U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Bearing Fault Detection: A Feature-based Approach

Paulo Roberto Rocha Sousa

Master's Degree Dissertation presented to:
Faculdade de Engenharia da Universidade do Porto

Supervisor at FEUP:
Prof. José Dias Rodrigues

Supervisor at Siemens:
Eng. Carina Freitas

OPorto, January 2018

Bearing Fault Detection: A Feature-based Approach

Paulo Roberto Rocha Sousa

Dissertation developed at Siemens PLM Software in Leuven and submitted to Faculdade de Engenharia da Universidade do Porto for the degree of Master in Mechanical Engineering

OPorto, January 2018

LaTeX

Bearing Fault Detection: A Feature-based Approach
P. Sousa
January, 2018
FEUP-U.PORTO

To my girlfriend, Joana
To my Mother

Strive for perfection in everything you do.
Take the best that exists and make it better.
When it does not exist, design it.
- Sir Henry Royce

Acknowledgements

After an intensive period of five months, today is the day: writing this note of thanks is the finishing touch on my dissertation. The people that came across my journey have not only contributed to this thesis, but also helped in shaping me into the individual that I am today.

I would like to express my gratitude to Eng. Carina Freitas and Kilian Hendricks for all the help, inspiration and the interesting topics they suggested that added value to my work. Their help revealed essential to the development of my thesis, not only by providing knowledge, but also for the remarkable guidance they provided.

I would like to acknowledge Siemens PLM software for the opportunity given. I would also like to thank everyone in the office for the company and environment that surrounded me.

I would also like to show gratitude to Faculdade de Engenharia da Universidade do Porto, for the amazing years I had while studying there and the great education and resources put available to its students. I would also like to thank Prof. José Dias Rodrigues for accepting this challenge from the start, for sparking my interest in mechanical vibrations and for encouraging me to be seek a rigorous attitude towards my work. I would also like to express my appreciation to Prof. Paulo Tavares de Castro for always being available to help me whenever I needed and whose perseverance and energy inspired me.

Because nothing is possible without friends, I would like to thank all of the *Batcave* crew: Nuno Coelho, Pedro Silva, Sérgio João Moreira, Hugo Moreira, Ana Pereira and many others for the friendship and support.

To my friend Paulo Morais, for all the help and advices during the degree. To João Trindade and Sérgio Esteves for the long-lasting friendship and brotherhood. To Tiago Martins for the unexpected friendship, brotherhood, support and the laughs whenever we most needed.

I would also like to give a special thanks to Laura Pinto and Custódio Oliveira, for all they have done for me.

To my girlfriend Joana Pinto, for the patience needed to put up with me, the understanding and support she provided, not only during this thesis but throughout my degree. Thank you for never giving up on me when a sane person would, for believing in me when everyone else thought otherwise, for always pushing and encouraging me to go further and be the best version of myself. Thank you for all the courage, love and companionship you gave me through these years. I dedicate this thesis to you.

To my mother, to whom I also dedicate this thesis. Thank you for everything, I hope you're proud up there. *Com Saudade.*

Abstract

Rolling element bearings are components widely used in industrial applications. Their failure leads to economical and logistical losses, considering production loss or repairing and replacement costs.

A suitable and predictive plan of maintenance would maximize times between intervals while constinuously monitoring the equipment, looking for a fault. If one is detected, then its evolution is monitored and replacement is incorporated in the maintenance plan according to the estimation of the remaining life time.

Even though many other techniques are available, vibration monitoring is still the most used technique for condition monitoring, due to its robustness and accuracy in fault identification. The tool developed in this work applies a series of different signal processing techniques to enhance the fault's signature signal: Angular resampling, Cepstral Editing Procedure, Cyclic Spectral Correlation and Envelope analysis with filtering finishing with feature calculation, or feature extraction.

Additionally, this project paper suggests a method for semi-automatic signal processing based in the identification of the frequency band and feature selection.

Furthermore this paper covers various papers on condition monitoring, signal processing of vibration signals, and finally feature selection and classification.

The final goal is, along with the signal processing method, to develop a method for class classification using *Python* that can classify a specific class with good accuracy.

Finally, some results are shown where one can extrapolate some conclusions about the efficiency of the different processes used in this method.

Resumo

Os rolamentos, de esferas ou de rolos, são amplamente utilizados em máquinas cujos componentes funcionem segundo movimentos de rotação. A falha de um simples rolamento leva a perdas económicas, custos de reparação e substituição e, por vezes, custos logísticos para a empresa, uma vez que leva à paragem do equipamento e, consequentemente, paragem da cadeia de produção.

Um plano de manutenção adequado, através de manutenção preditiva, maximiza os tempos em que o equipamento está em funcionamento, ao mesmo tempo que o monitoriza constantemente, procurando uma falha. Se tal for detectada, esta é monitorizada e a sua substituição é incluída no plano de manutenção, de acordo com a estimativa do tempo de vida útil do equipamento.

Apesar de muitas outras técnicas estarem disponíveis, o recurso à análise por vibração ainda é a principal técnica para a aplicação de monitorização de condição, devido à sua robustez e precisão para identificação de falhas. A ferramenta aqui desenvolvida aplica uma série de diferentes técnicas de processamento de sinal para que este seja melhorado: re-amostragem para o domínio angular, procedimento de edição cepstral, correlação espectral cíclica e análise de envelope com filtragem, acabando com o cálculo de indicadores ou extração de indicadores.

Para além disso, este documento sugere um método para processamento semi-automático de sinal e extração de recursos, envolvendo filtragem.

Adicionalmente, este documento cobre vários artigos sobre o tema *condition monitoring* (monitorização de condição), o processamento de sinais de vibração e finalmente a seleção e classificação dos indicadores.

O objetivo final é, juntamente com o método de processamento de sinal, desenvolver uma ferramenta para classificação de falha através de librarias disponíveis em *Python*, librarias essas que podem classificar classes com boa precisão.

Finalmente, alguns resultados são apresentados onde se podem extrapolar algumas conclusões sobre a eficiência de cada um dos processos usados neste método.

Keywords

- Condition Monitoring
- Vibration Analysis
- Envelope Analysis
- Cyclic Spectral Correlation
- Feature Extraction
- Feature Reduction
- Machine Learning
- Classification Methods
- Support Vector Machines
- Decision Trees
- Random Forest

Contents

List of Figures

List of Tables

List of Acronyms

CM Condition Monitoring

PM Preventive Maintenance

CBM Condition - based Maintenance

CBMS Condition - based Maintenance Strategy

CMS Condition Monitoring Systems

REB Rolling Element Bearings

FDD Fault Detection and Diagnosis

RMS Root Mean Square

BPFO Ball Pass Frequency of Outer Race

BPFI Ball Pass Frequency of Inner Race

BSF Ball Spin Frequency

PCA Principal Component Analysis

LDA Linear Discriminant Analysis

ICA Independent Component Analysis

RUL Remaining Useful Life

ML Machine Learning

SVM Support Vector Machines

LCF Low Cycle Fatigue

HCF High Cycle Fatigue

VHCF Very High Cycle Fatigue

MEMS Micro Electro-Mechanical Systems

AC Alternated Current

DC Direct Current

AE Acoustic Emission

COT Computer Order Tracking

CEP Cepstral Editing Procedure

FT Fourier Transform

FFT Fast Fourier Transform

CSC Cyclic Spectral Correlation

PSE Power Spectrum estimation

DSP Digital Signal Processing

PSD Power Spectrum Density

HT Hilbert Transform

P2P Peak-to-Peak

CF Crest Factor

$\boldsymbol{Went_{Log}}$ Logarithmic Entropy

$\boldsymbol{Went_{Sha}}$ Shannon Entropy

IF Impulse Factor

MF Margin Factor

SF Shape Factor

FM4 Fourth Order Figure of Merit

mRMR Minimal-Redundancy-Maximal-Relevance

PSVM Proximal Support Vector Machines

ANN Artificial Neural Networks

CNN Convolutional Neural Networks

DT Decision Tree

SRM Structural Risk Minimisation

ERM Empirical Risk Minimisation

NIPS Neural Information Processing System

SQ SpectraQuest

FC Central Frequency

Mic Microphone Sensor

$\boldsymbol{BR_x}$ Accelerometer: X direction

$\boldsymbol{BR_y}$ Accelerometer: Y direction

$\boldsymbol{BR_z}$ Accelerometer: Z direction

LR Load Right

O&M Operations and Maintenance

DL Deep Learning

NN Neural Networks

DFT Discrete Fourier Transform

LM Load Middle

PC Principal Components

CHAPTER 1

Introduction

1.1 Condition Monitoring and its importance for bearing analysis

Every mechanical equipment will develop a fault during its useful working life. Whether it is a light fault like a component that needs reconditioning or lubricant replacement or a more severe fault like one or more components that need replacement altogether, faults will eventually happen. The severity of the fault will depend on a wide range of factors. To tackle this, condition monitoring has been conducted through every equipment even if it is on a unconscious level: whether it is through noise analysis, lubricant check when it is being changed or other more primitive measures, this type of maintenance strategies have been performed since the beginning of Mechanical Engineering.

The field of Condition Monitoring (CM) is getting more and more attention because its methods are becoming recognised as the most efficient strategies for carrying out maintenance [1].

However, before going deeper into the subject of CM, it is important to have a clear perception of the meaning of the term CM. Even though each author defines it differently in the literature, one can define CM as:

> *"Condition Monitoring is a monitoring process or a sensitive tool that focuses on early detection of faults, failures and wear of machinery with the intention to minimize downtimes and maintenance costs, and consequently, maximize production."* [2]

Additionally, according to [3], CM can also be defined, in a more general way, as a process or tool that integrates technologies, specialised people, condition indicators and quality data/measurements to estimate the health condition of a machine and to make the best decisions about maintenance actions.

One sees an increasing trend in today's market: the increasing competition between industries to keep launching products that are both effective and reliable while meeting each market's regulations. To match this severe internal competition, industries have no other option than to demand engineers and designers to produce machines and processes able to withstand brutal work conditions, so that, sometimes

machines have operating cycles of weeks, or even months only stopping for programmed events [4].

The evolution of industry equipment requires better and better maintenance strategies. Maintenance work should not be performed without consideration but it should also be kept in mind that a machine or equipment cannot run without an effective maintenance strategy.

Maintenance through CM can be performed through a wide range of techniques from chemical effects analysis, wear debris analysis via oil analysis through thermography or vibration analysis. The latter is the most commonly technique used when it comes to applying a Condition - based Maintenance Strategy (CBMS) for its signal processing techniques have been studied for more than thirty years. Nowadays other techniques are being studied such as analysis using Acoustic Emission sensors, or Microphones. However vibration analysis is still applied in conjunction to these techniques so that the conclusions can be cross-checked and its information can be used to help when it comes to make decisions [5].

The dynamic nature of machines requires correct maintenance for them to perform reliably at peak performance [6].

Machines were originally operated until failure (Run-to-Break), prioritising maximum operating times between breakdowns. However, this meant that when failure does occur can be catastrophic and result in severe consequential damage. This is undesirable, especially when such failure can be prevented by experienced technicians, sometimes with decades of experience, who are able to detect the state of a machine through sound, touch or other types of inspection. The Run-to-Break strategy still has place in industries where the production losses are not high and the machine is not essential for production, for example sewing machines in the textile industry.

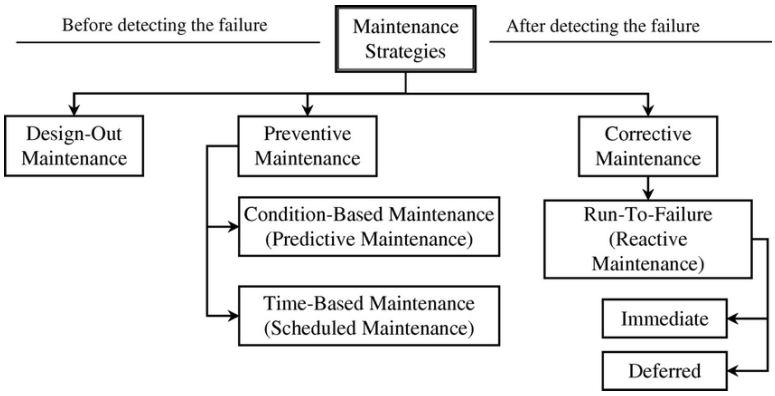


Figure 1.1: Different maintenance strategies [7].

The following maintenance strategy is the Preventive Maintenance (PM). This maintenance strategy tries to take into account this experience and applies it to routine minor checks that greatly influence machines' behaviour. This routine minor checks must be done at regular intervals considered shorter than the expected time between failures. This maintenance strategy usually assures that 1-2% of machines will experience failure during that time [1].

