them. In addition, it increases the complexity of the analyses.

Neural networks are stochastic, so at each new training a different model will be obtained, even if using the same dataset. It is worth mentioning that, in addition to the stochasticity of the networks, there is also the randomness in the generation and augmentation of the signals. To ensure that the model is stable and robust to variations in both network training and signal generation, the analyzes were repeated 10 times. At each analysis, new synthetic signals were randomly generated and the model trained. As in all dataset, the number of samples per condition were approximately equal, accuracy was used. For Case 2 where the test data are imbalanced, despite the use of accuracy, the confusion matrix is shown. To measure the variability after the 10 analyses, the standard deviation was calculated. The tests were performed using 2.2 GHz Intel Core i7 Dual-Core, 8 GB 1600 MHz DDR3, Intel HD Graphics 6000 1536 MB.

5.4.3 Analysis approaches

The success of an AI model is directly related to the quality and quantity of training data. On the other hand, a large amount of data implies acquisition costs, as well as a low quality implies a reduction in the model's accuracy. Being able to reduce the amount of real data needed, and generate data varied enough to make the model robust, is a solution to

enable the implementation of AI in the industrial application.

Besides the main objective of verifying the possibility of using a model trained with synthetic and augmented data to evaluate real signals and their explainability, other analyzes were performed.

First, the augmented data are analyzed, verifying if the behavior of the synthetic signal in the frequency domain, matches the available knowledge about fault detection in rotating machinery using vibration analysis. Subsequently, in order to reduce the amount of real data needed to generate the synthetic data, the minimum viable amount of data was analyzed to obtain a stable and robust model. This approach can be seen as an analogy of the study area called Few-Shot Learning, where few samples are used to train the model. Another analysis performed is in relation to the total amount of real samples needed in the training group to ensure the stability and robustness of the model. In addition, the possibility of working with 100% synthetic samples was analyzed, that is, where synthetic data are generated without dependence on the real sample. For comparison between training forms, supervised training was performed, using only real signals, which were divided into 70% for training and 30% for testing. Finally, the data obtained through XAI are analyzed to verify if the features learned by the model are capable of providing relevant information to the expert for decision making.

5.5 Results and discussion

5.5.1 Data exploration

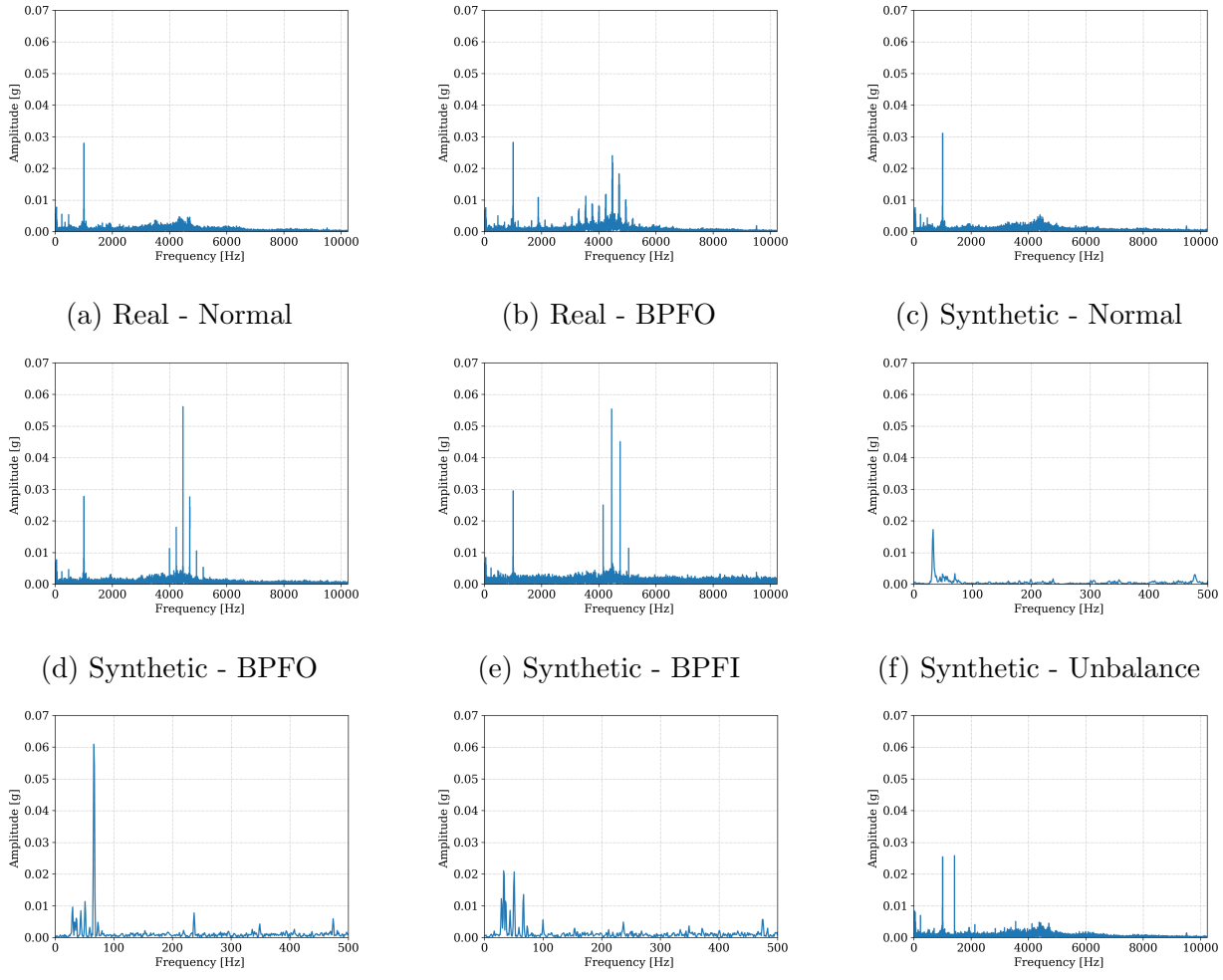
In this section the synthetic samples for Case 1-3 are analyzed, discussed and compared with the real samples. One sample per condition from the training database was randomly selected. Due to the randomness of the proposed approach, variations in signals may occur, being important to observe whether the deterministic characteristics of the conditions remain similar.

As the objective is only to visualize the signal characteristic, df = 1 hz was used, with the exception of Case 2, where df = 5.5 hz due to data limitation. For BPFO, BPFI and Gear Fault conditions, the full signal is presented to highlight the fault characteristic

at high frequency. In Unbalance, Misalignment, Looseness, a zoom was performed in the range from 0 to 500 hz to facilitate the visualization of the characteristic frequencies. For the normal condition, the complete signal is displayed.

Case 1 - Bearing Fault

In Fig. 5.3 the original signals for the two conditions (Normal and BPFO) and the synthetic signals for the seven created conditions (Normal, BPFO, BPFI, Unbalance, Misalignment, Looseness and Gear Fault) from Case 1 are presented.



(a) Real - Normal (b) Real - BPFO (c) Synthetic - Normal

(d) Synthetic - BPFO (e) Synthetic - BPFI (f) Synthetic - Unbalance

(g) Synthetic - Misalignment (h) Synthetic - Looseness (i) Synthetic - Gear Fault

Figure 5.3: Examples of vibration signals for Case 1, real and synthetic.

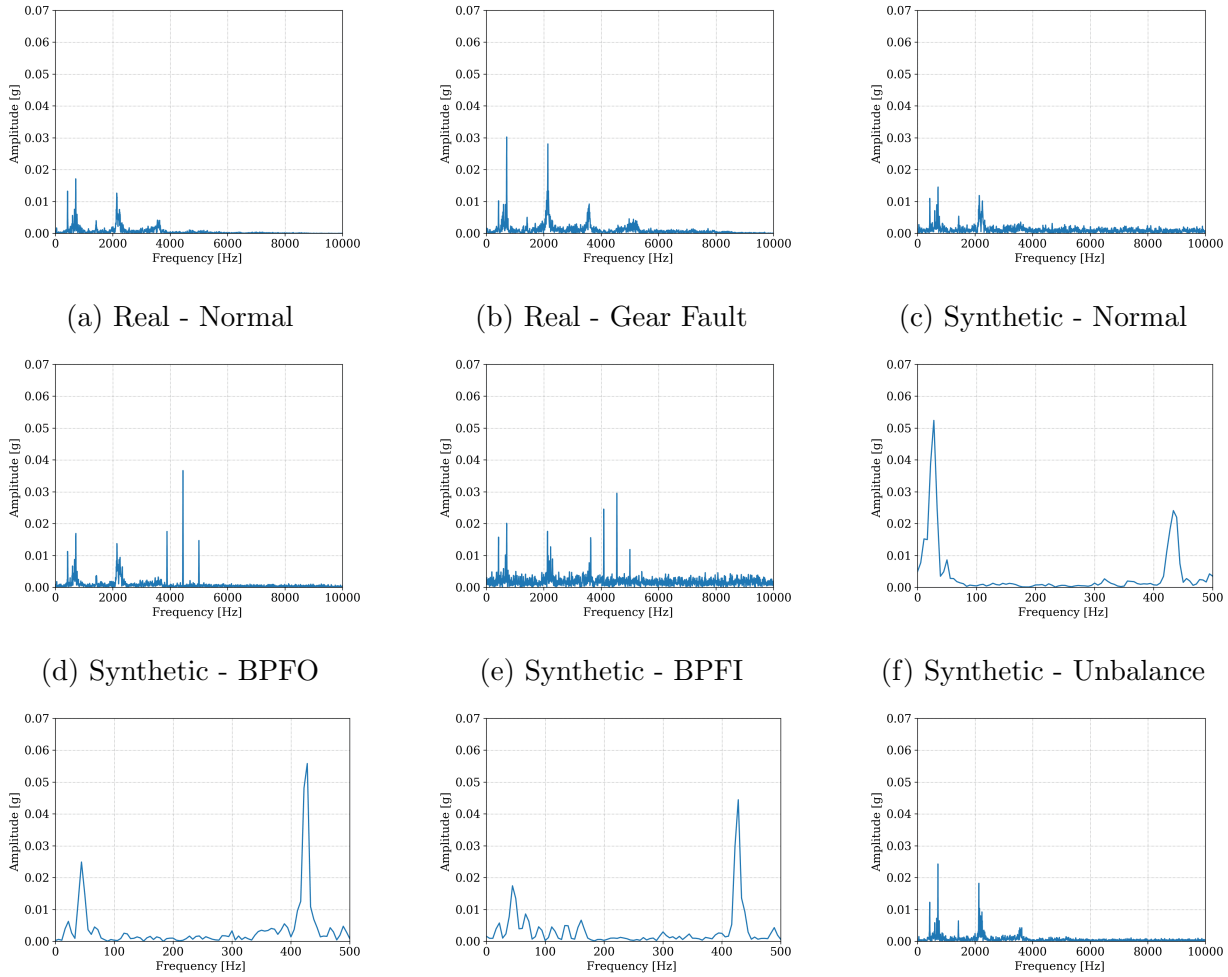
The real normal condition signal, Fig.5.3a does not present any frequency related to faults, showing only the characteristic behavior of the machine. On the other hand, in Fig. 5.3b there is the appearance of frequencies in the region of 3000 to 6000 Hz, related to excitation of the natural frequencies, caused by the fault in the outer race.

Analyzing the synthetic signals, it can be noted that for bearing faults, the impacts due to a small defect tend to excite the races natural frequencies (at high frequency) Fig.5.3d e 5.3e, being the fundamental frequency corresponding to the damaged element, BPFO = 236.4 hz and BPFI = 296.9 hz. For the unbalance, Fig.5.3f, there is a predominant increase in amplitude by 1 x fr (rotation frequency). In misalignment, there is an increase in 2 and 3 x fr, with emphasis on 2 x fr, Fig.5.3g. Finally, for the mechanical looseness, there is an increase in harmonics and the appearance of some sub-harmonics of fr, Fig.5.3h. Gear defect, Fig.5.3i show an appearance of frequencies related to GMF, hypothetically placed at 711 hz and its harmonics, since this dataset does not have gears.