The drawbacks of PM are that it doesn't take into account the fact that different components wear at different rates. As such, some components are replaced that could have run for more time while others are replaced at a tardier stage which can lead to failures that otherwise would have never occurred. This method is better suited for applications where the components wear at a predictable and constant rate that can be calculated if the operational regime is steady [1]. On the other hand, critical components such as bearings or gears must not be maintained with a PM strategy for these equipments do not wear at a constant and predictable rate.

Typically, different bearing faults develop at rates. Additionally, fault evolvement also depends on operating conditions and the fault type itself. The time between the development of a potential failure and a serious and catastrophic can be long. This allows the operator responsible for maintenance to establish a maintenance schedule that prevents catastrophic and suddenly failures.

However, these faults may go undetected at early stages [5].

According to Randall in [1], the steps for a maintenance strategy are the same for all the other applications and can be resumed in three steps:

1. Data Collection
2. Processing and Analysis
3. Diagnosis and Prognosis

Condition - based Maintenance: Advantages and Economic Benefits

Some companies still choose to see the *glass half-empty* by regarding a maintenance centre as a cost centre. However, as demonstrated by [8] there is a considerable amount of evidence that adopting a Condition - based Maintenance (CBM) or CBMS results in economic advantages in most industries. Some authors go even further by stating that a CBM or CBMS can convert maintenance to a profit centre [9, 10].

When it comes to wind power, its maintenance, especially when offshore, is known to be difficult and expensive. The reasons behind this problem are related to unexpected failures, spare part and equipment availability, and weather conditions that may lead to long down times. As demonstrated by [11], using Condition Monitoring Systems (CMS) results in not only economic benefits but also lower risks of consequential damage at failure while providing advantages for the planning of the maintenance. As stated in [11] using CMS returned an economic benefit of 190 000,00 € and reduced risk of high cost of failure.

Y. Wang studied in [12] the cost benefit of using CMS in correct maintenance decision making by analysing the interrelationship between CMS effectiveness and system downtime due to system failures. It was proven that adopting an optimal maintenance strategy can minimize Operations and Maintenance (O&M) costs.

Another topic that must be taken into account is the indirect cost of a machine that is not operating at its best performance. When an equipment is running with a failure that just started to develop, the running conditions will be different and the machine will not only wear other components but also use more energy than it should.

In essence it will not be efficient. This is important because nowadays engineers should thrive for a more sustainable world by creating machines that are reliable while using less and less resources.

1.2 Data Pre-Processing

Amongst other ways like acoustic emission analysis or oil debris analysis, vibration analysis has been extensively used in bearing diagnostics of rotating machinery and extensively studied for the past 40 years as proven by the papers using techniques found all the way back to the 1970s [13, 14].

Acquiring vibration signals involves the installation of a transducer, normally a linear accelerometer, as close as possible to the bearing that is being monitored so that the clearest signal can be obtained. The fault signal results from the vibration generated when the fault - either on the inner race, outer race or on the rolling elements - interacts with the other rolling surfaces, an aspect that will be covered further on this document [14].

However, when choosing methods for condition monitoring one must recognise that using tools such as vibration analysis (or any other form of analysis for that matter) result in data for the assessment of the bearing condition that are in raw state. For effective diagnosis and prognosis of bearing element bearings this data must be precisely treated in order to remove chances of the data being perturbed by surrounding equipments or other forms of signal contamination [15]. For instance, implementing vibration analysis for inferring the health condition of REB yields the information, to be treated further, in the form of signals that are often perturbed by surrounding noises - coming from other equipments - or interferences coming from mechanisms, present in the equipment itself, like gears or shafts rotating. As stated by Randall in [13], a major reason that explains why the relative weak bearing signals are masked - more obvious for very low rotational speeds - is the discrete frequency noise component coming from gears. Even in machines where gearboxes are not present, a very strong discrete frequency will exist and will contaminate the frequency bands where the bearing signal is usually dominant. On top of this, the signal coming from the source, the bearing, can be masked in a way that it loses its impulsiveness. This means that an accelerometer, or a microphone - if sound analysis is performed - is only mounted as close as possible to the component being tested. For the case of bearings - whether it is a faulty bearing or a healthy one - the signal is generated on the bearing itself and has to go through the structure from where the bearing is mounted to where the transducer is located. This is more relevant when studying the fault signal, what happens is that the fault pulses are modified when passing through a transmission path where the impulse response is long [13].

Another aspects to take into account are the speed fluctuations of not only the bearing, but also of the rolling element itself. This fluctuation in rotation speed - for the bearing - or when the ball or cylinder is slipping instead of spinning - for the REB - results in a considerable change of character of the measured signal [13, 15, 14].

In order to overcome these obstacles, signal processing has become a very important technique in order to obtain diagnostic information about bearings.

Different signal processing techniques have been proposed such as angular resampling to remove the speed fluctuations, Cyclostationary methods to determine the resonance frequencies at which the faults occur, Cepstrum analysis to remove the non-periodic components, [16] Hilbert transformation for demodulation of the signal, and many others as Akhand Rai et al. studied in [15]. Envelope analysis has also proved to be a very powerful method of analysing vibrational data since the analogue days. Today's capabilities of computer processing have shed new light on this already powerful method [13, 1].

After placing the correct transducers in place and acquiring the vibration signal of the bearing being studied, envelope analysis along with filtering is usually applied to the signal although different authors complement envelope analysis with different signal processing techniques [15].

1.3 Feature Extraction, Reduction and Selection

Feature Extraction

Once the stage of signal processing is performed, anomalies on REB are detected using features, or indicators, extracted from the vibrational signal and will give an idea of the condition of the bearing [16]. Fault Detection and Diagnosis (FDD) is easily done through the analysis of simple statistical features that are calculated on the time domain or on the frequency domain. Features like kurtosis or Root Mean Square (RMS) - among others constitute the features in time domain - or the correspondent amplitude values on the spectrum on the frequency areas of Ball Pass Frequency of Outer Race (BPFO), Ball Pass Frequency of Inner Race (BPFI), Ball Spin Frequency (BSF), when an outer, inner or ball fault develops, respectively - constitute the features on the frequency domain [17, 18, 19].

These features, or commonly known as Condition Indicators [6], are calculated after the signal is processed, which means that, if for envelope analysis the choice of the frequency envelope window (or central frequency) and window bandwidth (or frequency range) is important [20], it will be even more important because, as a consequence, the features can render as inconclusive. This means that the parameters for envelope analysis, or signal processing, must be chosen carefully in order to obtain features that might lead to a good characterisation of the condition of the REB.

Feature Reduction and Selection

Open literature gives plenty of evidence on the effectiveness of FDD of REB. Often, FDD of REB can be performed through one of two methods:

1. **Advanced signal processing techniques**: As stated above, signal processing techniques can be used in order to obtain information regarding the condition of a REB. Even though these tools can be used as a decision when it comes to diagnosis assessment, they do not give automatic results on the condition of the component, for these methods do not classify the fault nor recognise a pattern

when present. On top of this, analysis of vibrational signals and diagnosis decision making must be done by an expert, this means that an automatic diagnosis method cannot be implemented [21].

2. **Artificial intelligence tools and pattern recognition methods - Machine Learning**: Even though these methods are somewhat complex, they can be used to develop automatic bearing fault diagnosis systems, without human intervention. These methods can also be extended to be used to predict the Remaining useful life of a REB, through pattern recognition [22].

However, when calculating features on a signal - and specially when one doesn't know which feature correlates to a better indication of a fault of a component - it is good practice to calculate a high number of features, or indicators, so that the automatic classifier has a chance of finding the REB's condition - healthy or not.

For this purpose, various techniques have been proposed on the literature. Techniques such as PCA, Independent Component Analysis (ICA) and Linear Discriminant Analysis LDA are the most commonly used [22, 23, 24, 25, 21, 26], each with their own advantages and disadvantages. Jaouher Ben Ali et al. demonstrated in [22] that using feature reduction techniques improves the quality of the classifiers used. However, the ability to interpret the influence of individual features/variables decreases when feature reduction methods are applied. These techniques *save* feature's information in a set of components, and these components cannot be interpreted directly except for the amount of information kept compared to the original data - in case of PCA - or the information regarding the variance between the features and the components - in case of LDA. In addition, it should be taken into account that feature reduction comes at a cost: the loss of information compared to the original data. Using the PCA as an example, each of the resulting components represent a certain amount of information compared to the original situation. If, after applying PCA and using three components, these three components only amount to represent 70% of the original set of data, then one should think of using more components.

Another important step - parallel to feature reduction - to take into consideration is the step of feature selection. Effective FDD of REBs is accomplished by using the most meaningful and most responsive features to the fault presence and severity. It can be easily understood that using improper and inaccurate features reduces the overall accuracy and reliability of a possible health indicator method for REBs, moreover such method might be unable to predict the actual bearing condition. Removing irrelevant, redundant or noisy features also leads to better learning performance - higher learning accuracy - lower computational cost and a better interpretation of the model used [27, 28]. This step, just like the feature reduction step, becomes more important in problems with a high number of features.

This subject has received a lot of attention during the last years since there is a need for methods that are computationally efficient and accurate, yet sensitive to complex patterns of association so that features that hold good information are not mistakenly removed [29].

The supervised methods take into account the label to look for relationships between classes and the features and selects relevant features to distinguish between

classes. The unsupervised feature selection is considerably more complex once there is no knowledge regarding the labels. The semi-supervised methods create a similarity matrix and select features that best fit that matrix [28].

In addition, methods for feature selection can also be classified in terms of search strategies: Filter, Wrapper or Embedded methods. The filter methods distinguish the features via characterisation of data. Wrapper methods rely on machine learning algorithms to evaluate features, while the embedded methods construct a model of the data and select the features based on that model [28].

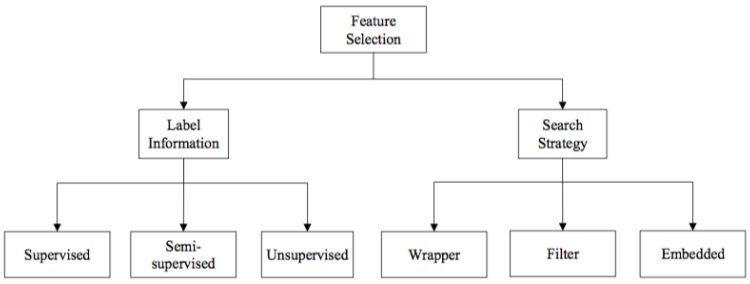


Figure 1.2: Feature Selection Classification [28].

1.4 Classification Methods

The final step of an *automated* method of FDD of REBs is the classification of the fault that developed on the bearing and, on a later stage, the recognition of a pattern for assessment of the fault severity and determination of the bearing's Remaining Useful Life (RUL).

It is already known that vibration analysis along with effective signal processing techniques allows one to extract quantitative information of the bearing vibrational signal characteristics. However, using this information - the features - to diagnose a fault that has developed on the REB presents a considerable challenge, for solving this requires the interpretation, classification and pattern recognition of the extracted data. In order to solve this, various Machine Learning (ML) algorithms have been successfully applied in the fields of fault classification, fault detection, condition monitoring and prognostics [30].

Today's processing power has enabled ML methods to be accurate and reliable and to be used in all fields from cancer research and marketing and sales applications to criminal justice [31, 32, 33].

Despite the existing efforts of explaining the *internals and workings* of ML methods, these and sometimes their output remain esoteric to the everyday user, often understood only by experts with years of training and development experience.

Given the fact that these methods' popularity is growing progressively and its use is growing considerably in the real world, understanding how these methods perform their intelligent decision-making has become increasingly important and critical [33, 34].

These methods are classified into three families, in accordance to the availability of label information: Supervised, methods that use labeled data, Semi-Supervised or

Unsupervised, where there is no need for labeled data. This is, if the classes of the data one has acquired - in this case, faulty or not - are available then it is possible to use between one of the three families (see figure 1.2) explained [28].

Figure 1.3: Example of a black box system [35].

On an industrial facility, this is no different. A maintenance engineer will most certainly feel disbelief if the tool he is using tells him that a component is faulty, without showing any evidence on why the tool classified the component as faulty. Engineers' attitude consists of questioning everything, hence the importance of relying on methods that are both easy to interpret and accurate.

Currently, ML methods are available through both comercial and open source packages, both offering optimized algorithms of numerous different methods [30]. One of the most relevant and used open source tools for ML is the scikit-learn package available on the Python scripting language [30, 36]. As already stated, this project tried to avoid methods that were complex to interpret. As such, a simple method of class classification, the Decision Tree, was used and directly compared to two more complex methods: SVM and Random Forest. Class, or fault, classification was accomplished with considerable accuracy using the Decision Tree method, which after running allows the user to see the model created and understand how class classification was performed.

1.5 Motivations

REBs are one of the most used components in machinery construction. Failure of REBs cause downtimes for component replacement or refurbishment which have consequences on the economic viability of large systems and chain operated equipments.

Component failures, however, can never be avoided, only postponed. Components will wear out, oil will loose its lubricant capability and one can only make plans for programmed maintenance. Here is where condition monitoring steps in, the constant supervision of a given component and, given a threshold, the warning of a developing failure and its consequent maintenance.Condition monitoring of REBs has been studied for many years having a considerable body of evidence that returns economic benefits, such study has resulted in a numerous signal processing techniques and approaches to this issue.

However, condition monitoring through vibration analysis can be an exhausting task performed only by an expert on vibrations, which doesn't allow the development of an automatic method of fault classification of malfunctioning REBs without user input. On the other hand, such tool cannot be created and used without proper interpretation and critic thinking. This text will present a semi-automated, fully interpretation-able, method for condition monitoring of REB for classifying faults.

1.6 Objectives

- Understand the fundamentals of condition monitoring, the different techniques used, and the state of the art. Clarify the use of vibration analysis in this area and where it fits with other techniques.
- Develop a semi-automatic method of frequency band selection for filtering for envelope analysis.
- Study the results differences between filtered signal and non-filtered signal.
- Different faults exist at different resonance frequencies that are dependent on speeds, as such, find a connection between frequency bands and, if possible, develop a frequency band for filtering that works for all situations.
- Extract features from the signal on the angle domain. Find relationships between features, study the methods available for relevant feature selection and apply the best suiting one.
- Apply a classification method in order to successfully classify a fault that might be present on a bearing being studied.

1.7 Structure

This document is structured into seven main chapters. The following chapter explains the vibration concepts of REBs, the different faults that might develop and the technologies and analysis that can be used to perform condition monitoring of said equipment. The third chapter explains the signal processing techniques that were used for signal enhancement and treatment. After this, the fourth chapter covers features and their analysis: which features were used and how they where calculated followed by an explanation of feature reduction methods and finishing on feature selection methods for redundant and *irrelevant* features elimination.

The fifth chapter clarifies the theory behind the classification methods used for fault classification. The sixth chapter shows the results obtained after applying signal processing techniques, feature reduction, feature selection and feature classification. The document ends with a final chapter stating the conclusions and future work on this topic.

CHAPTER 2

Rolling Element Bearings and Technologies

2.1 Bearing Defects Fundamentals

Most of the industries rely on equipments that must go through different motion conversions. When it comes to rotating motion, a key component is the REB whose main function is to support the rotating element while allowing it to rotate. This rotating element can either be a simple shaft or a shaft with a set of gears attached.

The precision rolling-element bearing of the twentieth century is a product of exacting technology and sophisticated science. Simple in form and concept and yet very effective in reducing friction and wear in a wide range of machinery. Normally this is the most precise component parts and are typically fabricated with tolerances that are around ten times more strict than the other machine components [5]. This equipment evolved immensely since its development and nowadays the project engineer developing a machine or component has a wide range of bearings to choose from.

The most common and most used types of bearings are the REBs and the journal bearings. The journal bearings work on principles that are somewhat different from the REBs. These work on the principle of hydrodynamic lubrication. In order to do so, a small *gap* - or tolerance - between the journal and the bearing is needed. Usually the rotating element - or journal - is supported on an outer bearing, usually made of bronze, for auto-lubricating purposes. Then, the gap between the bearing and the journal is filled with a lubricant characterized by a certain viscosity and while running at high speeds, a pressure is built up toward the centre of the shaft. This fluid pressure, is responsible for supporting the load on the shaft and helps reducing the friction, and consequently the wear. During the starting and stopping phases, usually when the shaft rotates at slower speeds, the shaft and the bearing are subject to metal-to-metal contact, which can lead to degradation and wear of both parts. Normally, the bearing is either coated or manufactured in a material that is softer than the shaft so that it degrades faster than the shaft. At a point, this element needs replacement and a solution to this problem are the hydrostatic bearings that have an external source of pressurised lubricant that avoids the wear between the journal and the bearing. This last type of bearings is used when the loads are very high. Another characteristic of journal bearings is that its damping ratio is controlled by the

lubricant. Consequently if the vibration response needs to be changed than one only needs to replace the lubricant used [5].

This text however will not cover this kind of bearings. The ones being studied on this document are the REBs, that work thanks to the rotating motion of a particular element between two circular races, the inner and outer. In order for this to work frictionless, the bearing is filled with a lubricant: either oil or grease or a mix of the two. To keep the rolling elements from hitting each other or getting out of their place is the cage, where the rolling elements are put into. The sealing has the objective of securing the lubricant inside and keeping external particles from getting into the tracks - and consequently developing a serious failure on either one of the races or on the surface of the rolling element.

These four elements are the constituents of a typical REB: the tracks or races, the rolling elements themselves, the cage and the sealing. Although different manufacturers have different design variations, this kind of bearing anatomy is the most common and simple. Figure 2.1 gives an illustration to the assembly of a REB.

Figure 2.1: Anatomy of a REB. Adapted from [37].

"A chain is only as strong as its weakest link."

The proverb can, of course, also be applied to industrial machinery, for an equipment is only as robust as its weakest component. It is easy to see why such simple component plays a vital role inside any mechanical equipment. This item, when faulty, causes shutdown of the entire equipment and, at times, the entire production chain if the next equipment depends on the preceding one. This of course results not only in maintenance costs but also in economic losses given that the production line will be on hold until that specific component is replaced.

As already stated, the main purpose of a REB is to provide relative positioning and rotational freedom while distributing a load between two structures: the shaft and the housing. If there is a need of distributing loads between surfaces, such action can be facilitated through the use of REBs between the sliding members. The friction encountered during the rotating motion is also reduced thanks to the use of lubricated rotating elements, even though such elements - rollers, tapered rollers or spherical balls - are afflicted with high stresses due to effective load transmission [38].

The rolling elements can have different geometries and, as already stated, are manufactured under very tight tolerances. However, no manufacturing process is 100%

perfect which means that some rolling elements are not perfectly spheric or cylindrical. Mohanty states that the best bearings available in the market are fabricated with radial imperfection of 3 to 5μm [39]. Due to this slight variation in geometry, when the bearing is in motion, a particular vibration signature is generated. Moreover, thanks to the same geometric variation of the rolling element, the radial load on the bearing is constantly changing with the shaft's rotation, which makes the signal amplitude modulated.

It is easy to understand that if there is a slight variation on the geometry of either the races or the rolling element, the vibration signature of the bearing will change. Let's assume that because of a faulty sealing, a grain of sand managed to get into the bearing. As the rolling element strikes this grain of sand, an impulse is generated which excites high frequency resonances between the bearing and the transducer, typically an accelerometer. With time, this grain of sand will start to develop other faults, either on the outer or inner races, by eroding the races' surface, creating a hole, or on the rolling elements' surface itself. This is merely an example, other situations may occur due to metal shavings or chips, etc.

Randall explained in his book [1] that the broadband bursts excited by the impulses are further modulated in amplitude by two factors:

- The strength of the bursts depends on the load applied to the rolling elements, and this is usually modulated by the rate at which the fault passes by the load zone;
- If the fault is moving - a rolling element fault, for example - the transfer function of the transmission path changes with respect to the fixed positions of the response transducers.

Given that the signal measured by the transducer will be modulated, it is important to demodulate the signals and analyse them on the frequency domain in order to perform diagnostic analysis. Figure 2.2 shows the envelope signals of the faulty signals generated by the different localised faults in the different components.

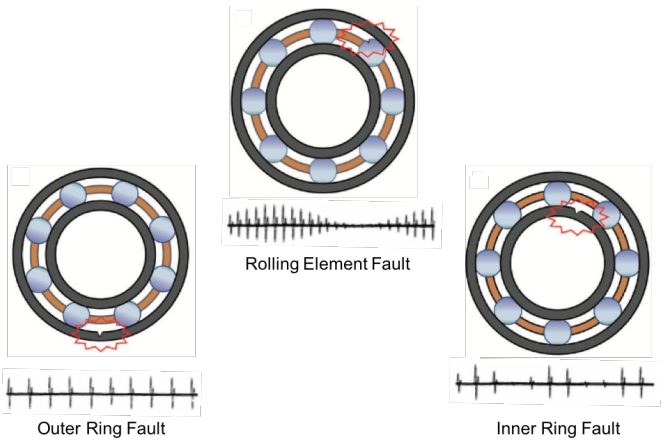


Figure 2.2: Envelope signals generated on a faulty bearing, adapted from [27, 5].

2.1.1 Fault Characteristic Frequencies

As already stated, for a particular bearing geometry, faults present on inner race, outer race or rolling element generate a vibration spectra different from the bearing without faults and with unique frequency components [30, 40, 41]. It is thanks to these specific frequencies and their magnitudes that one is able to analyse the vibration spectra obtained from the bearing and determine the condition of the bearing. These frequencies are not constant however, they are directly related to the shaft speed the bearing is attached to. In addition, Outer Race and Inner Race fault frequencies are also dependent on the number of balls present in the bearing.

Using the geometry clarified in figure 2.3, one can write the following equations

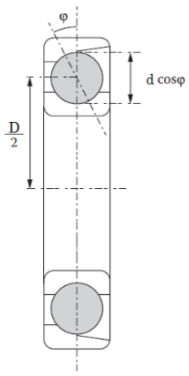


Figure 2.3: Ball bearing geometry [39].

describing the frequencies at which the different bearing faults occur:

$$BPFO = \frac{f_S \cdot N}{2}\left(1 - \frac{d\cos\phi}{D}\right) \tag{2.1}$$

$$BPFI = \frac{f_S \cdot N}{2}\left(1 + \frac{d\cos\phi}{D}\right) \tag{2.2}$$

$$BSF = \frac{f_S \cdot D}{2d}\left[1 - \left(\frac{d\cos\phi}{D}\right)^2\right] \tag{2.3}$$

$$FTF = \frac{f_S}{2}\left(1 - \frac{d\cos\phi}{D}\right) \tag{2.4}$$

Where f_S is the shaft speed in rpm and N is the number of balls in the bearing. One important aspect to be taken into account is that the BSF is the frequency with which the ball fault strikes the race, inner or outer and accordingly there are normally two shocks per ball rotation period. Consequently the even harmonics of BSF are often dominant, particularly in the envelope spectra [1].

However, these kinematic frequencies are described without taking into account the ball slip. In fact, slip must virtually occur because, given that the balls are not perfectly spherical the angle ϕ changes, depending on the position of each ball in the bearing as the ratio of local radial to axial load changes. Thus, each ball has a

different diameter and is trying to roll at a different speed. The cage as the function of ensuring a mean speed that is equal to all the balls, by causing a random amount of slip, varying between 1-2% both as a deviation from the calculated value and as a random variation around the mean frequency [1].

However small this random slip may be, it changes the character of the signal considerably and is the reason why envelope analysis extracts diagnostic information not available from frequency analyses of the raw signal. It also means that bearing signals can be considered as cyclostationary, an important aspect to be studied in chapter 3.

Additionally, the number of rolling elements and their position in the load zone change with bearing rotation, giving rise to a periodical variation of the total stiffness of the bearing assembly. This change in the total stiffness generates vibrations commonly known as varying compliance vibrations. When the bearing races are assumed as continuous systems, the changing direction of the contact forces applied by the rolling elements may cause flexural or ring-mode vibration of the races even if they are geometrically perfect [42].

As seen previously, it was shown that bearing defects cause a significant change in the vibration signature of the REB. The defects present on bearings can be categorised as distributed or local defects.

Examples of distributed defects are surface roughness, waviness, misaligned races and off-size rolling elements which are caused by manufacturing error, improper installation or abrasive wear [42].

Localized defects can be cracks, pits and spalls on the rolling surfaces. The most common cause of failure of rolling element bearing is the spalling of the races or of the rolling elements, and is caused by a fatigue crack under the surface of the metal. Fatigue failure can be caused by overloading or shock loading of the bearing during installation or running. In [42] other defects were found, such as electric pitting or cracks thanks to excessive shock loading.

Whichever fault may be, one common fact among all of them is that whenever an interaction between elements with a defect, abrupt changes in the contact stresses occur which generates a pulse of high amplitude and very short duration. Such impulse produces vibration and noise that are used to monitor and detect the presence of a defect's initial stage in the bearing [42, 5].

2.1.2 Distributed Defects

Surface Roughness

Even though engineers try to optimize manufacturing processes, there is no process with 100% of accuracy. This is not only important in terms of dimensions but in terms of surfaces and geometry. There are no perfectly flat surfaces or perfectly spherical objects and no surfaces are perfectly parallel among them. Real objects will have form errors, surface waviness and surface roughness due either to the manufacturing processes or created by previous operation.