Comparing the synthetic and original (real) signals for the conditions present in the dataset (normal and BPFO), the similarity of the fundamental characteristics is noted, ensuring that the model learns the behavior of the signal in relation to each operating condition of the equipment.

Case 2 - Gear Fault

In Fig. 5.4 the original signals for the two conditions (Normal and Gear Fault) and the synthetic signals for the seven created conditions (Normal, BPFO, BPFI, Unbalance, Misalignment, Looseness and Gear Fault) from Case 2 are presented.



(a) Real - Normal (b) Real - Gear Fault (c) Synthetic - Normal

(d) Synthetic - BPFO (e) Synthetic - BPFI (f) Synthetic - Unbalance

(g) Synthetic - Misalignment (h) Synthetic - Looseness (i) Synthetic - Gear Fault

Figure 5.4: Examples of vibration signals for Case 2, real and synthetic.

The real normal condition signal, Fig.5.4a shows evidence of 1 and 2 x GMF, which is normal when monitoring gearboxes. On the other hand, the original faulty signal, Fig.5.4b already presents, in addition to 1 and 2 x GMF, harmonics in 3 and 4 x GMF, and side bands, characterizing the appearance and progression of the fault.

Regarding the synthetic signals, the same analysis performed for Case 1 is valid for Fig.5.4, with the exception of Fig.5.4i which has GMF equal to 711 hz, being the real frequency of the gearbox. It can be noted that the behavior of the real and synthetic signals for each condition is similar, as proposed by the approach.

Case 3 - Mechanical Faults

In Fig. 5.5 the original signals for the four conditions (Normal, Unbalance, Misalignment and Looseness) and the synthetic signals for the seven created conditions

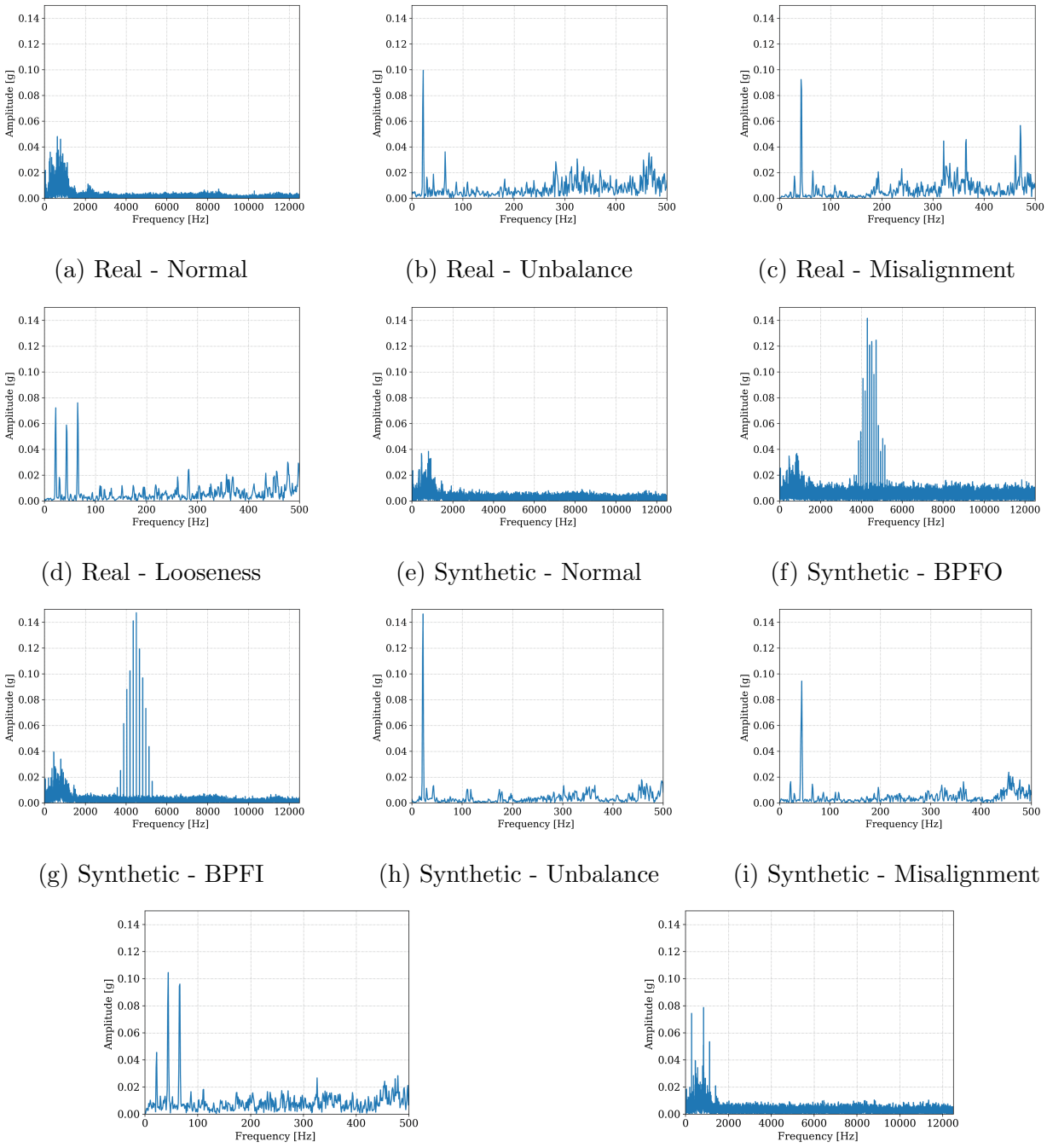
(Normal, BPFO, BPFI, Unbalance, Misalignment, Looseness and Gear Fault) from Case 3 are presented.

Note, in Fig. 5.5, that even for the normal condition, there is evidence of highlighted frequencies up to 2000 hz, which may be related to lack of precision in the adjustment, assembly and lubrication of the bearings, improper adjustment (tensioning) and/or wear of the pulleys, noise generated by the frequency inverter etc. The condition was purposely chosen to simulate an industrial situation. It is known that, not all companies have enough specialized technicians to carry out an assembly and adjustment within the World Class Maintenance (WCM) standards. Also, there is a lack of specialized equipment to perform laser alignment, dynamic balancing etc. Moreover, due to the need for quick intervention to release the equipment for operation, maintenance cannot be performed with the desired precision. Because of these problems, the vibration signal can present a series of frequencies and/or noises, which make it difficult for the predictive maintenance specialist to analyze the main condition.

Even in a scenario where the signal is heavily polluted, technicians are usually able to perform an accurate diagnosis, based on field experience. Thus, it was decided to keep the bench in the presented condition, to make the test closer to the industrial reality, instead of completely adjusting the bench to eliminate noise and other sources of excitation that are not related to the defects. Allowing, to assess whether the proposed methodology is capable of imitating human behavior, accurately identifying the operating conditions of the asset.

Again, analysis performed for Case 1 and 2 in relation to synthetic data is valid for Fig. 5.5, with the exception of Fig. 5.5f and 5.5g being the fundamental frequency corresponding to the damaged element, BPFO = 107.09 hz and BPFI = 155.7 hz.

As in the other cases, it can be noted that the described fault characteristics are evident in all the synthetically generated signals and in the original signals. This shows that the data augmentation process was able to increase the diversity of the dataset without losing its fundamental characteristics.



(a) Real - Normal (b) Real - Unbalance (c) Real - Misalignment

(d) Real - Looseness (e) Synthetic - Normal (f) Synthetic - BPFO

(g) Synthetic - BPFI (h) Synthetic - Unbalance (i) Synthetic - Misalignment

(j) Synthetic - Looseness (k) Synthetic - Gear Fault

Figure 5.5: Examples of vibration signals for Case 3, real and synthetic.

5.5.2 Number of total samples for training

To evaluate the influence of the total number of signals in the training group, 10 sets were created containing: 1050, 2100, 3150, 4200, 5250, 6300, 7350, 8400, 9450, 10500 signals.

As the objective was not to evaluate the efficiency of the model, but to compare the result obtained in relation to the amount of signals in the training group, 30 original signals were selected to perform the data augmentation step, and the results for accuracy and standard deviation are shown in Fig. 5.6.

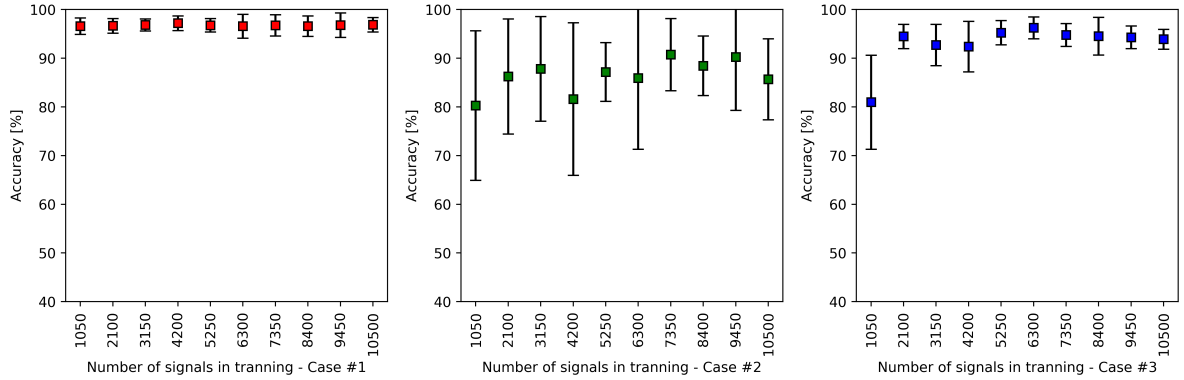


Figure 5.6: Results for each case varying the total amount of signals in the training set.

Analyzing Fig.5.6, it can be noted that there is no significant variation in relation to the amount of data used, with the exception of Case 3 in the set of 1050 signals. Such variation may have occurred due to the greater complexity of the signal present in the dataset, requiring a greater amount of signals in training for learning.

Note that there was no significant gain in the accuracy and stability of the model (less standard deviation) in increasing the number of signals above 5250 signals, and therefore, the amount was selected for the other analyses.

5.5.3 Number of real samples for training

Based on the previous analysis, 5250 signals were used in the training, aiming to balance the computational cost and the efficiency of the model.

Unlike the previous analysis, the number of real signals to generate the total amount of training was varied. The goal is to analyze whether, through zero or few data acquisitions, it is possible to train a model robust enough for fault diagnosis.

11 conditions were tested, using 0, 1, 2, 3, 5, 10, 15, 25, 30, 50 and 75 real signals to generate the 5250 signals in the training. The values were selected so that the number of times the signals were augmented resulted in integer values, with no need to discard excess signals. As Case 2 has only 104 real signals per condition, the range was set from 1 to 75 signals. The results obtained for accuracy and standard deviation are shown in Fig. 5.7.

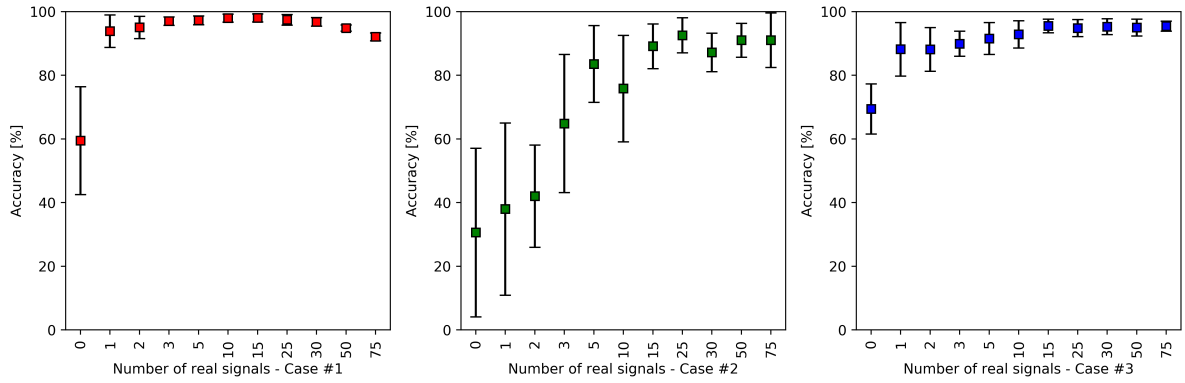


Figure 5.7: Results for each case varying the amount of real signals used to create the training set.

Unlike the other conditions, in the first test, all signals for the training group were generated synthetically, without performing the combination with the real signal (zero number of real signal). It can be seen in Fig.5.7, that with zero real signal the lowest performing test was achieved. This result was expected, since the signals were created, basically containing the machine rotation frequency, white noise and random frequencies with amplitude close to the noise. This characteristic does not represent the dynamic behavior of the machine, resulting in the lowest value obtained. Therefore, it can be concluded that using the chosen strategy, it is not possible to train a model using only synthetic data without the combination with the real signal. It is worth mentioning that, despite not being the focus of the work, other strategies, such as more complex modeling, can be tested to overcome this problem.

Analyzing Fig.5.7, it can also be seen that using up to two real signals to create the entire training dataset, the average accuracy is significantly lower than the others, and the model has a large standard deviation. This occurs because if one of the selected signals presents some noise or excitation not related to the dynamic behavior of the machine, such pattern will be replicated for at least 50% of the training set signals, leading the model to

error.

The increase in the number of signals tends to contribute to the improvement in the performance of the model, although it does not mean a linearly increasing behavior. It is noted that, in general, there is a stabilization of accuracy and standard deviation after 15 signals. In view of the computational and acquisition cost, and the results obtained with the increase in the amount of real data, it is suggested to be between 15 and 30 signals, a good working range for the proposed approach.

The maximum accuracy value reached in Case 1 was 98.1% and standard deviation of 1.2% with 15 signals. For Case 2, the maximum accuracy value was 92.5% and standard deviation of 5.5% with 25 signals. Case 3 obtained the highest accuracy value with 15 signals, being 95.5% and standard deviation of 2.1%. Smaller values of standard deviation were obtained for other amounts of signals, however none of them had a difference greater than 0.5% in relation to those presented.

The differences in accuracy and standard deviation obtained between each case are in accordance with what was expected in the study. Case 1 is a bearings remaining useful life test, where data were collected continuously, without the bench being dismantled to insert the defects. As there were no changes in the test bench, variations in the dynamic behavior of the machine are not expected, other than those caused by the wear of the components (since it is isolated). This makes the fault identification simpler for the model.

On the other hand, in Case 2 and 3, the benches were dismantled to insert the faults. Therefore, small changes in relation to the real signal used as a reference for generating the synthetic data may occur during assembly and adjustment, generating a reduction in accuracy. This condition represents a situation where the system was trained with a machine signal in a similar condition, but not identical to the current one (which is a transfer learning problem definition).

The two analyzes are extremely interesting, and show the applicability of the system in the industrial scenario. In Case 1, it addresses the monitoring of a machine, where the system was installed, trained, and over time was able to identify a fault in the asset. In Case 2 and 3, the model was initially installed and trained. Hypothetically, the machine underwent some corrective maintenance, and the same model, without being retrained,

continued to monitor the asset, and was able to identify the faults present. The fact that there is no need to retrain the model for small changes in the asset or reestablish the operating condition through corrective maintenance, overcomes a major problem in applications that is retraining. It is worth mentioning that for major interventions or changes in projects that modify the dynamic behavior of the machine, the model will tend to reduce its accuracy if not retraining.

The confusion matrix for the best values obtained in Case 1 is presented, containing the number of signals per class and the standard deviation for the 10 tests, in addition to the value in percentage, Table 5.1. Where in the column are the True Labels and in the row are the Predicted Labels.

Table 5.1 – Confusion Matrix - Case 1

Case 1	Normal	BPFO	BPFI	Unbalance	Misalignment	Looseness	Gear Fault
Normal	**52.01 % (0.51%)**	0.72 % (0.41%)	0 % (0%)	0 % (0%)	0.41 % (0.51%)	0.10 % (0.10%)	0 % (0%)
	504 (5)	7 (4)	0 (0)	0 (0)	4 (5)	1 (1)	0 (0)
BPFO	0.62 % (0.92%)	**46.03 % (0.92%)**	0 % (0%)	0 % (0%)	0 % (0%)	0.10 % (0.10%)	0 % (0%)
	6 (9)	**446 (9)**	0 (0)	0 (0)	0 (0)	1 (1)	0 (0)
BPFI	0 % (0%)	0 % (0%)	**0 % (0%)**	0 % (0%)	0 % (0%)	0 % (0%)	0 % (0%)
	0 (0)	0 (0)	**0 (0)**	0 (0)	0 (0)	0 (0)	0 (0)
Unbalance	0 % (0%)	0 % (0%)	0 % (0%)	**0 % (0%)**	0 % (0%)	0 % (0%)	0 % (0%)
	0 (0)	0 (0)	0 (0)	**0 (0)**	0 (0)	0 (0)	0 (0)
Misalignment	0 % (0%)	0 % (0%)	0 % (0%)	0 % (0%)	**0 % (0%)**	0 % (0%)	0 % (0%)
	0 (0)	0 (0)	0 (0)	0 (0)	**0 (0)**	0 (0)	0 (0)
Looseness	0 % (0%)	0 % (0%)	0 % (0%)	0 % (0%)	0 % (0%)	**0 % (0%)**	0 % (0%)
	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	**0 (0)**	0 (0)
Gear Fault	0 % (0%)	0 % (0%)	0 % (0%)	0 % (0%)	0 % (0%)	0 % (0%)	**0 % (0%)**
	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	**0 (0)**

In Table 5.1, it can be seen that the model confused normal signals with BPFO, misalignment and looseness. On the other hand, the faulted signals (BPFO) were confused with the normal condition and looseness only. The confusion between normal and BPFO may still be acceptable, due to the run to failure test feature and manual label. Therefore, signals referring to the beginning of the fault may present similar characteristics. The other classifications were model errors. Analyzing the result only by class, the normal signals were classified correctly 97.67% of the time, and faulty signals (BPFO) 98.45%. The confusion matrix for Case 2 is shown in Table 5.2.

Table 5.2 – Confusion Matrix - Case 2

Case 2	Normal	BPFO	BPFI	Unbalance	Misalignment	Looseness	Gear Fault
Normal	**7.90 % (0.98%)**	0 % (0%)	0 % (0%)	0.11 % (0.11%)	0.11 % (0.11%)	0.11 % (0.11%)	0.44 % (0.76%)
	72 (9)	0 (0)	0 (0)	1 (1)	1 (1)	1 (1)	4 (7)
BPFO	0 % (0%)	**0 % (0%)**	0 % (0%)	0 % (0%)	0 % (0%)	0 % (0%)	0 % (0%)
	0 (0)	**0 (0)**	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
BPFI	0 % (0%)	0 % (0%)	**0 % (0%)**	0 % (0%)	0 % (0%)	0 % (0%)	0 % (0%)
	0 (0)	0 (0)	**0 (0)**	0 (0)	0 (0)	0 (0)	0 (0)
Unbalance	0 % (0%)	0 % (0%)	0 % (0%)	**0 % (0%)**	0 % (0%)	0 % (0%)	0 % (0%)
	0 (0)	0 (0)	0 (0)	**0 (0)**	0 (0)	0 (0)	0 (0)
Misalignment	0 % (0%)	0 % (0%)	0 % (0%)	0 % (0%)	**0 % (0%)**	0 % (0%)	0 % (0%)
	0 (0)	0 (0)	0 (0)	0 (0)	**0 (0)**	0 (0)	0 (0)
Looseness	0 % (0%)	0 % (0%)	0 % (0%)	0 % (0%)	0 % (0%)	**0 % (0%)**	0 % (0%)
	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	**0 (0)**	0 (0)
Gear Fault	5.49 % (3.29%)	0 % (0%)	0 % (0%)	0.88 % (2.19%)	0.22 % (0.32%)	0.11 % (0.11%)	**84.63 % (7.68%)**
	50 (30)	0 (0)	0 (0)	8 (20)	2 (3)	1 (1)	**771 (70)**

In Case 2, Table 5.2, normal condition and gear fault were confused with unbalance, misalignment, looseness and gear defect. As the gear fault characteristics are different from the errors presented, it can be concluded that there is a model error, despite the low amount. Analyzing by class, normal signals were classified correctly 91.14% of the time, and faulty signals 92.67%. The confusion matrix for Case 3 is shown in Table 5.3.

Table 5.3 – Confusion Matrix - Case 3

Case 3	Normal	BPFO	BPFI	Unbalance	Misalignment	Looseness	Gear Fault
Normal	**24.54 % (0.17%)**	0.04 % (0.09%)	0.02 % (0.02%)	0.06 % (0%)	0.04 % (0%)	0.02 % (0.02%)	0.06 % (0.09%)
	1223 (9)	2 (5)	1 (1)	3 (0)	2 (0)	1 (1)	3 (5)
BPFO	0 % (0%)	**0 % (0%)**	0 % (0%)	0 % (0%)	0 % (0%)	0 % (0%)	0 % (0%)
	0 (0)	**0 (0)**	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
BPFI	0 % (0%)	0 % (0%)	**0 % (0%)**	0 % (0%)	0 % (0%)	0 % (0%)	0 % (0%)
	0 (0)	0 (0)	**0 (0)**	0 (0)	0 (0)	0 (0)	0 (0)
Unbalance	0.08 % (0.04%)	0.02 % (0.02%)	0 % (0%)	**24.82 % (0.39%)**	0 % (0%)	0.16 % (0.39%)	0 % (0%)
	4 (2)	1 (1)	0 (0)	**1247 (20)**	0 (0)	8 (20)	0 (0)
Misalignment	0.62 % (0.65%)	0 % (0%)	0 % (0%)	0.02 % (0%)	**22.71 % (2.46%)**	1.71 % (2.40%)	0.02 % (0%)
	31 (33)	0 (0)	0 (0)	1 (0)	**1141 (124)**	86 (121)	1 (0)
Looseness	0 % (0%)	0 % (0%)	0 % (0%)	1.53 % (0.85%)	0.10 % (0.13%)	**23.44 % (0.75%)**	0 % (0%)
	0 (0)	0 (0)	0 (0)	77 (43)	5 (7)	**1178 (38)**	0 (0)
Gear Fault	0 % (0%)	0 % (0%)	0 % (0%)	0 % (0%)	0 % (0%)	0 % (0%)	**0 % (0%)**
	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	**0 (0)**

Table 5.3 presents the results obtained for Case 3. The normal signals were confused with all types of faults, despite the low amount. The unbalance signals were confused with normal condition, BPFO and looseness. Misalignment was confused with normal condition, unbalance, looseness and gear fault. Finally, looseness was confused only with unbalance and misalignment. Among the classes, the normal condition was the one

with the highest accuracy rate with 99.04%. The misalignment showed the highest error, resulting in 90.56%. The unbalance showed 98.97% and looseness 93.49%. The results show the greatest difficulty in classifying misalignment and looseness, which was expected, since some harmonics (such as: 2x and 3x) can be highlighted in both condition, making the analysis difficult.

5.5.4 Supervised training with real signals only

A comparative analysis was performed to verify the difference in accuracy and robustness of the model using the proposed approach and with supervised training using only real data.

In supervised training, all samples have labels and the signals used for each class refer to real fault conditions. Although in general, it presents excellent results, it is necessary to have samples of all fault conditions and they need to be previously labeled. Assumptions that often make industrial applications unfeasible.

The real signals were divided into 70% training and 30% testing, and as in the proposed approach, they were executed 10 times due to the randomness of the model. All available signals were used. The accuracy and standard deviation for each test were: 98.1% (0.53%), 98.9% (0.86%) and 99.8% (0.11%), for Case 1, 2 and 3 respectively.

The result obtained for Case 1 it was the same using the two types of training (supervised with real signals and proposed approach). The only difference was lower the standard deviation (0.53% compared to 1.2% in the proposed methodology). It is believed that the results were similar due to the lower complexity of the dataset, and consequently ease of identification of the fault. Moreover, the faults have more specific pattern compared to the other cases.