As one can visualize in figure 2.4, for the real contact area, there is less contact between surfaces. Additionally, the border of the real contact area may be wider than

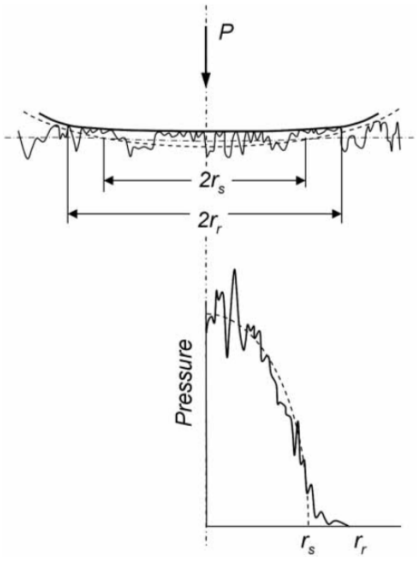


Figure 2.4: Differences in contact pressure distribution between real and ideal surfaces [43, 5].

that of the ideal surface [43]. In addition, figure 2.4 shows that the real surface contact leads to higher pressures.

Surface roughness is one of the parameters set by project engineers on their different projects. This information is in the technical drawing of a certain equipment and usually there are different levels of surface roughness requirement for a certain project. In the assembly regions, where different components are in direct contact, there will be a higher requirement for lower surface roughness for this exact reason.

Surface roughness exists because the material will not be completely *flat* in its entirety. One way of assessing the roughness of a certain surface is through the arithmetic average R_a which is measured through the performing and arithmetic average of a number of measurements in the direction of the normal vector of a real surface (see equation 2.5).

$$R_a = \frac{1}{n} \sum_{i=1}^{n} |y_i| \tag{2.5}$$

Film Thickness

Between other factors, surface wear will be largely influenced not only by the lubricant used but also by *how much* lubricant is used. If one pictures a gap on a rolling contact, there will most certainly be some kind of liquid inside the gap, trapped between the surfaces. The thickness of the liquid in this gap is important. The liquid being caught in that gap, when the surfaces are rolling or sliding, will create a lubricant film that aims to separate the contact surfaces and reduce the contact pressure [43].

The lubrication mechanism and its effect is chosen based on the value of λ, that refers to the ratio between the film thickness h and the combined surface roughness in

the contact surface R_a [43, 5].

$$\lambda = \frac{h}{R_A} \tag{2.6}$$

$$R_A = \frac{1}{2}\left(R_{a_1} + R_{a_2}\right) \tag{2.7}$$

Based on the value of λ, the lubrication regimes can be divided into boundary lubrication, mixed lubrication, elasto-hydrodynamic lubrication and hydrodynamic lubrication [43, 5].

The first regime, boundary lubrication, characterized by $\lambda < 1$, occurs when operating conditions' hydrodynamic action is too weak, or insufficient, for separating the contact surfaces. One example of such situation is low-speed applications where the surfaces in contact are merely wetted by the lubricant.

Moreover, the load is carried mainly by solid-to-solid contacts and this regime of lubrication relies on lubricant and surface properties to generate boundary films when running [43, 5]. As a consequence to this, some wear is expected due to tribochemical or mechanical wear [43]. In the case of REBs this kind of lubrication can occur locally in micro-slip zones and at the end planes of axially loaded rollers. One should note that boundary lubricated contacts lack any damping effect, and boundary lubricated contacts can act as a vibration sources [43, 5]. The following regime, the mixed lubrication with $1 < \lambda < 3$, is characterized by boundary lubricated conditions similar to elastohydrodynamic or hydrodynamic lubrication. In this regime, lubrication happens by either a solid-like or viscous-like boundary film [43, 5].

The last two regimes, elastohydrodynamic $3 < \lambda < 10$, and hydrodynamic, $\lambda > 10$, are based on Reynold's equation for hydrodynamic oil pressure build-up and a combination of elastic deformation at the contact surfaces of low geometric conformity. In the elastohydrodynamic regime the film thickness is slightly higher than the combined surface roughness, while in hydrodynamic regime the film thickness is much higher than the referred roughness. In rolling element bearings, hydrodynamic lubrication occurs in contacts between rolling elements and their cages and, also between roller ends and bearing race flanges for roller bearings with axially loaded rollers. The higher film thickness achieved in this regime provides a significant degree of damping for dynamic loads [43].

Fatigue Wear

Fatigue wear follows the same originating principles of fatigue cracks. On an airplane, for example, several components such as the wings, the rotor blades or even the fuselage, are subject to fatigue which has caused plenty of headaches on aviation engineers. The cyclical application of pressure on the metal of a given component eventually causes it to develop a crack. A REB is no exception, fatigue cracks on bearings can be due to a wide range of causes like, the dynamic working conditions on the bearing, the shear stress, plastic deformations or contaminating particles present in the lubricant, sometimes external particles that somehow managed to go over the sealing and made their way into the REB itself. A fatigue crack begins below the surface of the metal and propagates towards the surface until a piece of metal breaks away to leave a small pit or spall. A fatigue failure can be expedited by overloading or

shock loading of the bearings during running and installation [43].

According to [44] structural fatigue of machines can be categorized in three regimes:

- Low Cycle Fatigue (LCF) refers to the fatigue regime in which the number of cycles to failure, N_f, is less than 10^5 cycles, and the failure is initiated from the surface slip bands.
- High Cycle Fatigue (HCF) corresponds to a N_f that varies from 10^5 to 10^7 cycles. In this regime, fatigue failure can be attributed to both surface slip bands and subsurface inclusions.
- Very High Cycle Fatigue (VHCF) corresponds to a number of cycles to failure of 10^7 or more. Common to high strength steels this type of fatigue failure is mostly due to material in-homogeneities such as non-metallic inclusions.

Gabelli states in [44] that usually, rolling bearings have a very high number of stress cycles to failure - no less than 10^7 and typically close to 10^9 - and are induced by the over-rolling of the rolling elements on the bearing races. Most bearings usages generate typically 2×10^9 to 3×10^9, while heavier applications require bearings able to withstand up to 3×10^{11} stress cycles.

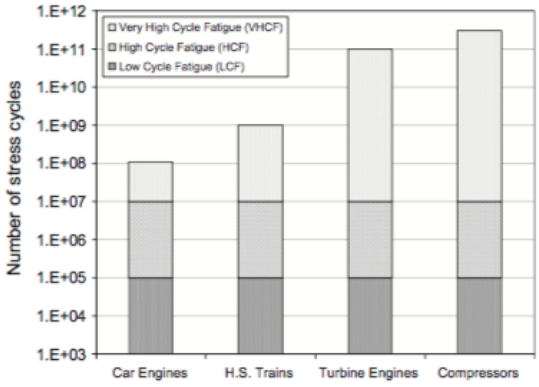


Figure 2.5: Bearing applications and their typical ranges of service lives [44].

As one can observe from figure 2.5 bearings for gas compressors, diesel engines superchargers for instance, involve a very larne number of stress cycles. These machines will usually run 5 years of service life, with 24h of continuous operation at high speeds (10 000 rpm to 30 000 rpm) which results in required stress cycles of $3,4 \times 10^{11}$. However, it should be noted that small or medium size bearings are better suited to withstand very high stress cycles [44]. There are large size bearings that go through longer service life expectancies, in some cases more than 10 years in continuous operations. Although, these develop lower number of stress cycles, meaning that the dynamic loads are not as great as the small to medium size bearings are subjected to.

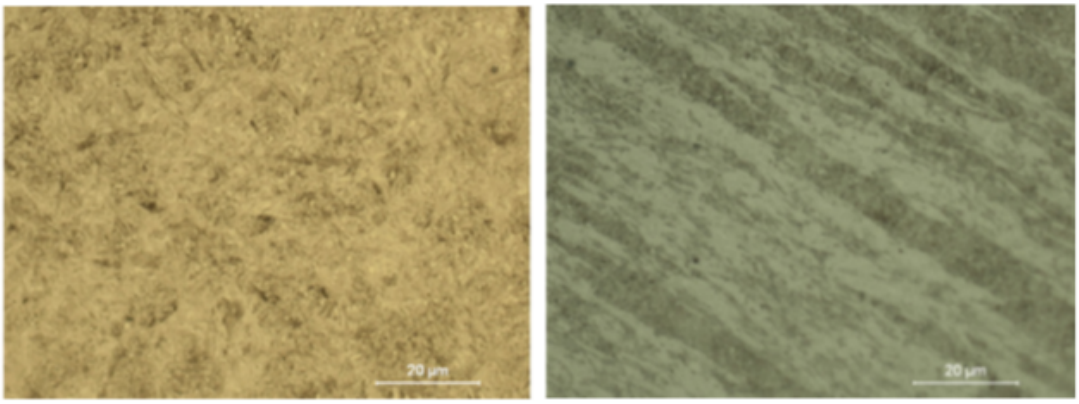
As an example, bearings used in large driving systems of industrial manufacturing plants require continuous operation and can reach typically 12 years of uninterrupted service. However they only withstand $6,6 \times 10^9$ stress cycle during this running period. This lower number of stress cycles is explained by the relative slower speeds at which the large size bearings operate [44].

Typically, the service life of rolling bearings is always in the VHCF regime which goes beyond 10^7 cycles and may even reach 10^{11} cycles in some applications. If the over-rolling contact pressure exceeds the elastic limit of the material, decay of the steel microstructure can develop in the VHCF regime. If the contact pressure is below the elastic limit of the bearing, only localized fatigue damage will develop at stress risers such as pores or inclusions.

The ISO 281:2007, on the other hand, covers the concept of a fatigue limit stress in the assessment of the fatigue life of rolling bearings. In case of a rolling contact this model reads [44]:

$$\ln \frac{1}{S} = A \cdot N^e \cdot \frac{(\tau_0 - \tau_u)^c}{{z_0}^h} \cdot b \cdot z_0 \cdot l \tag{2.8}$$

A common method to specify the general rules of limit stress conditions in rolling bearings is to set a threshold condition to the Hertzian contact pressure resulting from the contact between the rolling elements and the races. For instance, the static load capacity of rolling bearings is defined by the maximum stress level of the Hertzian contact. This simple rule can then be transformed to a static load rating of a given specific bearing geometry. In a similar way, the ISO 281:2007 fatigue limit stress is defined in terms of the corresponding maximum Hertzian pressure (σ_H) of the rolling contact. For modern bearings made of high quality bearing steels and good manufacturing practices, the value of σ_H is standardized to a nominal value of approximately $15000MPa$ [44].



a) b)

Figure 2.6: Outer race microstructure with a) $2,3 \times 10^{10}$ stress cycles and Hertzian pressure of 2,3 GPa and b) $2,3 \times 10^9$ stress cycles and Hertzian pressure 3,3 GPa. Severe microstructure decay is visible indicating that the bearing on the right is damaged [44].

Adhesive and Abrasive Wear

When the bearing's lubrication is poor, REBs may suffer from adhesive wear at roller ends and in micro-slip zones. In these sliding contacts, due to the lack of lubrication strong adhesive junctions between surface asperities may be formed due to frictional heating which can lead to adhesive wear.

On the other hand, solid particles can cause local stress peaks and shorten the life of the bearing. Even if these particles are smaller than the mean film thickness they can cause abrasive wear, acting as micro-abrasive elements. Furthermore, abrasive wear may take place when rough surfaces are in contact, for instance as micro-abrasion between rolling elements and their cages [44].

2.1.3 Localized Defects

Cracks

Cracks may form in bearing rings for various reasons. The most common cause is rough treatment when the bearings are being mounted or dismounted.

Hard blows, applied directly to the ring, may cause fine cracks to develop, resulting in pieces of the ring breaking off when the bearing is put into service. Other causes such as excessive interference, excessive load, shock load or heat generation can also lead to development of cracks.

Smearing, phenomenon of material being transferred from one surface to the other which occurs mainly in situations with poor lubrication, may also produce cracks. Cracks of this type can produce fractures right across the rings [45].

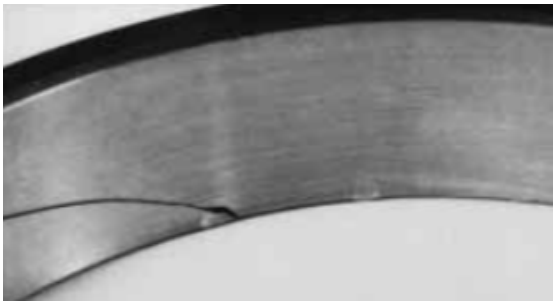
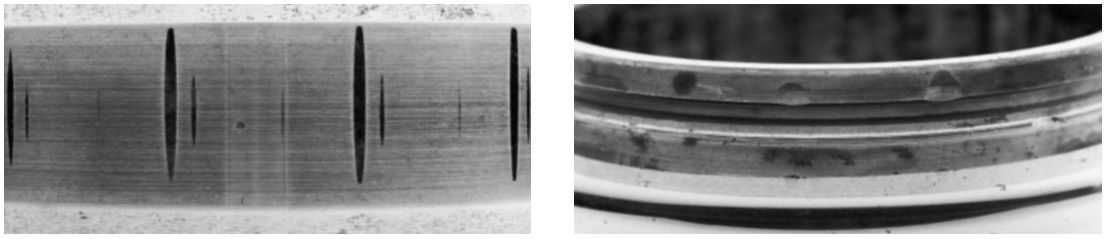


Figure 2.7: Example of a fractured outer ring, on a self-aligning ball bearing [45].