In Cases 2 and 3, supervised training with real data obtained higher results and lower standard deviation than the proposed approach, as expected. In Case 2 the precision was improved from 92.5% to 98.9% and the standard deviation decreased from 5.5% to 0.86%. In Case 3, the accuracy was 95.5% with the proposed approach becoming 99.8% using only the real signals, and the standard deviation went from 2.1% to 0.11%.

The results confirm the expected superiority of supervised training with only real

data. On the other hand, it is worth mentioning that obtaining real and labeled data of all faults can be a complex and time-consuming task in the industrial applications, problems that are overcome by the proposed approach.

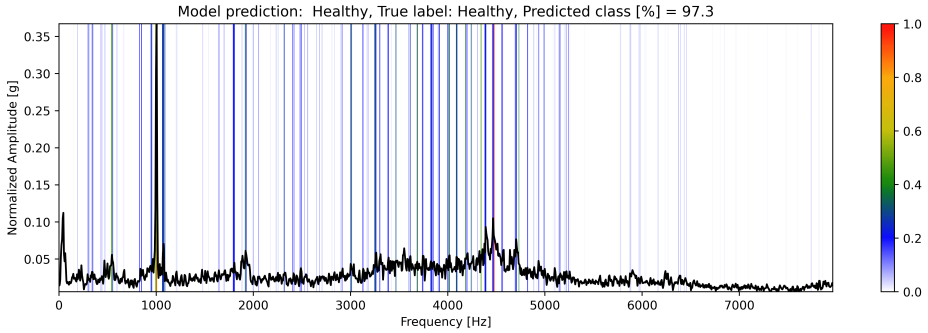
5.5.5 Explainable Artificial Intelligence (XAI)

Understanding how the AI model arrived at the end result is critical to its application in the industry. The models need a minimum of interpretability, so that users can trust in the predictions.

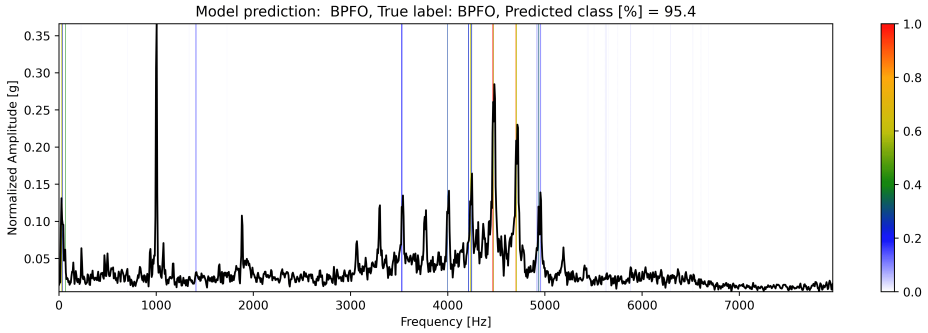
In addition to allowing us to understand the most relevant parameters used by the model for the final classification, the analysis also allows us to evaluate the learning of the model.

The model's interpretability makes it possible to understand which were the most relevant features used for the prediction, and also to evaluate the model's learning.

The gradients used by the Grad-CAM method after the classification of the sample by the 1D CNN are evaluated. The heat map with the most relevant frequencies for the prediction are presented. A random sample was selected for each type of classification in each case, to exemplify the most relevant features, and validate if they are in accordance with the available literature on faults in vibration analysis of rotating machines. The results obtained for Case 1 are shown in Fig.5.8.



(a) Normal

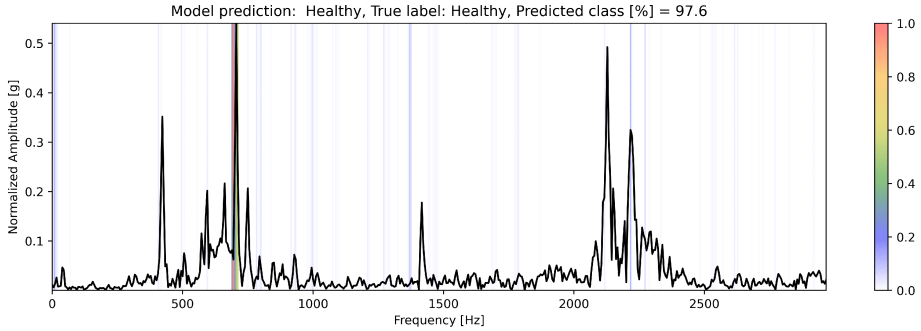


(b) BPFO

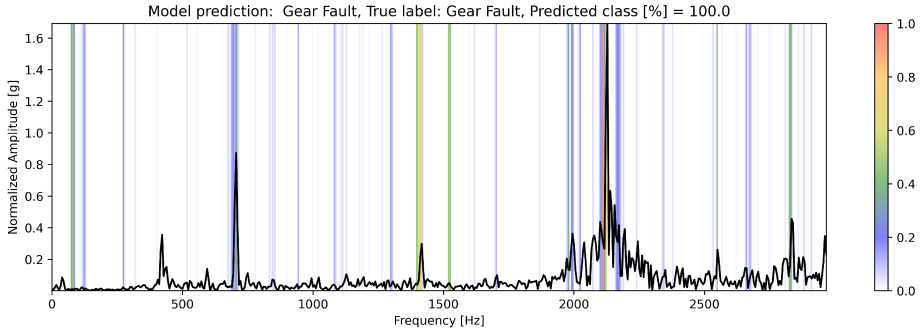
Figure 5.8: XAI analysis - Case 1

Analyzing Fig.5.8a, it is verified that the normal signal uses several frequencies of the signal, not having a specific frequency or region. This was to be expected, as no faults are present, the entire signal referring to a normal machine condition. In the faulty condition, Fig.5.8b, the high frequency regions are excited, since it is a bearing fault. Consequently, the model uses the frequencies associated with the BPFO as the most relevant for indicating the defect. The analysis corresponds to a manual analysis performed by a rotating machinery specialist.

The results obtained for Case 2 are shown in Fig.5.9.



(a) Normal

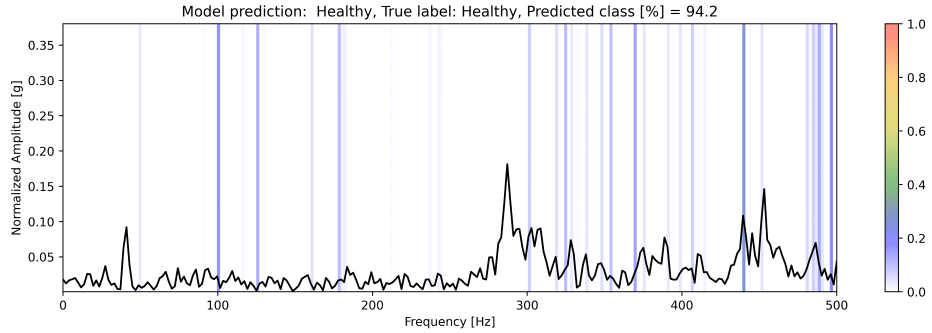


(b) Gear Fault

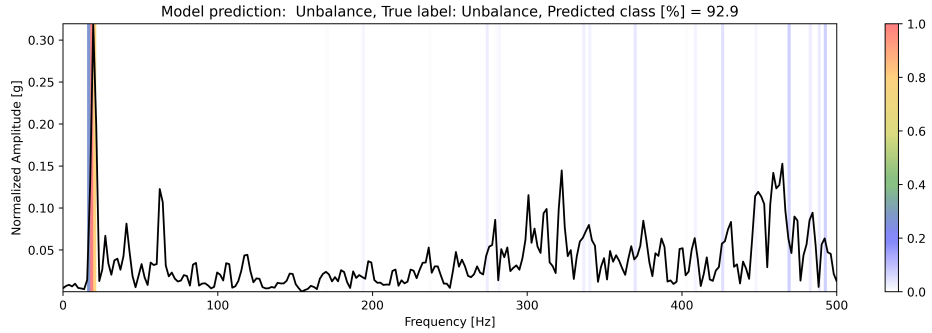
Figure 5.9: XAI analysis - Case 2

Case 2, Fig.5.9a, the model uses 1xGMF as a reference for determining the no-fault condition. It is normal to see peaks in GMF, even when the system does not present a fault, as identified by the method. On the other hand, when sidebands start to appear and harmonics appear in greater evidence, it is quite likely that a gear fault is happening. Note that in Fig.5.9b, the method identified the harmonics of the GMF and its sidebands as the most relevant frequencies, again validating the manual analysis of an expert.

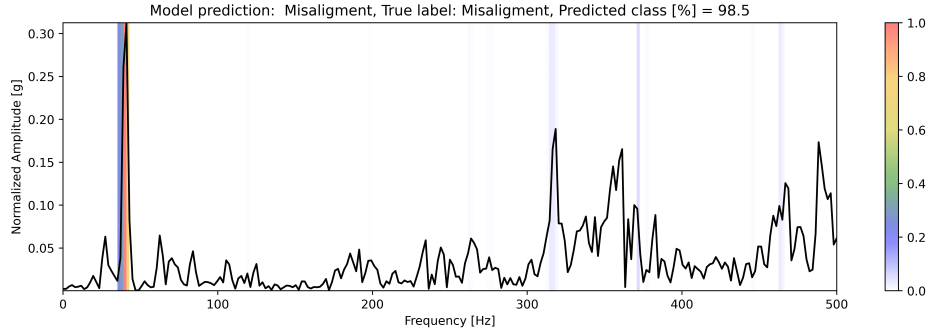
The results obtained for Case 3 are shown in Fig.5.10. The signals are zoomed in the 0 to 500 hz range for easy viewing.



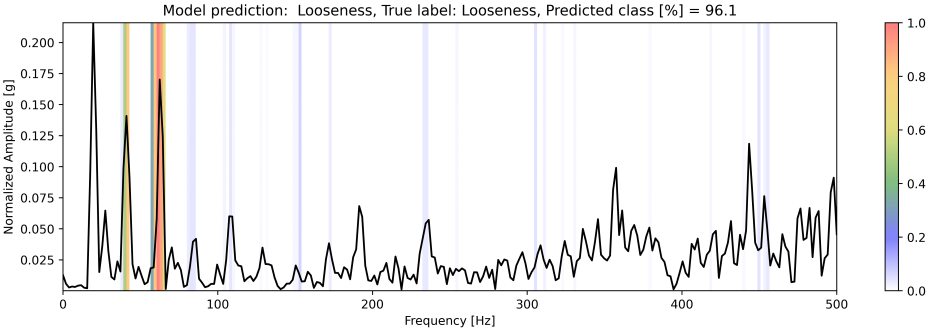
(a) Normal



(b) Unbalance



(c) Misalignment



(d) Looseness

Figure 5.10: XAI analysis - Case 3

For the normal condition, Fig.5.10a, the model did not identify a specific frequency as expected. On the other hand, in Fig.5.10b, there is unbalance, where normally 1 x fr is dominant in the signal. Once again, the model presented the same analysis by an expert,

indicating 1 x fr as the most relevant frequency. In Fig.5.10c, the most relevant frequency is 2 x fr, which is exactly one of the misalignment characteristics, where 2 x fr is greater than 1 x fr, and there may or may not be harmonics of 3x, 4x, 5x etc. The mechanical looseness can present itself in some ways, being one of them the excitation of the fr harmonics. In Fig.5.10d it is noted that the model used the rotation harmonics as a reference, with emphasis on 3 x fr, again confirming the analysis based on human knowledge.

5.6 Conclusions

This chapter presents a new generic and interpretable approach to classifying faults in rotating machinery based on transfer learning from augmented synthetic data to real rotating machinery, namely FaultD-XAI (Fault Diagnosis using eXplainable AI).

The six-stage scheme is adopted: i) Data Acquisition; ii) Signal Generation; iii) Data Augmentation; iv) Signal Processing; v) Fault Diagnosis; vi) Explainable Artificial Intelligence (XAI). From the acquired vibration signals, the synthetic faults are generated and augmented. Subsequently, they are transformed to the frequency domain, and used to train the AI model. After training, the model is able to classify real fault signals, without them having been used in training. Finally, explainability allows analyzing the predictions obtained.

The results show that it is possible to train a model for fault diagnosis in rotating machinery, without having labeled data for all faults, overcoming this major problem in industrial applications. In addition, the proposed approach makes it possible to identify the most relevant features used by the model to perform the prediction, overcoming another major problem, which is to trust the result obtained by black-box models.

In summary, the best results (accuracy and standard deviation) reached in Case 1, 2 and 3 were: 98.1% (1.2%), 92.5% (5.5%) and 95.5% (2.1%), respectively. Using the traditional supervised training method and real data, the results obtained were: 98.1% (0.53%), 98.9% (0.86%) and 99.8% (0.11%), for Case 1, 2 and 3 respectively. Therefore, comparing both results, it can be concluded that the approach presents a promissory proposal for fault diagnosis. Moreover, the model's explainability analysis shows that it was able to learn the most relevant characteristics of each condition, and that the proposed approach is

capable of supporting the expert's final decision-making.

As FaultD-XAI does not require labeled data of all possible machine operating conditions, and only knowledge currently available on fault diagnosis through vibration analysis, the methodology has many possible engineering applications. Future work will explore the approach to remaining useful life (RUL) estimation of rotating machinery components (fault prognosis Susto et al. (2014)), using synthetic data, transfer learning and XAI.

CHAPTER VI

CONCLUSIONS

This thesis presented the study and development of different methodologies for monitoring rotating machinery based on explainable artificial intelligence and vibration analysis.

The methodologies applied made it possible to use synthetic and real experimental data and analyze them quickly and with low computational cost, making the applications viable for the industrial reality. In addition to creating a new dataset to study defects such as misalignment, unbalance and mechanical looseness, which will be publicly available for further research.

The results and discussions show the feasibility of applying the proposed methodologies in the monitoring of rotating machinery.

6.1 Main Conclusions

The main conclusions about the work are presented below.

- The results show that it is possible to use vibration signals for unsupervised fault detection in rotating machinery. Furthermore, the feature extraction process is fundamental for the success of ML models. Techniques to reduce the size of input features have been proposed, showing their feasibility.

- A new approach that allows to automatically select the relevant frequency bands for

an analysis of rotating machinery signals, and to obtain the respective ranking of importance is presented. Through the analysis of the entropy and rms value of the signal, the BRF is obtained allowing the classification of relevance of the band. In automatic systems (e.g., Artificial Intelligence frameworks) the use of the method allows analyzing only bands that contain relevant information, and in manual analyses, it provides the vibration specialist with guidance on the main focuses of the analysis.

- The explainable artificial intelligence approach proposed allows the unsupervised fault detection in rotating machinery. And, in addition to providing explainability about the models used, the methodology provides relevant information for root cause analysis, or even unsupervised fault classification.

- The results for FaultD-XAI, show that it is possible to train a model for fault diagnosis in rotating machinery, without having labeled data for all faults, overcoming this major problem in industrial applications. In addition, the proposed approach makes it possible to identify the most relevant features used by the model to perform the prediction, overcoming another major problem, which is to trust the result obtained by black-box models.

6.2 Other Conclusions

In addition to the main conclusions mentioned above, the following conclusions can be highlighted.

- Among the studied techniques, PCA had the best performance, being second only to the manual selection of features based on the expert's knowledge. The raw signal can also be reduced in size, and later used in ML models for fault detection. It is noteworthy, however, that the amount of principal components needed to represent the problem tends to be greater, which can lead to a reduction in the robustness and assertiveness of the system. Feature trend analysis is an interesting tool to visually verify variations in the system and can be used by the specialist in conjunction with the anomaly score obtained by the model for monitoring.

The work avoids the introduction of irrelevant or correlated features in ML models, reducing data storage space and computational time to train the models. This, allows the specialist to monitor the features in a summarized manner, which is not always possible when there are many monitoring variables. Thus, decision-making is supported by solid indicators, essential for application in an industrial scenario.

- The results obtained for the different dataset show that the BRF is able to identify the relevance of bands both for synthetically generated data, with and without the presence of noise, and for two real dataset.

  Due to its unsupervised characteristic, the methodology can be applied within frameworks as a pre-feature extraction method in data analytics and artificial intelligence applications, avoiding extracting features from irrelevant bands of the signal. It can also be used to verify the correct operation of sensors (quality of the acquired signal), contributing to the reliability of wireless/remote and IoT (Internet of Things) monitoring systems.

- In general, the results obtained for anomaly detection, can be summarized as follows: i) for AD, in Case 1, a maximum F1-Score value of 99.45% was obtained, with the three main models being MCD, HBOS and IF. In Case 2, a maximum value of 99.84% was obtained, and the best models were: MCD, kNN and IF. In Case 3, the maximum value obtained was 99.22%, and the main models being: HBOS, IF and MCD. ii) For the unsupervised classification, in Case 1 a maximum accuracy of 99.57% was obtained using IF, and having HBOS and FB as main models. For Case 2, a maximum value of 96.72% was reached, with the Ensemble, kNN and CBLOF models having the best results. For the root cause analysis, the features that most correlate with the fault under study are presented, as for example in the case of unbalance, where 55.83% of the samples indicated 1xfr as the most important features, in line with the fault behavior, which helps the specialist to eliminate features that are not related to the fault. iii) The comparison between Local-DIFFI and SHAP showed greater similarity for all cases, especially for Case 3, with a Kendall-Tau Distance value equal to 0.455. In terms of computational performance, Local-DIFFI presented a superiority

of approximately 6.5–8.0x over SHAP.

Different state-of-the-art ML algorithms in anomaly detection were studied showing the possibility to change models according to the dataset. The new approach can be applied to different types of faults just by modifying the extracted features associated with a potential fault as shown for the 3 datasets studied. Since the approach does not require previously labeled data, and only knowledge currently available on fault detection through vibration analysis, the methodology has many possible industrial applications.

- In summary, the best results for fault diagnosis (accuracy and standard deviation) reached in Case 1, 2 and 3 were: 98.1% (1.2%), 92.5% (5.5%) and 95.5% (2.1%), respectively. Using the traditional supervised training method and real data, the results obtained were: 98.1% (0.53%), 98.9% (0.86%) and 99.8% (0.11%), for Case 1, 2 and 3 respectively. Therefore, comparing both results, it can be concluded that FaultD-XAI presents a promissory proposal for fault diagnosis. Moreover, the model's explainability analysis shows that it was able to learn the most relevant characteristics of each condition, and that the proposed approach is capable of supporting the expert's final decision-making.

  As FaultD-XAI does not require labeled data of all possible machine operating conditions, and only knowledge currently available on fault diagnosis through vibration analysis, the methodology has many possible engineering applications.

6.3 Future Works

- Future works will explore the possibility of using the BRF value as an anomaly detection / drift concept detection method. Considering the variation in the classification of a band from relevant to irrelevant during a time series may indicate variations or anomalies in the system. Or suggesting that the model needs to be retrained due to a new distribution of the data.

- In addition, the BRF will be studied as a monitoring feature over time (trend analysis) in order to assess its sensitivity as an indicator of the component's end-of-life. Future

developments also include studying the behavior of the method in time series of different applications (e.g., electroencephalogram (EEG) etc).

- New studies will also focus on understanding how the model relates input features with significance to perform anomaly detection, and thus be able to work on dimensionality reduction in a more optimized way.

- Other possibilities are to explore the study of remaining useful life (RUL) estimation of rotating machinery components (fault prognosis), using synthetic data, transfer learning and XAI.

- And finally, to study the domain adaptation and transfer learning associated with methods for model interpretability to improve the applicability of the proposed approaches in different industrial scenarios.

Bibliography

Abdeljaber, O., Avci, O., Kiranyaz, M., Boashash, B., Sodano, H. and Inman, D. (2017), '1-d cnns for structural damage detection: verification on a structural health monitoring benchmark data', *Neurocomputing* **52**. https://doi.org/10.1016/j.neucom.2017.09.069.

Abid, F. B., Sallem, M. and Braham, A. (2020), 'Robust interpretable deep learning for intelligent fault diagnosis of induction motors', *IEEE Transactions on Instrumentation and Measurement* **69**, 3506–3515. https://doi.org/10.1109/TIM.2019.2932162.

Ai, Y.-T., Guan, J.-Y., Fei, C. W., Tian, J. and Zhang, F. L. (2017), 'Fusion information entropy method of rolling bearing fault diagnosis based on n-dimensional characteristic parameter distance', *Mech.Syst.Sig. Process.* **88**, 123–136. https://doi.org/10.1016/j.ymssp.2016.11.019.

Amruthnath, N. and Gupta, T. (2018), 'A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance', *5th ICIEA, Singapore* pp. 355– 361. https://doi.org/10.1109/IEA.2018.8387124.

Antoni, J. (2006), 'The spectral kurtosis: a useful tool for characterising non-stationary signals', *Mech.Syst.Sig. Process.* **20**, 282–307. https://doi.org/10.1016/j.ymssp.2004.09.001.

Antoni, J. (2007), 'Fast computation of the kurtogram for the detection of transient faults', *Mech.Syst.Sig. Process.* **21**, 108–124. https://doi.org/10.1016/j.ymssp.2005.12.002.

Antoni, J. (2016), 'The infogram: entropic evidence of the signature of repetitive transients', *Mech.Syst.Sig. Process.* **74**, 73–94. https://doi.org/10.1016/j.ymssp.2015.04.034.

Appana, D., Ahmad, W. and Kim, J.-M. (2017), 'Speed invariant bearing fault characterization using convolutional', *Neural Networks* **40**, 189–198. https://doi.org/10.1007/978-3-319-69456-6_16.

Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M. and Inman, D. (2018), 'Wireless and real-time structural damage detection: a novel decentralized method for wireless sensor networks', *J. Sound Vib.* . https://doi.org/10.1016/j.jsv.2018.03.008.

Barbariol, T., Feltresi, E. and Susto, G. A. (2020), 'Self-diagnosis of multiphase flow meters through machine learning-based anomaly detection', *Energies* **13**(12), 3136. https://doi.org/10.3390/en13123136.

Barszcz, T. and JabŁonski, A. (2011), 'A novel method for the optimal band selection for vibration signal demodulation and comparison with the kurtogram', *Mech.Syst.Sig. Process.* **25**, 431–451. https://doi.org/10.1016/j.ymssp.2010.05.018.

Bolón Canedo, V., Sánchez Maroño, N. and Alonso Betanzos, A. (2013), 'A review of feature selection methods on synthetic data', *Knowl Inf Syst* **34**, 483 – 519. https://doi.org/10.1007/s10115-012-0487-8.

Bousdekis, A., Magoutas, B., Apostolou, D. and Mentzas, G. (2018), 'Review, analysis and synthesis of prognostic-based decision support methods for condition based maintenance', *Journal of Intelligent Manufact.* **29**, 1303–1316. https://doi.org/10.1007/s10845-015-1179-5.

Bozchalooi, I. and Liang, M. (2007), 'A smoothness index-guided approach to wavelet parameter selection in signal de-noising and fault detection', *J. Sound Vib.* **308**, 246–267. https://doi.org/10.1016/j.jsv.2007.07.038.

Breunig, M. M., Kriegel, H.-P., Ng, R. T. and Sander, J. (2000), 'Lof: identifying density-based local outliers', *ACM* **29**, 93–104. https://doi.org/10.1145/342009.335388.

Brito, J. N. and Pederiva, R. (2002), 'Using artificial intelligence tools to detect problems in induction motors', *In Proceedings of the 1st International Conference on Soft Computing and Intelligent Systems (International Session of 8th SOFT Fuzzy Systems Symposium)*

and 3rd International Symposium on Advanced Intelligent Systems (SCIS and ISIS 2002) **1**, 1–6.

Brito, J. N. and Pederiva, R. (2003), 'A hybrid neural/expert system to diagnose problems in induction motors', *Proceedings of 17th International Congress of Mechanical Engineering* **17**, 1–9.