False Brinelling or defects caused by vibration

When a bearing is not running, there is no lubricant film between the rolling elements and the raceways. The absence of lubricant film results in metal to metal contact and if vibrations are present or induced on the main equipment, produce small relative movements of the rolling elements and rings. As a result of these movements, small particles break away from the surfaces and this leads to the formation of depressions in the raceways.

This damage is known as false brinelling, sometimes also referred to as washboarding. Balls produce sphered cavities while rollers produce fluting [45].



a) False brineling on outer raceway. b) False brineling on inner raceway.

Figure 2.8: Example of damage created by vibration, or false brineling [46].

True Brinelling

True brinelling occurs when the loads applied on the bearing exceed the elastic limit of the ring material. Marks resulting from this defect appear as indentations in the raceways which will increase the bearing's vibration and noise when under operation. A more severe brinelling fault will eventually lead to fatigue failure. This type of fault can be caused by severe impact or static overload.

In order to minimize the chances of brinelling, bearings must be assembled and disassembled carefully. In addition, the loads applied on the bearings should be studied thoroughly to avoid overload. Figure 2.9 shows an example of brinelling caused by static overload on the inner race of a rolling element bearing.

Figure 2.9: Brinelling caused by static overload [47].

Corrosion

For every component that is made of metal, water or other corrosive agent, can prove to be a severe problem. Corrosion wear will form if water or corrosive agents reach the inside of the bearing in such quantities that the lubricant cannot provide protection for the steel surfaces. This process will soon lead to deep seated rust. Another type of corrosion is fretting corrosion [45].

a) b)

Figure 2.10: Corrosion defect on a) outer ring of cylindrical roller bearing and b) inner ring of a ball bearing [45].

If the thin oxide film is penetrated, oxidation will proceed deeper into the material. An example of this is the corrosion that occurs when there is relative movement between the bearing ring and shaft, or housing, on account of the fit being too loose. This type of damage is called fretting corrosion. The relative movement may also cause small particles of material to become de break away from the surface. These particles oxidise quickly when exposed to the atmosphere.

As a result of the fretting corrosion, the bearing rings may not be evenly supported and this has a detrimental effect on the load distribution in the bearings. Rusted areas also act as fracture notches.

Pitting & Spalling

As a first note, no common definitions have been established to distinguish spalling from pitting in the literature. In most of the literature, spalling and pitting have been used indiscriminately, and in some other literature, spalling and pitting were used to designate different severities of surface contact fatigue. For instance, Tallian defined in [48] *spalling* as macroscale contact fatigue caused by fatigue crack propagation and reserved *pitting* as surface damage caused by sources other than crack propagation.

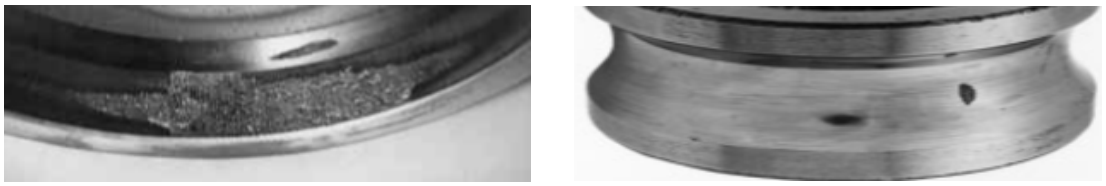
The pitting phenomena is noted when deep craters appear in the surface of the bearing components and are a result of the fatigue cracks that originated in the subsurface. These cracks are a result of the fatigue process and propagate from the subsurface to the surface causing material particles to break away from the surface which will eventually be the cause of abrasive wear. Beside the fatigue process, these cracks can also be caused by moisture in the lubricant as an example of poor lubrication [5].

According to an informative document from the bearing manufacturer SKF [45], spalling occurs as a result of normal fatigue, i.e. the bearing has reached the end of its expected useful life.

If this type of defect is discovered at an early stage, when the damage is not too extensive, it is possible to diagnose its cause and take the requisite action to prevent a recurrence of the issue.

However, when spalling has proceeded to a certain stage, it makes its presence known in the form of noise and vibrations, which serve as a warning that it is time to change the bearing.

The causes of premature spalling may be heavier external loading than anticipated, preloading on account of incorrect fits, oval distortion caused by the shaft or housing being misaligned or axial compression, for instance as a result of thermal expansion.



a) b)

Figure 2.11: Defects detected on a ball bearing: a) spalling on an outer ring [45] and b) pitting on an inner ring [46].

2.2 Sensor Technologies

Based on the idea that a significant change from a reference signal is indicative of a developing failure, condition monitoring systems consist of a combinations of sensors and signal processing equipment and techniques that provide continuous indications of component condition based on analysis techniques such as vibration, acoustics, oil, strain measurement and thermography [49].

García Márquez verified in [49] that using good data acquisition methods and appropriate signal processing, faults can be detected while components are operational, and appropriate actions can be planned in time to prevent damage or severe failure of critical components. In accordance, maintenance tasks can be planned and scheduled more efficiently, resulting in increased reliability, availability, maintainability and safety whilst downtime, maintenance and operational costs are reduced.

This section will be an overview of the techniques used in order to perform condition monitoring on rolling element bearings, the most commonly used being vibration and acoustic emission analyses [30]. Sound analysis is also a powerful technique found on the literature, however not so much explored [50].

Regardless of the technique, the capability of a condition monitoring system relies upon two basic elements: the number and type of sensors, and the associated signal processing and enhancement techniques used to extract important information from the various signals [49].

It is clear that each different condition monitoring methods has their own advantages and disadvantages - Table 2.1 - and should be properly and carefully chosen according to the specific application [50].

Table 2.1: Summary table of different bearing condition monitoring methods [50].

Monitoring Schemes	Major Advantages	Major Disadvantages
Vibration monitoring	Reliable; Standardized (ISO 10816)	Expensive; Intrusive; Subject to sensor failures;
Chemical analysis	Directly monitoring the bearing and its oil	Limited to bearings with closed-loop oil supply system; Specialist knowledge required
Temperature measurement	Standard available in some industries (IEEE 841);	Embedded temperature detector required; Other factors may cause temperature to rise
Acoustic emission	High signal-to-noise ratio	AE sensor required; Specialist knowledge required
Sound measurement	Easy to measure	Background noise must be shielded
Laser displacement	Other way to measure bearing vibration	Laser sensor required; Difficult implementation
Stator current monitoring	Inexpensive; Non-intrusive; Easy to implement	Sometimes low signal-to-noise ratio; Still in development stage

2.2.1 Vibration analysis

Vibration analysis continues to be the most popular technology employed in condition monitoring, especially for rotating equipment. It is advised that different sensors are used for different frequencies: position transducers are used for the low-frequency range, velocity sensors in the middle frequency area, accelerometers in the high frequency range and spectral emitted energy sensors for very high frequencies [49]. According to [51] transducers for vibration condition monitoring can be classified into two main categories:

- Accelerometers - Seismic devices that are normally mounted on the machine structure and whose output is a measure of the absolute vibration of the structure;
- Displacement transducers - Relative displacement transducers that measure the vibratory displacement and the mean position between rotating and non-rotating elements of the machine.

The more commonly used type of transducers to measure dynamic force and vibration are the seismic transducers, which are usually modelled as a mass-spring-damper system, Figure 2.12. The flatness of the response frequency range depends on the ratio between the frequency of the dynamic phenomena to measure, and the natural frequency of the transducer [52].

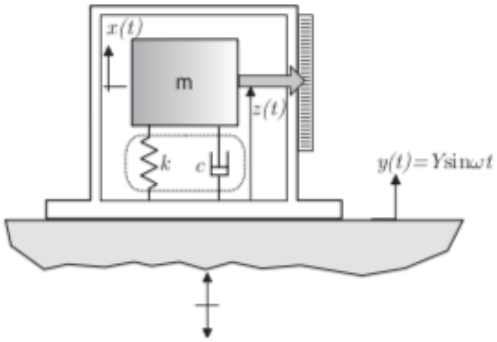


Figure 2.12: A model of a seismic transducer [52].

The theory behind this type of transducers falls on the model of mass-spring-damper system with one degree of freedom. Consequently, a vibration transducer is composed by a mass-spring-damper system placed on a vibrating object, Figure 2.12. The vibration is measured by the measuring the displacement of the mass of the transducer, the seismic mass, relatively to the vibrating object where the transducer is mounted on and whose movement the transducer is solidary with [52].

Denoting the seismic mass as m, the elastic element with a rigidity of k, and the damping constant of the damper as c, then, if an harmonic movement is assumed, the vibrating movement of the system $y\,(t)$ can be written as [52]:

$$y\,(t) = Y \sin(\omega t) \tag{2.9}$$

if x is defined as the movement of the seismic mass, then:

$$m\ddot{x} + c\,(\dot{x} - \dot{y}) + k\,(x - y) = 0 \tag{2.10}$$

The relative movement between the mass relatively to the box of the transducer can be defined as $z(t)$:

$$z\,(t) = x\,(t) - y\,(t) \tag{2.11}$$

hence, equation 2.10 can be rewritten as:

$$m\ddot{z} + c\dot{z} + kz = -m\ddot{y} \tag{2.12}$$

If one substitutes 2.9 into 2.12 and derives 2.12 in order of time, it can be rewritten as:

$$m\ddot{z} + c\dot{z} + kz = m\omega^2 Y \sin(\omega t) \tag{2.13}$$

The homogeneous solution to 2.12 is:

$$z\,(t) = Z\,(\omega) \sin(\omega t - \phi) \tag{2.14}$$

where its amplitude, $Z(\omega)$ and phase ϕ are given by:

$$Z(\omega) = Y\frac{\beta^2}{\sqrt{(1-\beta^2)^2 + (2\xi\beta)^2}} \quad \phi = \arctan\frac{2\xi\beta}{1-\beta^2} \tag{2.15}$$

A Seismic Transducer can implemented in the form of a vibrometer or an accelerometer. The vibrometer has a low natural frequency, so that it is lower than the frequency of the vibrating movement one wishes to measure. The accelerometer, on the other hand, has a high natural frequency, so that it is higher than the frequency of the vibrating movement one wishes to measure. Consequently, vibrometers are usually denoted as low frequency transducers, whereas accelerometers are usually known as high frequency transducers [52].

If the seismic mass transducer is implemented via an accelerometer, then, the homogeneous solution to equation 2.12 is slightly different than the one seen in equation 2.15. If one derives expression 2.12 in order to time one obtains the following expression:

$$z(t)\,\omega_n^2 = \frac{1}{\left((1-\beta^2)^2 + (2\xi\beta)^2\right)^{\frac{1}{2}}}\left\{-\omega^2 Y\sin(\omega t - \phi)\right\} \tag{2.16}$$

If, in the above expression, the condition is valid:

$$\frac{1}{\left((1-\beta^2)^2 + (2\xi\beta)^2\right)^{\frac{1}{2}}} \cong 1 \tag{2.17}$$

then equation 2.16 can be rewritten as:

$$-z(t)\,\omega_n^2 \cong \omega^2 Y\sin(\omega t - \phi) \tag{2.18}$$

Comparing equation 2.18 to $\ddot{y}(t) = -\omega^2 Y\sin(\omega t)$, one can verify that the term $-z(t)\,\omega_n^2$ gives the acceleration of the base, or vibrating object, $\ddot{y}$ subtrated by the phase, or angular difference, ϕ. Thus, the instrument can record or give directly the value of $\ddot{y} = -z(t)\,\omega_n^2$ as long as condition 2.17 is verified. The time delay between the measurement itself and the acceleration is given by $t' = \frac{\phi}{\omega}$. If $\ddot{y}$ is composed by only one harmonic component, then the difference in phase is not important.

The value of the first member of condition 2.17 is shown in figure 2.13 as a function of the ratio of frequencies $\beta = \frac{\omega_n}{\omega}$.

Given that condition 2.17 is verifiable for low values of β, the accelerometer's natural frequency (or resonance frequency) has to be high when compared to the frequency of the vibrating phenomena to measure. From the definition of resonance frequency $\omega_n = \sqrt{\frac{k}{m}}$ one can verify that the accelerometer must have high rigidity an low mass, which is why the accelerometers are considered as transducers that do not affect the acceleration to measure. Due to their reduced dimensions and mass, along with very high sensitivity, accelerometers are the preferred transducers for measuring vibrations in mechanical applications [52].