Brito, L. C., Susto, G. A., Brito, J. N. and Duarte, M. A. V. (2022), 'An explainable artificial intelligence approach for unsupervised fault detection and diagnosis in rotating machinery', *Mechanical Systems and Signal Processing* **163**, 108105. https://doi.org/10.1016/j.ymssp.2021.108105.

Brito, L., Susto, G., Brito, J. and Duarte, M. (2021), 'Fault detection of bearing: An unsupervised machine learning approach exploiting feature extraction and dimensionality reduction', *Informatics* **8(4)**, 85. https://doi.org/10.3390/informatics8040085.

Cao, P., Zhang, S. and Tang, J. (2018), 'Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning', *IEEE Access* **6**, 26241–26253. 10.1109/ACCESS.2018.2837621.

Carletti, M., Masiero, C., Beghi, A. and Susto, G. A. (2019), Explainable machine learning in industry 4.0: evaluating feature importance in anomaly detection to enable root cause analysis, *in* '2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)', IEEE, pp. 21–26. https://doi.org/10.1109/SMC.2019.8913901.

Carletti, M., Terzi, M. and Susto, G. A. (2020), 'Interpretable anomaly detection with diffi: Depth-based feature importance for the isolation forest', *arXiv preprint arXiv:2007.11117* pp. 1–12.

Chen, H. and Lee, C. (2020), 'Vibration signals analysis by explainable artificial intelligence (xai) approach: Application on bearing faults diagnosis', *IEEE Access* **8**, 134246–134256. https://doi.org/10.1109/ACCESS.2020.3006491.

Chen, J., Randall, R. B., Feng, N., Peeter, B. and der Auweraer, H. V. (2013), Automated

diagnostics of internal combustion engine using vibration simulation, *in* 'Proceedings of the ICSV20, Bangkok'.

Cheng, G., Chen, X., Li, H., Li, P. and Liu, H. (2016), 'Study on planetary gear fault diagnosis based on entropy feature fusion of ensemble empirical mode decomposition', *Measurement* **91**, 140–154. https://doi.org/10.1016/j.measurement.2016.05.059.

Chuya-Sumba, J., Alonso-Valerdi, L. and Ibarra-Zarate, D. I. (2022), 'Deep-learning method based on 1d convolutional neural network for intelligent fault diagnosis of rotating machines', *Appl. Sci.* **12**, 2158. https://doi.org/10.3390/app12042158.

Ciabattoni, L., Ferracuti, F., Freddi, A. and Monteriù, A. (2018), 'Statistical spectral analysis for fault diagnosis of rotating machines', *IEEE Transactions on Industrial Electronics* **65**(5), 4301–4310. https://doi.org/10.1109/TIE.2017.2762623.

Cui, L., Huang, J., Zhang, F. and Chu, F. (2019), 'Hvsrms localization formula and localization law: Localization diagnosis of a ball bearing outer ring fault', *Mech.Syst.Sig. Process.* **120**, 608–629. https://doi.org/10.1016/j.ymssp.2018.09.043.

Dalvand, F., Dalvand, S., Sharafi, F. and Pecht, M. (2017), 'Current noise cancellation for bearing fault diagnosis using time shifting', *IEEE Transactions on Industrial Electronics* **64**(10), 8138–8147. https://doi.org/10.1109/TIE.2017.2694397.

Diallo, M., Mokeddem, S., Braud, A., Frey, G. and Lachiche, N. (2021), 'Identifying benchmarks for failure prediction in industry 4.0', *Informatics* **8**(8), 68. https://doi.org/10.3390/informatics8040068.

Ding, X. and He, Q. (2017), 'Energy-fluctuated multiscale feature learning with deep convnet for intelligent spindle bearing fault diagnosis', *IEEE Trans. Instrum. Meas.* **66**, 1926–1935. https://doi.org/10.1109/TIM.2017.2674738.

Doshi-Velez, F. and Kim, B. (2017), 'Towards a rigorous science of interpretable machine learning', *arXiv preprint arXiv:1702.08608* .

D.S.Alves, G.B.Daniel, H.F.deCastro, T.H.Machado, K.L.Cavalca, Gecgel, O., Dias, J. P. and Ekwaro-Osire, S. (2020), 'Uncertainty quantification in deep convolutional neural

network diagnostics of journal bearings with ovalization fault', *Mechanism Mach. Theory* **149**, 103835. https://doi.org/10.1016/j.mechmachtheory.2020.103835.

Du, M., Liu, N. and Hu, X. (2019), 'Techniques for interpretable machine learning', *Communications of the ACM* **63**(1), 68–77. https://doi.org/10.1145/3359786.

Duan, L., Xie, M., Wang, J. and Bai, T. (2018), 'Deep learning enabled intelligent fault diagnosis: Overview and applications', *J. Intell. Fuzzy Syst* **35**, 5771–5784. https://doi.org/10.3233/JIFS-17938.

Endo, H. and Randall, R. (2007), 'Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter', *Mech.Syst.Sig. Process.* **21**, 906–919. https://doi.org/10.1016/j.ymssp.2006.02.005.

Eren, L., Ince, T. and Kiranyaz, S. (2019), 'A generic intelligent bearing fault diagnosis system using compact adaptive 1d cnn classifier', *J. Signal Process. Syst.* **91**, 179–189. https://doi.org/10.1007/s11265-018-1378-3.

Feng, F., Wu, C., Zhu, J., Wu, S., Tian, Q. and Jiang, P. (2020), 'Research on multitask fault diagnosis and weight visualization of rotating machinery based on convolutional neural network', *Journal of the Brazilian Society of Mechanical Sciences and Engineering* **42(11)**, 1–14. https://doi.org/10.1007/s40430-020-02688-6.

Gecgel, O., Dias, J., Ekwaro-Osire, S., Alves, D. S., Machado, T. H., Daniel, G. B. and Cavalca, K. L. (2021), 'Simulation-driven deep learning approach for wear diagnostics in hydrodynamic journal bearings', *Journal of Tribology* **143**, 8. https://doi.org/10.1115/1.4049067.

Goodfellow, I., Bengio, Y. and Courville, A. (2016), *Deep Learning*, MIT Press. `http://www.deeplearningbook.org`.

Grezmak, J., Wang, P., Sun, C. and Gao, R. X. (2019), 'Explainable convolutional neural network for gearbox fault diagnosis', *Procedia CIRP* **80**, 476–481. https://doi.org/10.1016/j.procir.2018.12.008.

Grezmak, J., Zhang, J., Wang, P., Loparo, K. A. and Gao, R. X. (2020), 'Interpretable convolutional neural network through layer-wise relevance propagation for machine fault diagnosis', *IEEE Sensors Journal* **20**, 3172–3181. https://doi.org/10.1109/JSEN.2019.2958787.

Guo, L., Lei, Y., Xing, S., Yan, T. and Li, N. (2019), 'Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data', *IEEE Transactions on Industrial Electronics* **66**(9), 7316–7325. https://doi.org/10.1109/TIE.2018.2877090.

Handikherkar, V. C. and Phalle, V. M. (2021), 'Gear fault detection using machine learning techniques-a simulation-driven approach', *International Journal of Engineering* **34**, 212–223. https://doi.org/10.5829/ije.2021.34.01a.24.

Hans-Peter Kriegel, M. S. and Zimek, A. (2008), 'Angle-based outlier detection in high-dimensional data', *In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* **14**, 444–452. https://doi.org/10.1145/1401890.1401946.

Hasan, M. and Kim, J. (2018), 'Bearing fault diagnosis under variable rotational speeds using stockwell transform-based vibration imaging and transfer learning', *Appl. Sci.* **8**, 2357. https://doi.org/10.3390/app8122357.

Hasegawa, T., Ogata, J., Murakawa, M. and Ogawa, T. (2017), 'Adaptive training of vibration-based anomaly detector for wind turbine condition monitoring', *Annual Conference on PHM Society* pp. 1–8. https://doi.org/10.36001/ijphm.2017.v8i2.2634.

Hasegawa, T., Ogata, J., Murakawa, M. and Ogawa, T. (2018), 'Tandem connectionist anomaly detection: Use of faulty vibration signals in feature representation learning', *IEEE PHM, Seattle* pp. 1–7. https://doi.org/10.1109/ICPHM.2018.8448450.

Hawkins, D. (1980), *Identification of outliers*, Chapman and Hall. https://doi.org/10.1007/978-94-015-3994-4.

H.Chen, Hu, N., Cheng, Z., Zhang, L. and Y.Zhang (2019), 'A deep convolutional neural network based fusion method of two-direction vibration signal data for health state identification of planetary gearboxes', *Measurement* **146**, 268–278. https://doi.org/10.1016/j.measurement.2019.04.093.

He, K., Zhang, X., Ren, S. and Sun, J. (2016), 'Deep residual learning for image recognition', *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* . https://doi.org/10.1109/CVPR.2016.90.

He, Z., Xu, X. and Deng, S. (2003), 'Discovering cluster-based local outliers', *Pattern Recognition Letters* **24(9-10)**, 1641–1650. https://doi.org/10.1016/S0167-8655(03)00003-5.

Hebda-Sobkowicz, J., Zimroz, R., Pitera, M. and Wyłomańska, A. (2020), 'Informative frequency band selection in the presence of non-gaussian noise – a novel approach based on the conditional variance statistic with application to bearing fault diagnosis', *Mech.Syst.Sig. Process.* **145**, 106971. https://doi.org/10.1016/j.ymssp.2020.106971.

Hendrickx, K., Meert, W., Mollet, Y., Gyselinck, J., Cornelis, B., Gryllias, K. and Davis., J. (2020), 'A general anomaly detection framework for fleet-based condition monitoring of machines', *Mechanical Systems and Signal Processing* **139**, 106585. https://doi.org/10.1016/j.ymssp.2019.106585.

Hoang, D. and Kang, H. (2019), 'A survey on deep learning based bearing fault diagnosis', *Neurocomputing* **335**, 327–335. https://doi.org/10.1016/j.neucom.2018.06.078.

Ince, T., Kiranyaz, S., Eren, L., Askar, M. and Gabbouj, M. (2016), 'Real-time motor fault detection by 1-d convolutional neural networks', *IEEE Trans. Ind. Electron* **63**, 7067–7075. https://doi.org/10.1109/TIE.2016.2582729.

ISO (2009), *Mechanical Vibration — Evaluation Of Machine Vibration By Measurements On Non-Rotating Parts — Part 3*, ISO 10816-3:2009 edn, International Organization for Standardization.

Jia, F., Lei, Y., Lu, N. and Xing, S. (2018), 'Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding

via visualization', *Mechanical Systems and Signal Processing* **110**, 349–367. https://doi.org/10.1016/j.ymssp.2018.03.025.

Jolliffe, I. (1986), *Principal Component Analysis*, New York: Springer-Verlag. https://doi.org/10.1007/978-1-4757-1904-8.

Jutten, C. and Hérault, J. (1991), 'Blind separation of sources, part i: an adaptive algorithm based on neuromimetic architecture', *Signal Process* **24**. https://doi.org/10.1016/0165-1684(91)90079-X.

Khan, A., Hwang, H. and Kim, H. S. (2021), 'Synthetic data augmentation and deep learning for the fault diagnosis of rotating machines', *Mathematics* **9**, 2336. https://doi.org/10.3390/math9182336.

Khan, S. and Yairi, T. (2018), 'A review on the application of deep learning in system health management', *Mech. Syst. Signal Process* **107**, 241–265. https://doi.org/10.1016/j.ymssp.2017.11.024.

Kim, J. and Kim, J.-M. (2020), 'Bearing fault diagnosis using grad-cam and acoustic emission signals', *Appl. Sci.* **10**, 2050. https://doi.org/10.3390/app10062050.

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M. and Inman, D. J. (2021), '1d convolutional neural networks and applications: A survey', *Mechanical Systems and Signal Processing* **151**, 107398. https://doi.org/10.1016/j.ymssp.2020.107398.

Kiranyaz, S., Gastli, A., Ben-Brahim, L., Alemadi, N. and Gabbouj, M. (2018), 'Real-time fault detection and identification for mmc using 1d convolutional neural networks', *IEEE Trans. Ind. Electron* . https://doi.org/10.1109/TIE.2018.2833045.