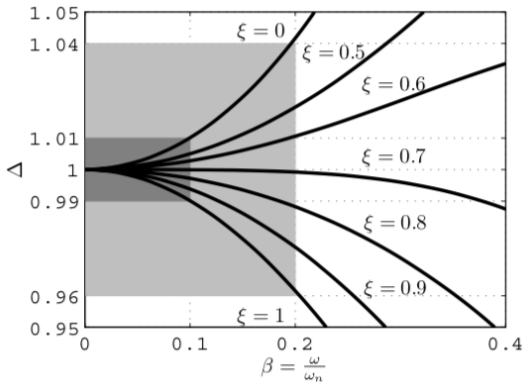


Figure 2.13: Responce of an accelerometer [52].

Decision on what type of transducer to use falls on what one wants to measure, obtain or analyze, specifically: acceleration, velocity and/or displacement. If one chooses to use an accelerometer, its output can be processed in order to obtain any of the referred physical quantities. If acceleration isn't important for the work being developed, then, a velocity transducer can be used, whose output can be integrated to yield displacement. Finally, if one's needs are working with displacement only, then a non-contacting probe, whose output is directly proportional to the relative displacement between the rotating and non-rotating elements of the machine, is sufficient.

Additionally, each one of these type of sensors has different characteristics. Given this, the decision on which one to use must be a balance between what is needed and the sensors' characteristics required for the project. Figure 2.14 shows guidelines between the dynamic ranges of each type of transducer and the frequency [51].

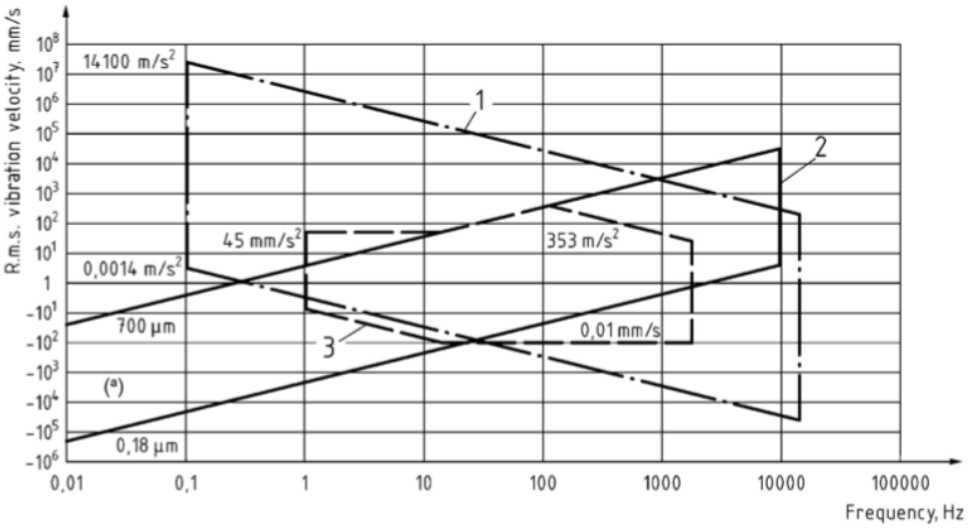


Figure 2.14: Dynamic range versus frequency range of vibration transducers for typical condition monitoring applications [51].

Where 1, 2 and 3 stand for Piezo-electric accelerometer, Eddy-current proximity probe and Electro-mechanical velocity transducer, respectively [51]. This text will only cover accelerometers as a form of measuring vibration, for it was the transducer chosen to acquire vibration signals for this project.

Other aspects, perhaps even more important, to take into consideration when choosing an accelerometer, are its specifications and if they meet the requirements for the analysis. These characteristics can be:

- **Sensitivity** - Defined as the ratio of change in the input, or acceleration, to the change in the output signal. Sensitivity is specified at a particular supply voltage and is typically expressed in mV/g [53];
- **Frequency Range** - For reliable condition monitoring, measuring equipment shall be capable of covering a wide frequency range in order to encompass not only shaft rotational frequencies and harmonics, but also frequencies due to other components, such as bearings, gears, seals, blades or vanes. The linear frequency range of the system should generally cover frequencies from 0,2 times the lowest rotational frequency to 3,5 times the highest excitation frequency of interest, which generally not exceed 10 kHz [51];
- **Resolution** - As in the smallest acceleration that can be detected by the transducer;
- **Number of axis** - there are single, double and triaxial accelerometers. The three axis accelerometer will measure acceleration in three orthogonal directions.

Furthermore, [51] states that the location and placement of the accelerometer are of utmost importance, to accurately acquire a vibrational signal from the machine under analysis. According to this standard, accelerometers should be attached to the machine through stud-mounting procedure (see figure 2.15). This permits the transfer of high frequency signals with little or no signal loss between the origin and the transducer.

Figure 2.15: Stud-mounting procedure according to the manufacturer PCB Piezotronics.

If such mounting procedure is not possible, adhesives, glue or magnetic

couplings are used to fix the sensors. The adhesives however, should have high stiffness characteristics when cured. The technique adopted for this thesis' measurement campaign was to glue the accelerometer to the bearing housing.

Another technique is to use a permanent magnet, however in order to effectively use magnets the mounting surfaces' flatness, is also a very important aspect to consider.

Furthermore, the addition of mass to the accelerometer - such as an adhesive or magnetic mounting base - lowers the resonant frequency of the sensing system and may affect the accuracy and limits of the accelerometer usable frequency range. Given this fact, these mounting techniques should be used carefully.

Although manufacturers use different terminologies for their products, four main techniques for sensing acceleration are piezoelectric, piezoresistive, capacitive or servo accelerometers. These can be further split into Alternated Current (AC) or Direct Current (DC) types. The main differences are that AC accelerometers cannot measure static acceleration but are appropriate for dynamic tests while DC accelerometers can detect constant acceleration such as gravity making them able to measure vibrations of 0 Hz, theoretically [53].

The most popular AC accelerometers are equipped with piezoelectric elements as their sensing mechanism, while the most popular DC accelerometers are capacitive MEMS and piezoresistive sensors types [53].

Piezoelectric Accelerometers

These devices have an extended region of flat frequency response range, a large linear amplitude range and excellent durability. These properties are due to the use of a piezoelectric material as the sensing element for the sensor. Piezoelectric generate an output electrical signal that is proportional to the stress applied to the material.

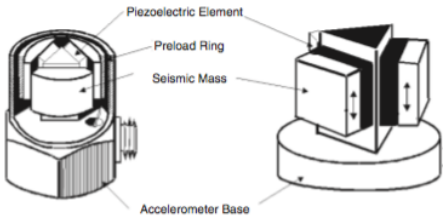


Figure 2.16: Construction of a piezoelectric accelerometer [54].

The piezoelectric element acts as a spring, with a natural stiffness k, and is responsible for connecting the base of the accelerometer to the seismic masses, m. When motion is induced to the base of the accelerometer, the piezoelectric material forces the masses to follow this motion. This causes a slight deformation of the piezoelectric element, translating the deformation to a strain. Consequently an electric charge proportional to the acceleration is generated.

The response frequency of the sensor is determined by its resonant frequency, which can be obtained by applying $\omega = \sqrt{\frac{k}{m}}$. Figure 2.17 shows the typical response frequency of piezoelectric accelerometers.

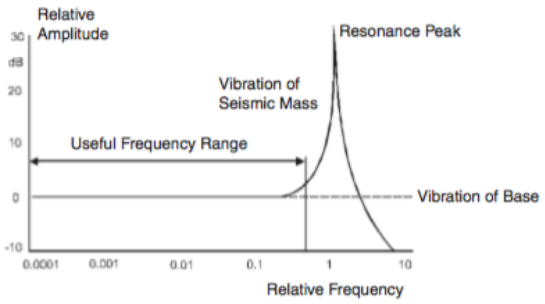


Figure 2.17: Typical response frequency of a piezoelectric accelerometer [54].

Piezoelectric accelerometers can be broken down into two main categories: internally amplified accelerometers, IEPE (Internal Electronics Piezoelectric), which contain built-in microelectronics for signal conditioning, or charge mode accelerometers that contain only the self-generating piezoelectric sensing element and have a high impedance charge output signal [54]. Other designs can even detect accelerations within several directions, using piezoelectric materials that can be subjected to shear deformations, which can again be translated into an acceleration measurement.

Piezoresistive Accelerometers

Single-crystal silicon is often used in manufacturing accelerometers. It is an anisotropic material whose atoms are organized in a lattice having several axes of symmetry [54].

Piezoresistive accelerometers rely on silicon both as the flexural element and as the transduction element, since the strain gauges are diffused directly into the flexure. The advantage of using this type of accelerometers is a high resonant frequency, that optimizes its frequency response, due to their relative high stiffness [54].

Capacitive Accelerometers

Capacitive accelerometers are similar to piezoresistive accelerometers which measure a change across a bridge. However, instead of measuring a change in resistance, they measure a change in capacitance. The sensing element consists of two parallel plate capacitors acting in a differential mode. These capacitors require built-in electronic circuit and operate in a bridge configuration. Additionally, these sensors are dependent on a carrier demodulator circuit or its equivalent to generate an electrical output that is proportional to acceleration [54].

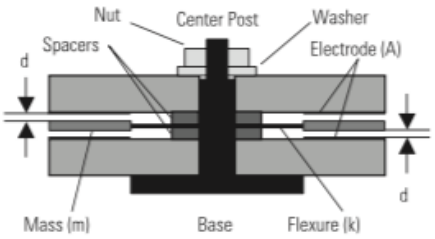


Figure 2.18: Construction of a capacitive accelerometer [54].

The main advantage of capacitive accelerometers is to measure low level - less than 2 g's - and low frequency acceleration with the capability of withstanding high shock levels, typically 5 000 g's or greater. Some of the disadvantages of the capacitive accelerometer are limited high frequency range, a relatively large phase shift and higher noise floor than an equivalent piezoelectric accelerometer [54].

Table 2.2 shows typical accelerometer characteristics for each of the types of accelerometers studied.

Table 2.2: Typical accelerometer characteristics.

Accelerometer Type	Frequency Range	Sensitivity	Measurement Range	Dynamic Range
Piezoelectric	0,5 Hz to 50 kHz	0,05 mV/g to 10 V/g	0,000001 g's to 100 000 g's	∼120 dB
Piezoresistive	0 to 10 000 Hz	0,0001 mV/g to 10 mV/g	0,001 g's to 100 000 g's	∼80 dB
Capacitive	0 to 1 000 Hz	10 mV/g to 1 V/g	0,00005 g's to 1 000 g's	∼90 dB

2.2.2 Acoustic Emission analysis

Wind turbines rotate at relatively slow speeds. This sets a limitation in early fault diagnosis using vibration monitoring method because it proves rather difficult to detect low frequencies of slow moving bearings. AE sensing detects the surface stress waves generated when faulty components come into contact. This technology has been as a suitable enhancement to the classic vibration techniques for condition monitoring of roller-bearings, gearboxes and wind turbines, specially for early detection of faults on these equipments [49, 55, 56].

Sources of AE in rotating machinery include impacting, cyclic fatigue, friction, turbulence, material loss, cavitation, leakage, among others. For instance, ball bearings rolling over a faulty outer race with a crack, or generally surface imperfections, will generate AE waves that can be sensed through the application on an AE sensor [55].

Tandon showed in [56] the measurement and interpretation of AE parameters for

fault detection in radially loaded ball bearings at different speed ranges. Additionally, Tan covered the application of AE for the detection of bearing failures in [57].

Acoustic monitoring is somewhat similar to vibration monitoring. However, whereas vibration sensors are mounted on the component involved in order to detect movement, acoustic sensors are attached with flexible glue, with low attenuation properties, and can record ultrasound, up to 100 kHz, directly [49]. García Márquez also finds that AE sensors have been used successfully not only in the monitoring of bearings and gearboxes but also for damage detection in blades of a wind turbine [49].

AE was originally developed for non-destructive testing of static structures, however, over the years its application has been extended to health monitoring of rotating machines [55]. In spite of the advantages given by using AE sensors, when compared to vibration sensors, limitations in the successful application of AE technique for condition monitoring of rotating machinery are partly due to the difficulty in processing, interpreting and classifying the intelligent information from the acquired data. The main obstacle with this technique is the attenuation of the signal, hence the need for the AE sensor to be close to its source. Another important factor to take into consideration when choosing an AE sensor is the frequency dynamic range. Given that there are different AE sensors with different dynamic ranges, choosing a sensor with a frequency range compatible with the application is important.

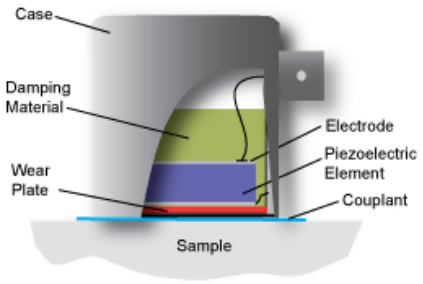


Figure 2.19: General *anatomy* of an AE sensor.