Kiranyaz, S., Ince, T. and Gabbouj, M. (2016), 'Real-time patient-specific ecg classification by 1-d convolutional neural networks', *IEEE Trans. Biomed. Eng.* **63**, 664–675. https://doi.org/10.1109/TBME.2015.2468589.

Knorr, E. and Ng, R. (1998), 'Algorithms for mining distance-based outliers in large datasets', *Proceedings of the 24rd International Conference on Very Large Data Bases* **24**, 392–403.

Kriegel, H.-P., Kröger, P., Schubert, E. and Zimek, A. (2009), Loop: local outlier probabilities, *in* 'Proceedings of the 18th ACM conference on Information and knowledge management', pp. 1649–1652. https://doi.org/10.1145/1645953.1646195.

Krizhevsky, A., Sutskever, I. and Hinton, G. (2012), 'Imagenet classification with deep convolutional neural networks', *Adv. Neural Inf. Process. Syst.* **25**, 1097–1105.

Kumar, P. and Hati, A. S. (2020), 'Review on machine learning algorithm based fault detection in induction motors', *Arch Computat Methods Eng* pp. 1–12. https://doi.org/10.1007/s11831-020-09446-w.

Lazarevic, A. and Kumar, V. (2005), 'Feature bagging for outlier detection', *In Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining* p. 157–166. https://doi.org/10.1145/1081870.1081891.

Lecun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998), 'Gradient-based learning applied to document recognition', *Proceedings of the IEEE* **86 (11)**, 2278–2324. https://doi.org/10.1109/5.726791.

Lee, J., Qiu, H., Yu, G. and Lin, J. (2007), 'Bearing dataset', *IMS, University of Cincinnati, NASA Ames Prognostics Data Repository, Rexnord Technical Services, Moffett Field, CA, Available online: https://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/bearing (accessed on 14 November 2021)* .

Lei, Y. (2017), *Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery*, Elsevier Inc. https://doi.org/10.1016/B978-0-12-811534-3.00006-8.

Lei, Y., He, Z., Zi, Y. and Hu, Q. (2007), 'Fault diagnosis of rotating machinery based on multiple anfis combination with gas', *Mechanical Systems and Signal Processing* **21**(5), 2280 – 2294. https://doi.org/10.1016/j.ymssp.2006.11.003.

Lei, Y., Jia, F., Lin, J., Xing, S. and Ding, S. (2016), 'An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data', *IEEE Transactions on Industrial Electronics* **63**, 3137–3147. https://doi.org/10.1109/TIE.2016.2519325.

Lei, Y., Lin, J., He, Z. and Zuo, M. J. (2013), 'A review on empirical mode decomposition in fault diagnosis of rotating machinery', *Mech. Syst. Signal Process* **35**, 108–126. https://doi.org/10.1016/j.ymssp.2012.09.015.

Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N. and Nandi, A. K. (2020), 'Applications of machine learning to machine fault diagnosis: A review and roadmap', *Mechanical Systems and Signal Processing* **138**, 106587. https://doi.org/10.1016/j.ymssp.2019.106587.

Lei, Y. and Zuo, M. J. (2009), 'Gear crack level identification based on weighted k nearest neighbor classification algorithm', *Mechanical Systems and Signal Processing* **23**(5), 1535 – 1547. https://doi.org/10.1016/j.ymssp.2009.01.009.

Lei, Y., Zuo, M. J., He, Z. and Zi, Y. (2010), 'A multidimensional hybrid intelligent method for gear fault diagnosis', *Expert Systems with Applications* **37**(2), 1419 – 1430. https://doi.org/10.1016/j.eswa.2009.06.060.

Li, C., Cabrera, D., Sancho, F., Sanchez, R. V., Cerrada, M. and de Oliveira, J. V. (n.d.), 'One-shot fault diagnosis of 3d printers through improved feature space learning', *IEEE Transactions on Industrial Electronics* .

Li, C., Zhang, S., Qin, Y. and Estupinan, E. (2020), 'A systematic review of deep transfer learning for machinery fault diagnosis', *Neurocomputing* **407**, 121–135. https://doi.org/10.1016/j.neucom.2020.04.045.

Li, S., Liu, G., Tang, X., Lu, J. and Hu, J. (2017), 'An ensemble deep convolutional neural network model with improved d-s evidence fusion for bearing fault diagnosis', *Sensors* **41**, 1729. https://doi.org/10.3390/s17081729.

Li, T., Zhao, Z., Sun, C., Cheng, L., Chen, X., Yan, R. and Gao, R. X. (2019), 'Waveletkernelnet: An interpretable deep neural network for industrial intelligent diagnosis', *arXiv:1911.07925v3* pp. 1–9.

Li, X., Zhang, W. and Ding, Q. (2019), 'Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism', *Signal Processing* **161**, 136–154. https://doi.org/10.1016/j.sigpro.2019.03.019.

Li, X., Zhang, W., Ding, Q. and Sun, J.-Q. (2020), 'Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation', *J Intell Manuf* **31**, 433–452. https://doi.org/10.1007/s10845-018-1456-1.

Li, Y., Li, G., Yang, Y., Liang, X. and Xu, M. (2018), 'A fault diagnosis scheme for planetary gearboxes using adaptive multi-scale morphology filter and modified hierarchical permutation entropy', *Mech.Syst.Sig. Process.* **105**, 319–337. https://doi.org/10.1016/j.ymssp.2017.12.008.

Li, Y., Wang, X., Liu, Z., Liang, X. and Si, S. (2018), 'The entropy algorithm and its variants in the fault diagnosis of rotating machinery: A review', *IEEE Access* **6**, 66723–66741. https://doi.org/10.1109/ACCESS.2018.2873782.

Li, Y., Yang, Y., Li, G., Xu, M. and Huang, W. (2017), 'A fault diagnosis scheme for planetary gearboxes using modified multi-scale symbolic dynamic entropy and mrmr feature selection', *Mechanical Systems and Signal Processing* **91**, 295 – 312. https://doi.org/10.1016/j.ymssp.2016.12.040.

Lin, C. J. and Jhang, J. Y. (2021), 'Bearing fault diagnosis using a grad-cam-based convolutional neuro-fuzzy network', *Mathematics* **9(13)**, 1502. https://doi.org/10.3390/math9131502.

Liu, F. T., Ting, K. M. and Zhou, Z.-H. (2008), 'Isolation forest', *Proceedings of the 2008 IEEE International Conference on Data Mining (ICDM'08),IEEE* p. 413–422. https://doi.org/10.1109/ICDM.2008.17.

Liu, F. T., Ting, K. M. and Zhou, Z.-H. (2012), 'Isolation-based anomaly detection', *ACM Trans. Knowl. Discov. Data* **6**, 1–39. https://doi.org/10.1145/2133360.2133363.

Liu., L., Zhi, Z., Zhang, H., Guo, Q., Peng, Y. and Liu, D. (2019), 'Related entropy theories application in condition monitoring of rotating machineries', *Entropy* **21**, 1061. https://doi.org/10.3390/e21111061.

Liu, Q., Zhang, J., Liu, J. and Yang, Z. (2022), 'Feature extraction and classification algorithm, which one is more essential? an experimental study on a specific

task of vibration signal diagnosis', *Int. J. Mach. Learn. and Cyber.* **13**, 1685–1696. https://doi.org/10.1007/s13042-021-01477-4.

Liu, R., Yang, B., Zio, E. and Chen, X. (2018), 'Artificial intelligence for fault diagnosis of rotating machinery: A review', *Mechanical Systems and Signal Processing* **108**, 33 – 47. https://doi.org/10.1016/j.ymssp.2018.02.016.

Liu, T., Chen, J., Dong, G., Xiao, W. and Zhou, X. (2013), 'The fault detection and diagnosis in rolling element bearings using frequency band entropy', *Proc. Inst. Mech. Eng. C, J. Mech. Eng. Sci.* **227**, 87–99. https://doi.org/10.1177/0954406212441886.

Liu, Z., Jin, Y., Zuo, M. J. and Peng, D. (2019), 'Accugram: A novel approach based on classification to frequency band selection for rotating machinery fault diagnosis', *ISA Transactions* **95**, 346–357. https://doi.org/10.1016/j.isatra.2019.05.007.

Lu, C., Wang, Z. and Zhou, B. (2017), 'Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification', *Adv. Eng. Inform* **32**, 139–151. https://doi.org/10.1016/j.aei.2017.02.005.

Luo, S., Cheng, J., Zeng, M. and Yang, Y. (2016), 'An intelligent fault diagnosis model for rotating machinery based on multi-scale higher order singular spectrum analysis and ga-vpmcd', *Measurement* **87**, 38–50. https://doi.org/10.1016/j.measurement.2016.01.006.

M, L. S. and Su-In, L. (2017), 'A unified approach to interpreting model predictions', *Advances in NIPS* **30**, 4765–4774.

Meneghetti, L., Terzi, M., Del Favero, S., Susto, G. A. and Cobelli, C. (2018), 'Data-driven anomaly recognition for unsupervised model-free fault detection in artificial pancreas', *IEEE Transactions on Control Systems Technology* . https://doi.org/10.1109/CDC.2018.8619048.

Miao, Y., Wang, J., Zhang, B. and Li, H. (2022), 'Practical framework of gini index in the application of machinery fault feature extraction', *Mech.Syst.Sig. Process.* **165**, 108333. https://doi.org/10.1016/j.ymssp.2021.108333.

Miao, Y., Zhao, M. and Lin, J. (2017), 'Improvement of kurtosis-guided-grams via gini index for bearing fault feature identification', *Measur. Sci. Technol.* **28**, 125001. https://doi.org/10.1088/1361-6501/aa8a57.

Molnar, C. (2020), *Interpretable Machine Learning*, Lulu.com.

Moshrefzadeh, A. and Fasana, A. (2018), 'The autogram: an effective approach for selecting the optimal demodulation band in rolling element bearings diagnosis', *Mech.Syst.Sig. Process.* **105**, 294–318. https://doi.org/10.1016/j.ymssp.2017.12.009.

Obuchowski, J., Wyłomańska, A. and Zimroz, R. (2014), 'Selection of informative frequency band in local damage detection in rotating machinery', *Mech Syst Signal Process* **48**, 138–152. https://doi.org/10.1016/j.ymssp.2014.03.011.

Ogata, J. and Murakawa, M. (2016 in Bilbao, Spain.), 'Vibration-based anomaly detection using flac features for wind turbine condition monitoring', *8th European Workshop on Structural Health Monitoring (EWSHM 2016)* **July 5-8,2016**.

Pan, S. J. and Yang, Q. (2010), 'A survey on transfer learning', *IEEE Trans. Knowl. Data Eng.* **22**, 213–237. https://doi.org/10.1109/TKDE.2009.191.

Pan, Y., Chen, J. and Li, X. (2008), 'Spectral entropy: A complementary index for rolling element bearing performance degradation assessment', *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* **223**, 1223–1231. https://doi.org/10.1243/09544062JMES1224.

Pevn'y, T. (2016), 'Loda: lightweight on-line detector of anomalies', *Machine Learning* **102(2)**, 275–304. https://doi.org/10.1007/s10994-015-5521-0.

Qiu, H., Lee, J., Lin, J. and Yu, G. (2006), 'Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics', *Journal of Sound and Vibration* **289**(4), 1066 – 1090. https://doi.org/10.1016/j.jsv.2005.03.007.

Rai, A. K. and Dwivedi, R. K. (2020), Fraud detection in credit card data using unsupervised machine learning based scheme, *in* '2020 International Conference on

Electronics and Sustainable Communication Systems (ICESC)', IEEE, pp. 421–426. https://doi.org/10.1109/ICESC48915.2020.9155615.

Ramaswamy, S. and K.S., R. R. (2000), 'Efficient algorithms for mining outliers from large data sets', *Sigmod Record* **29**, 427–438. https://doi.org/10.1145/335191.335437.

Randall, R. B. and Antoni, J. (2011), 'Rolling element bearing diagnostics—a tutorial', *Mechanical Systems and Signal Processing* **25**, 485–520. https://doi.org/10.1016/j.ymssp.2010.07.017.

Rostaghi, M. and Azami, H. (2016), 'Dispersion entropy: A measure for time-series analysis', *IEEE Signal Process. Lett.* **23**, 610–614. https://doi.org/10.1109/LSP.2016.2542881.

Rousseeuw, P. J. and Driessen, K. V. (1999), 'A fast algorithm for the minimum covariance determinant estimator', *Technometrics* **41(3)**, 212–223. https://doi.org/10.1080/00401706.1999.10485670.

Saeki, M., Ogata, J., Murakawa, M. and Ogawa, T. (2019), 'Visual explanation of neural network based rotation machinery anomaly detection system', *IEEE International Conference on Prognostics and Health Management, San Francisco* pp. 1–4. https://doi.org/10.1109/ICPHM.2019.8819396.

Samuel, P. D. and Pines, D. J. (2005), 'A review of vibration-based techniques for helicopter transmission diagnostics', *Journal of Sound and Vibration* **282**(1), 475 – 508. https://doi.org/10.1016/j.jsv.2004.02.058.

Sassi, S., Badri, B. and Thomas, M. (2007), 'A numerical model to predict damaged bearing vibrations', *J. Vib. Contr.* **13**, 1603–1628. https://doi.org/10.1177/1077546307080040.

Saufi, S. R., Ahmad, Z. A. B., Leong, M. S. and Lim, M. H. (2019), 'Challenges and opportunities of deep learning models for machinery fault detection and diagnosis: A review', *IEEE Access* **7**, 122644–122662. https://doi.org/10.1109/ACCESS.2019.2938227.

Schmidt, S., Heyns, P. S. and Gryllias, K. C. (2021), 'An informative frequency band identification framework for gearbox fault diagnosis under time-varying

operating conditions', *Mechanical Systems and Signal Processing* **158**, 107771. https://doi.org/10.1016/j.ymssp.2021.107771.

Schubert, E., Wojdanowski, R., Zimek, A. and Kriegel, H.-P. (2012), On evaluation of outlier rankings and outlier scores, *in* 'Proceedings of the 2012 SIAM International Conference on Data Mining', SIAM, pp. 1047–1058. https://doi.org/10.1137/1.9781611972825.90.

Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J. and Williamson, R. C. (2001), 'Estimating the support of a high-dimensional distribution', *Neural computation* **13(7)**, 1443–1471. https://doi.org/10.1162/089976601750264965.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D. (2017), 'Grad-cam: Visual explanations from deep networks via gradient-based localization', *2017 IEEE International Conference on Computer Vision (ICCV)* pp. 618–626. https://doi.org/10.1109/ICCV.2017.74.

Shannon, C. E. (1948), 'A mathematical theory of communication', *Bell Syst. Tech. J.* **27**, 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Shen, F., Chen, C., Yan, R. and Gao, R. (2015), Bearing fault diagnosis based on svd feature extraction and transfer learning classification, *in* 'Prognostics and System Health Management Conference, Beijing, China (2015', PHM, pp. 1–6.

Singh, M. and Shaik, A. G. (2019), 'Faulty bearing detection, classification and location in a three-phase induction motor based on stockwell transform and support vector machine', *Measurement* **131**, 524 – 533. https://doi.org/10.1016/j.measurement.2018.09.013.

Smith, W. A. and Randall, R. B. (2015), 'Rolling element bearing diagnostics using the case western reserve university data: A benchmark study', *Mechanical Systems and Signal Processing* **64-65**, 100 – 131. https://doi.org/10.1016/j.ymssp.2015.04.021.

Sobie, C., Freitas, C. and Nicolai, M. (2018), 'Simulation-driven machine learning: Bearing fault classification', *Mechanical Systems and Signal Processing* **99**, 403–419. https://doi.org/10.1016/j.ymssp.2017.06.025.

Stetco, A., Dinmohammadi, F., Zhao, X., Robu, V., Flynn, D., Barnes, M., Keane, J. and Nenadic, G. (2019), 'Machine learning methods for wind turbine condition monitoring: A review', *Renewable Energy* **133**, 620 – 635. https://doi.org/10.1016/j.renene.2018.10.047.

Strange, H. and Zwiggelaar, R. (2014), *Open Problems in Spectral Dimensionality Reduction*, Briefs in Computer Science. https://doi.org/10.1007/978-3-319-03943-5.

Susto, G. A., Schirru, A., Pampuri, S., McLoone, S. and Beghi, A. (2014), 'Machine learning for predictive maintenance: A multiple classifier approach', *IEEE transactions on industrial informatics* **11**(3), 812–820. https://doi.org/10.1109/TII.2014.2349359.

Sánchez, R.-V., Lucero, P., J.Macancela, Alonso, H., Cerrada, M., Cabrera, D. and Castejón, C. (2020), 'Evaluation of time and frequency condition indicators from vibration signals for crack detection in railway axles', *Applied Sciences* **10**, 12:4367. https://doi.org/10.3390/app10124367.

Sánchez, R.-V., Lucero, P., Vásquez, R., Cerrada, M., Macancela, J.-C., and Cabrera, D. (2018), 'Feature ranking for multi-fault diagnosis of rotating machinery by using random forest and knn', *Journal of Intelligent and Fuzzy Systems* **34**, 3463–3473. https://doi.org/10.3233/JIFS-169526.

Tang, S., Yuan, S. and Zhu, Y. (2020), 'Data preprocessing techniques in convolutional neural network based on fault diagnosis towards rotating machinery', *IEEE Access* **8**, 149487–149496. https://doi.org/10.1109/ACCESS.2020.3012182.

Tenenbaum, J., de Silva, V. and Langford, J. (2000), 'A global geometric framework for nonlinear dimensionality reduction', *Science* **290**, 2319–2322. https://doi.org/10.1126/science.290.5500.2319.

VanDerMaaten, L. and Hinton, G. (2008), 'Visualizing data using t-sne', *Jour. of Machine Learning Research* **9**.

von Birgelen, A., Buratti, D., Mager, J. and Niggemann, O. (2018), 'Self-organizing maps for anomaly localization and predictive maintenance in cyber-physical production systems', *Procedia CIRP* **72**, 480–485. https://doi.org/10.1016/j.procir.2018.03.150.

Wang, B., Lei, Y., Li, N. and Li, N. (2020), 'A hybrid prognostics approach for estimating remaining useful life of rolling element bearings', *IEEE Transactions on Reliability* **69**, 401–412. https://doi.org/10.1109/TR.2018.2882682.

Wang, D. (2018*a*), 'Some further thoughts about spectral kurtosis, spectral l2/l1 norm, spectral smoothness index and spectral gini index for characterizing repetitive transients', *Mech.Syst.Sig. Process.* **108**, 360–368. https://doi.org/10.1016/j.ymssp.2018.02.034.

Wang, D. (2018*b*), 'Spectral l2/l1 norm: a new perspective for spectral kurtosis for characterizing non-stationary signals', *Mech.Syst.Sig. Process.* **104**, 290–293. https://doi.org/10.1016/j.ymssp.2017.11.013.

Wang, G., Zhao, Y., Zhang, J. and Ning, Y. (2021), 'A novel end-to-end feature selection and diagnosis method for rotating machinery', *Sensors* **21**, 2056. https://doi.org/10.3390/s21062056.

Wang, X., Mao, D. and Li, X. (2021), 'Bearing fault diagnosis based on vibro-acoustic data fusion and 1d-cnn network', *Measurement* **173**, 108518.

Wang, Y., Lu, C., Liu, H. and Wang, Y. (2016), 'Fault diagnosis for centrifugal pumps based on complementary ensemble empirical mode decomposition, sample entropy and random forest', *Proc. 12th World Congr. Intell. Control Autom. (WCICA)* p. 1317–1320. https://doi.org/10.1109/WCICA.2016.7578401.

Wei, Y., Li, Y., Xu, M. and Huang, W. (2019), 'A review of early fault diagnosis approaches and their applications in rotating machinery', *Entropy* **21**(4), 409. https://doi.org/10.3390/e21040409.

Wu, X., Zhang, Y., Cheng, C. and Peng, Z. (2021), 'A hybrid classification autoencoder for semi-supervised fault diagnosis in rotating machinery', *Mechanical Systems and Signal Processing* **14**, 107327. https://doi.org/10.1016/j.ymssp.2020.107327.

Yan, X. and Jia, M. (2019), 'Intelligent fault diagnosis of rotating machinery using improved multiscale dispersion entropy and mrmr feature selection', *Knowledge-Based Systems* **163**, 450–471. https://doi.org/10.1016/j.knosys.2018.09.004.

Yang, B., Lei, Y., Jia, F. and Xing, S. (2019), 'An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings', *Mechanical Systems and Signal Processing* **122**, 692–706. https://doi.org/10.1016/j.ymssp.2018.12.051.

Yibing, L., Li, Z., Li, J. and Xiangyu, Z. (2019), 'Fault diagnosis of rotating machinery based on combination of deep belief network and one-dimensional convolutional neural network', *IEEE Access* **7**, 165710–165723. https://doi.org/10.1109/ACCESS.2019.2953490.

Yu, S., Wang, M., Pang, S., Song, L. and Qiao, S. (2022), 'Intelligent fault diagnosis and visual interpretability of rotating machinery based on residual neural network', *Measurement* **196**, 111228. https://doi.org/10.1016/j.measurement.2022.111228.

Zak, G., Wyłomanska, A. and Zimroz, R. (2015), 'Application of alpha-stable distribution approach for local damage detection in rotating machines', *Journal of Vibroengineering* **17**, 2987–3002.

Zhang, K., Li, Y., Scarf, P. and Ball, A. (2011), 'Feature selection for high-dimensional machinery fault diagnosis data using multiple models and radial basis function networks', *Neurocomputing* **74**(17), 2941 – 2952. https://doi.org/10.1016/j.neucom.2011.03.043.

Zhang, W., Peng, G., Li, C., Chen, Y. and Zhang, Z. (2017), 'A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals', *Sensors* **17**, 425. https://doi.org/10.3390/s17020425.

Zhang, X., Zhang, Q., Chen, M., Sun, Y., Qin, X. and Li, H. (2018), 'A two-stage feature selection and intelligent fault diagnosis method for rotating machinery using hybrid filter and wrapper method', *Neurocomputing* **275**, 2426 – 2439. https://doi.org/10.1016/j.neucom.2017.11.016.

Zhang, Y., Hutchinson, P., Lieven, N. and Nunez-Yanez, J. (2019), 'Adaptive event-triggered anomaly detection in compressed vibration data', *Mechanical Systems and Signal Processing* **122**, 480–501. https://doi.org/10.1016/j.ymssp.2018.12.039.

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P. and Gao, R. X. (2019), 'Deep learning and

its applications to machine health monitoring', *Mech. Syst. Signal Process* **115**, 213–237. https://doi.org/10.1016/j.ymssp.2018.05.050.

Zhao, Y., Nasrullah, Z. and Li, Z. (2019), 'Pyod: A python toolbox for scalable outlier detection', *Journal of Machine Learning Research* **20**(96), 1–7.

Zheng, J., Pan, H. and Cheng, J. (2017), 'Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines', *Mech.Syst.Sig. Process.* **85**, 746–759. https://doi.org/10.1016/j.ymssp.2016.09.010.

Zheng, J., Pan, H., Yang, S. and Cheng, J. (2018), 'Generalized composite multi-scale permutation entropy and laplacian score based rolling bearing fault diagnosis', *Mech.Syst.Sig. Process.* **99**, 229–243. https://doi.org/10.1016/j.ymssp.2017.06.011.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. and Torralba, A. (2016), Learning deep features for discriminative localization, *in* 'Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016', IEEE, p. 2921–2929.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H. and He, Q. (2021), 'A comprehensive survey on transfer learning', *Proceedings of the IEEE* **109**, 43–76. https://doi.org/10.1109/JPROC.2020.3004555.

Zocco, F., Maggipinto, M., Susto, G. A. and McLoone, S. (2021), 'Greedy search algorithms for unsupervised variable selection: A comparative study'.