2.2.3 Sound analysis

Since the early times that *hearing* and *touching* human senses have been used, by experienced technicians, to asses if a machine is running with a faulty component. Specialized technicians could even fine tune older equipments by applying fine adjustments to it, hearing it work for a few seconds, and if needed proceed to apply finer adjustments. One can have already a simple idea that some faulty components do generate noise.

Nowadays however, this is analysed using a more scientific method. Microphones are now used to perform sound monitoring on industrial equipment. Using microphone analysis to this end has the main advantage of being a non-intrusive method when comparing to vibration analysis for example.

However, in noisier environments, like factories that rely on a large number of noisy machines, background and unwanted noise from surrounding components must be shielded. Skipping this step will corrupt the noise coming from the component of

interest, yielding incorrect results. Therefore, sound monitoring is not applicable for processing facilities having many electric machines in one room without this issue being addressed [58].

A microphone provides an analog output signal which is proportional to the variation in acoustic pressure acting upon a flexible diaphragm. This electrical signal is then used to transmit, record or measure the characteristics of the acoustic signal.

Measurement microphones differ from those used for audio applications since their primary role is to electrically reproduce the sound waveform without distortion and with a linear relationship between the voltage out and the pressure sensed by the microphone diaphragm.

This precision must be maintained over a wide range of frequencies and amplitudes when measuring sound waves arriving from different angles. Furthermore, they are expected to maintain this degree of precision over a range of temperature and barometric pressure variations.

A microphone's sensitivity is the relationship between the output voltage and the acoustic pressure sensed by the diaphragm and is expressed in units of mV/Pa.

The magnitude of the sensitivity is important because it is inherently related to electrical noise of the measurement system, since one cannot properly measure a voltage that is near the voltage noise floor of the instrument itself. Hence, for a given microphone and measurement system, the magnitude of the sensitivity will establish the minimum sound pressure, which can be accurately measured [54].

Any change on the sensitivity produces a distortion in the output signal compared to the acoustic signal originally generated on the component being studied. Thus, the following parameters are essential in a microphone measurement [54]:

- The sensitivity should be nearly constant over the range of frequencies to be measured. Thus, a quality measurement microphone should have a "flat" frequency response.
- The sensitivity should be nearly constant over a wide range of sound pressure levels. This is expressed as linearity, since a constant sensitivity would produce a straight-line graph of output voltage versus sound pressure.
- The sensitivity should be nearly constant over a wide range of temperature and barometric pressure, so as to not be affected by changes in room temperature or pressure.

Common Microphone Types

- **Dynamic microphones** - An electrical coil connected to the diaphragm is moved through an electrical field, generating a voltage proportional to the velocity of the moving element. Dynamic microphones have characteristics that are good for audio applications, but have high sensitivity to vibration, limited dynamic range and their frequency response is generally not adequate for measurement applications.

- **Piezoelectric Microphones** - The effect of sound pressure acting on the diaphragm is transmitted to a piezoelectric element which generates a charge

proportional to the original pressure. These microphones are often used in very high pressure situations, measuring explosive waves pressure for instance. In addition, due to their robust design, they can be used to measure small dynamic pressure fluctuations that precede large static pressures. In general, the noise floor of these microphones limit their use for precision sound measurement applications.

- **Condenser (capacitive) microphones** - the relative movement of the diaphragm to a fixed backplane produces a variation in capacitance that is proportional to the deflection of the diaphragm. A built-in electrical circuit in the microphone preamplifier converts the capacitance fluctuation to a voltage variation. The characteristics of condenser microphones, such as high sensitivity, wide dynamic range, flat frequency response, low internal noise, low distortion and high stability, make them the design of choice for measurement microphones [54].

CHAPTER 3

Signal Processing: Pre-Processing Techniques Applied

This section aims to clarify and give an insight into the signal processing and enhancement techniques used on this project. At the end of this chapter the method applied will be explained along with illustrations in order to clarify the adopted procedure. However, before proceeding to clarify the techniques used, one fundamental idea behind signal acquisition is the sampling frequency, which will be further explained.

Sampling & Aliasing

The basic pre and post processing operations shall be properly executed to enable further, often complex operations on collected measuring data, obtained from measuring system [53]. In signal processing, sampling is the transformation of a continuous signal to a discrete one, a technique that has been extensively used for signal reconstruction.

Whenever a signal is being measured, the number of samples taken in one second of the signal is called sampling frequency f_s, expressed in Hz. The sampling frequency used is related to the maximum frequency measured of the signal, i.e. its frequency bandwidth. This rate should be set in such a way to enable reconstruction of analog signal from discrete data without losing information from the original signal [53]. An illustration on this concept is found on Figure 3.1:

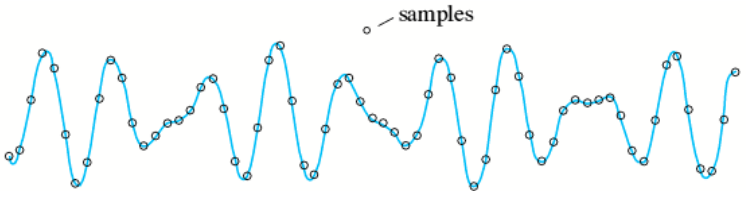


Figure 3.1: Input signal and samples taken [59].

For this conversion to be possible without losing valuable information, the Shannon-Kotielnikov condition must be fulfilled [53]. This condition states that the

sampling rate cannot be less than twice of the highest frequency - the Nyquist frequency - of a measured signal. If this condition is not fulfilled, then the phenomenon of aliasing occurs, see Figure 3.2.

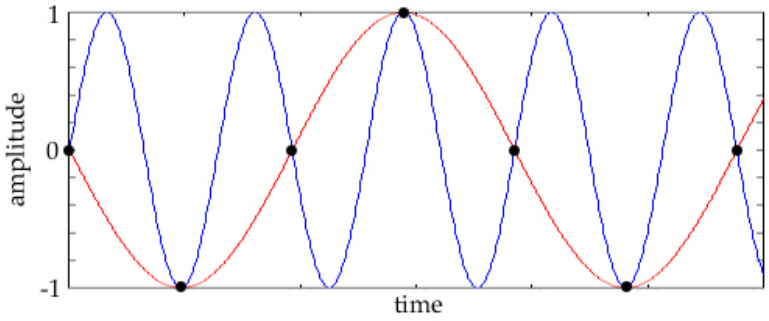


Figure 3.2: Aliasing phenomenon. The blue curve represents the measured signal and the red curve is the reconstructed signal [60].

It is stated in [53] that for condition monitoring, one must oversample a measured signal, as in the minimum sampling rate being much higher than the one given from the Shannon-Kotielnikov condition. The same source recommends that, when acquiring stationary signals from low speed condition monitoring, the sampling rate should be 5 to 10 times higher than the Nyquist frequency.

3.1 Order Tracking - Angular Resampling

In real working conditions, bearings operate under speed fluctuations which will result in a non-stationarity character of the bearing vibration signal and, consequently, the usual techniques for vibration analysis that are based on the assumption of constant rotating speed cannot be applied [61, 62].

In order to overcome this obstacle, order tracking algorithms and techniques have been proposed to remove the effects of speed fluctuation of REBs and smearing of the spectrum by resampling the original vibration at a constant angle increment and converting the non-stationary signal in time domain to a stationary signal one in angular domain [62].

Order tracking is a frequency analysis method that uses multiples of the running speed (orders), instead of absolute frequencies (Hz), as the frequency base. Order tracking is useful for machine condition monitoring because it can easily identify speed-related vibrations such as shaft defects and bearing wear [61].

This technique also allows to see how the intensity of the different harmonics changes over a big speed range [1]. Randall states in [1] that if a constant amplitude signal, synchronous with the rotation of a shaft, for instance, is sampled at a rate per revolution, then the digital samples are indistinguishable from the samples of a sinusoid, resulting in a line in the spectrum. On the other hand, of normal temporal sampling is used, the spectrum will spread over a range that corresponds to the fluctuation in shaft speed.

To avoid this, it is necessary to generate a sampling signal from a tacho or shaft

encoder signal that is synched to the shaft speed, as known as order analysis [1]. However, using this approach in order to apply order analysis, has the disadvantage of having a limited time response that might not be sufficient to detect random speed fluctuations along one cycle [1].

The best method is to use the angular resampling technique. To resample each record digitally, based on the corresponding period of the tacho signal, so as to achieve sampling for uniform increments in shaft rotation angle. In other words, the signal is resampled at a constant angle, but continuously changing with time [6]. Figure 3.3 shows an example of performing angular resampling through Computer Order Tracking (COT) and its stages: (a) determine the number of pulses over time; (b) obtain the cumulative angle, $\phi(t)$, of the shaft over time; (c) calculate the resampling time instants for the constant angular increments θ and (d) signal resampling by interpolation.

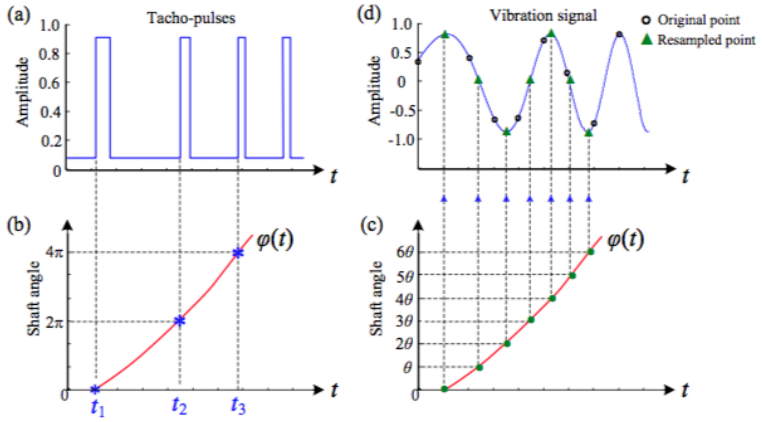


Figure 3.3: Angular resampling, through Computer Order Tracking [63].

As found in [1], based on digital interpolation, there are several ways of performing angular resampling. One way is to increase the sample rate by a large factor, and then select the nearest sample to each theoretical interpolated position. Increasing the sample rate by an integer factor can be achieved in two ways. In the time domain, it can be done by inserting the appropriate number of zeros in between each actual sample, and then applying a digital low-pass filter to limit the frequency range to the original maximum, thus smoothing the curve.

In the frequency domain the same result can be accomplished by filling the spectrum with zeros in the centre (i.e. around the Nyquist frequency) and then inverse transforming the increased (two-sided) spectrum to the same increased number of time samples.

On the other hand, more accurate interpolation, not limited to a ratio of integer numbers, can be achieved by fitting a curve to a group of samples and then calculating the value of the polynomial at the interpolated positions.

Another approach to angular resampling would be to use phase demodulation of a signal synchronous to the phenomenon of interest, shaft speed for example, to obtain a mapping of shaft rotation angle versus time.

One must take into consideration that if order tracking is being performed

directly on an analogue signal, it must be ensured that the signal is adequately low-pass filtered to prevent aliasing, in particular when resampling at a lower frequency [1].

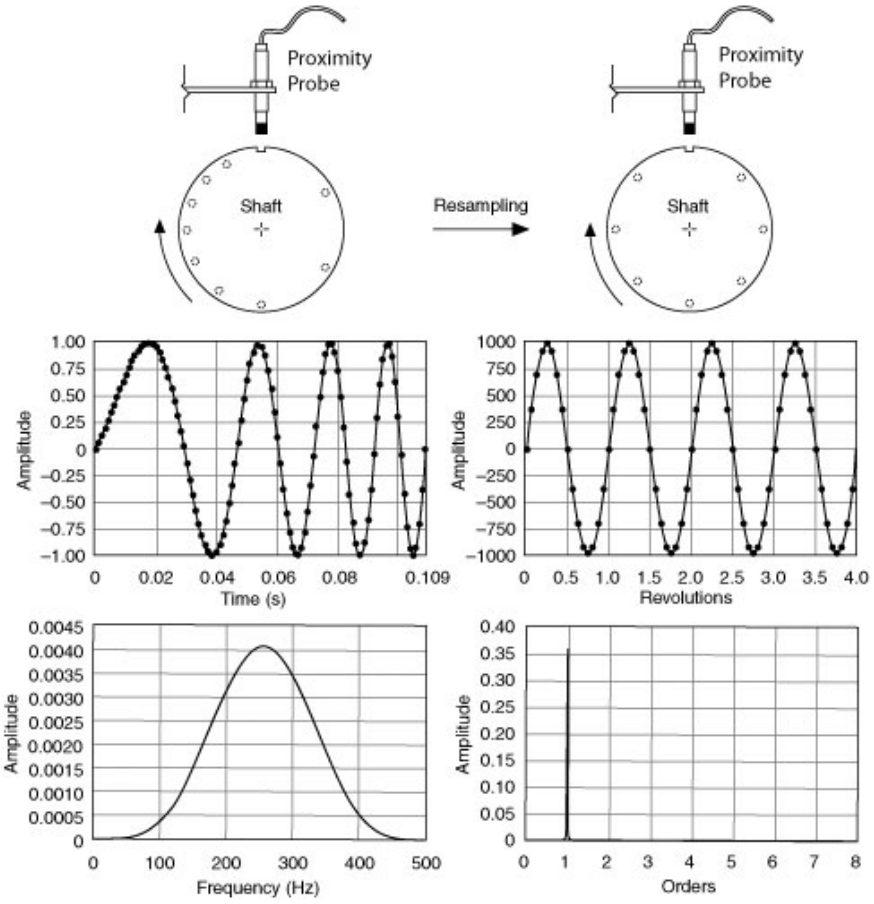


Figure 3.4: Angular Resampling according to the manufacturer of testing equipment National Instruments.

3.2 Removal of deterministic components - Cepstral Editing procedure analysis

Removal of deterministic signals (i.e. discrete components), which are typically dominating components in machinery vibration signals, is an important step in the diagnosis of rolling element bearing faults. There are numerous techniques found in the literature that allow the removal of these components: linear prediction, adaptive noise cancellation, self-adaptive noise cancellation, discrete/random separation, time synchronous averaging and Cepstral Editing Procedure (CEP) [64].

The CEP has become widely used for removing these unwanted components and has been extensively studied lately due to its easy interpretation and implementation, as well as its relatively low computational cost yet good performance [65].

The challenge of removing the unwanted deterministic components is important

for these contribute with high-energy dominating signals, originating from gears, screws or shafts, and mask the wanted, non-deterministic - i.e. random - signals that bearing faults generate. These unwanted signals appear as discrete components in the frequency domain.

Given this, the removal of these dominant components must be applied before further signal processing techniques [65].

Before going into detail of the specific cepstral method applied in this project, a few basic ideas about cepstrum analysis should be looked into. The word cepstrum is originated by reversing the first syllable of spectrum, which was justified because the cepstrum is the spectrum of a spectrum [1].

Following this pattern of reversing the first syllable, terminology also has different names found on table 3.1

Table 3.1: Cepstrum Terminology.

Frequency Analysis	Quefrequeny Analysis
Harmonic	Rahmonic
Filter	Lifter
Low-pass filter	Short-pass filter
High-pass filter	Long-pass filter
Magnitude	Gamnitude
Phase	saphe

Cepstrum

The Cepstrum method was initially presented as the power spectrum of the logarithm of the power spectrum. Later, the term *power cepstrum* was redefined as the inverse Fourier Transform (FT) of the log power spectrum, partly because it is more logical to use the inverse transform between a function of frequency and a function of time and also because it is then reversible to the power spectrum [1]. This redefinition however, was only possible after the publication of the paper, and algorithm for the Fast Fourier Transform (FFT). What distinguishes the autocorrelation function from the cepstrum, is that the cepstrum is the logarithmic conversion of the spectrum before the second transform [1].

Different variations, or formulations, of cepstrum analysis exist. Different than the power cepstrum, the complex cepstrum was defined as the inverse FT of the complex algorithm of the complex spectrum [1, 16]. The real cepstrum is defined by setting the phase of the phase of the complex cepstrum to zero [16].

Matematically, where $F_{XX}(f)$ is a power spectrum, the original definition of the (power) cepstrum can be written as:

$$C_p(\tau) = \left|\Im\left\{\log\left(F_{XX}(f)\right)\right\}\right|^2 \tag{3.1}$$

and if one recalls the definition of the complex cepstrum, than it becomes:

$$C_c(\tau) = \Im^{-1}\left\{\log\left(F(f)\right)\right\} = \Im^{-1}\left\{\ln\left(A(f)\right) + j\phi(f)\right\} \tag{3.2}$$

where:

$$F(f) = \Im\{(f(t))\} = A(f)\, e^{j\phi(f)} \tag{3.3}$$

in terms of the spectrum's amplitude and phase. Although *complex cepstrum* is called complex, it is real because the log amplitude of the spectrum is even and the phase of the spectrum is odd [16].

Given this, the *new* power cepstrum is then given by:

$$C_p(\tau) = \Im^{-1}\{\log(F_{XX}(f))\} \tag{3.4}$$

and for the spectrum of a single record (as in 3.5) it can be expressed as:

$$C_p(\tau) = \Im^{-1}\{2\log(A(f))\} \tag{3.5}$$

As it has already been said, the real complex is obtained by setting the phase of the power cepstrum to zero in Eq. 3.7:

$$C_p(\tau) = \Im^{-1}\{\ln(A(f))\} \tag{3.6}$$

which is just a scaled version of (3.4) [16].

An important note is that the complex cepstrum requires the phase $\phi(f)$ to be unwrapped to a continuous function of frequency. Because of this, it cannot be used with stationary signals, consisting of discrete frequency components. These signals' phase is undefined at intermediate frequencies. It cannot be used also with signals with stationary random components, whose phase is random. The complex cepstrum can only be used with well-behaved functions such as impulse responses, where the phase is well-defined [16].

Cepstral Editing Procedure

The distinctive property of periodic signals in the cepstrum domain is that they appear as narrow impulses with a number of rahmonics. Based on this property Randall proposed a method for removing these periodic signals from the cepstrum [66, 67]. The same author also found [68] that the CEP method enhances bearing fault-related signals more significantly when compared with the other methods for discrete components elimination that were stated previously.

The method used, and suggested by Randall, removes the periodic components from the cepstrum by editing - liftering - the magnitude of the real cepstrum at quefrencies of interest. This liftering procedure is performed using information about the shaft speed and machinery configuration [65].

According to [69] the log amplitude spectrum of stationary signals can be edited using the real cepstrum, and then the edited spectrum amplitudes combined with the original phase spectrum are used to return to the time domain. On a more concise level, Randall describes this CEP method as a pre-whitening operation consisting in setting an amplitude of zero for the real cepstrum, at determined frequencies. After this, once the signal has been transformed back to the frequency domain, the resulting signal is recombined with the phase of the original signal and inverse transformed to

time domain. This is equivalent to applying series of liftering operations around the quefrencies of the deterministic excitations. This procedure results in deleting almost completely the effect from these deterministic components on the signal as well as removing the resonance effects [66].

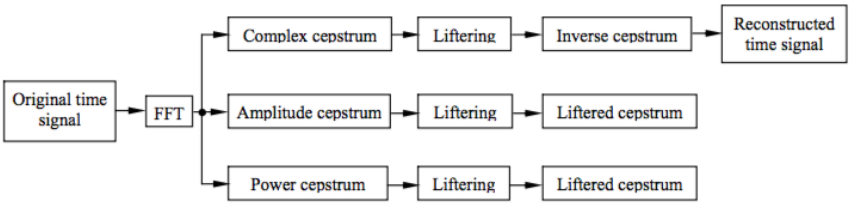


Figure 3.5: Schematic of performing cepstrum analysis [65].

At the same time, the bearing damage related components, cyclostationary of the second order and not strictly periodic, do not present a strong peak in the absolute value of the cepstrum and will not be affected by the liftering. Enhancing the effect of these signals, which are the effective signals that one wants to obtain can also be implemented simply by dividing the Fourier transformed signal by its absolute value and transforming all of this again to the time domain, all of this while avoiding the transformation to the cepstral domain [66]:

$$x_{cpw} = IFT\left\{\frac{FT\left(x\right)}{\left|FT\left(x\right)\right|}\right\} \tag{3.7}$$

3.3 Frequency Bands Identification Methods

3.3.1 Cyclic Spectral Correlation

Cyclostationarity covers a subclass of non-stationary signals which exhibit some cyclical behaviour. Such signals are generally not periodic but random in their waveform, yet inherently generated by some periodic mechanism. To avoid confusion, the period of a cyclostationary signal is referred as cycle. An example of a cyclostationary signal is a random noise that is amplitude modulated by a periodic function [70].

Aplication of Cyclostationary analysis proves to be interesting to the diagnostics of machine vibration signals. This is because some machine signals, even though being almost periodic, are not exactly phase-locked to shaft speeds, and thus even after processed in order to compensate for speed fluctuation cannot be extracted by synchronous averaging [71]. Typical examples of cyclostationary signals are the combustion events in IC engines, which vary from cycle to cycle, and impulsive signals from faults in rolling element bearings, which are affected by minor but randomly varying slip.

In such cases, the signals are not strictly periodic, but cyclostationary (of second

order), meaning that their second-order statistics such as autocovariance function are periodic [71].

Moreover, the assumption that a vibration signal is cyclostationary affords much more information than the usual and simplistic assumption of stationarity, since it provides the tools to simultaneously analyse the content of a signal, for instance its spectral content, and the characterisation of how this content evolves periodically in time [70]. In REBs, cyclostationarity is symptomatic to the presence of faults, proved by the occurrence of repetitive shocks when a defect impacts a rolling surface - a series of repetitive shocks may be seen as a signal periodically amplitude modulated in time. The rate of repetition of these shocks (the cyclic frequency) reveals the origin of the fault, and the cyclostationarity intensity may serve to indicate its severity [70].

A signal is cyclostationary of degree N if its Nth moment is periodic about any period. In other words, a cyclostationary signal is described by a periodically time-varying probability density function. Furthermore, cyclostationary analysis is another spectrum method applied to fault detection of gears and bearings, and takes advantage of the stochastic process nature of the vibration signal [72]. In a strict sense, a (quasi-) cyclostationarity signal $X[n]$ is a signal whose joint probability density function is a (quasi-) periodic function of time.

Given that CEP method is applied in order to remove the discrete frequency components from the signal, which correspond to first-order components, only second-order components will be left in the signal to be analysed. This is the reason why second-order descriptors are used to perform cyclostationary analysis of bearings. J. Antoni shows [73] that cyclosationary analysis can either be done using time descriptors, more precisely the instantaneous auto-correlation function or the cyclic auto-correlation function, or through frequency descriptors, such as the spectral correlation or the cyclic power spectrum. Randall, on the other hand, states that in practice it is best to use the frequency domain to estimate spectral correlations [1]. In other words, perform cyclostationary analysis (cyclostationary analysis consists in a series of correlations). Hence the reason for the analysis performed in this project being on frequency domain, or more precisely, angular domain. Moreover, Randall and J. Antoni worked extensively on this topic and the descriptor J. Antony used for bearing diagnosis assessment was the spectral correlation.

According to [1, 70], by denoting $X_L(f)$ as the Fourier transform of a cyclostationary signal $x(t)$ evaluated over an interval of length L, CSC can be expressed as:

$$S_x(f;\alpha) = \lim_{L\to\infty} \frac{1}{L} E\left\{ X_L\left(f + \frac{\alpha}{2}\right) X_L\left(f - \frac{\alpha}{2}\right)\right\} \tag{3.8}$$

where E denotes the expected value.

The term $S_x(f;\alpha)$ is coined the cyclic power spectrum (signal-units^2/Hz). The physical meaning of the frequency f is correspondent to the time-lag τ, indicating the frequency of the carrier signal, where as frequency α is named cyclic frequency. Intuitively, the cyclic power spectrum may be interpreted as providing the distribution of the frequency content of signal $x(t)$ that statistically repeats itself with rate α [70].

The approach for this technique is the direct visual inspection of the magnitude of the cyclic power spectral density displayed as a graph over the f versus α frequency domain.

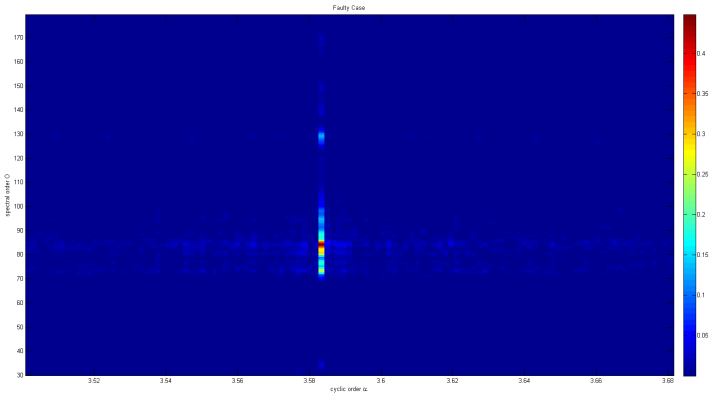


Figure 3.6: Example of graph analysis with CSC method.

As can be seen in Figure 3.6, this highlights with the central frequency to filter around using envelope analysis with filtering. Given this, the most interesting advantage of applying this procedure is that it can be used alongside envelope analysis. Moreover, the graph 3.6 itself can be used as a visual feature for further analysis. If the classification algorithms classify the situation as faulty, one can quickly verify this graph and see whether or not there is a highlighted area indicating failure or something that might indicate a potential failure.

3.3.2 Wavelet Analysis

In cases of very low speeds, as for cranes or wind turbines, the energy levels generated by the impacts (because of the faults) are low. Hence, applying CSC to these cases becomes a matter of getting the settings right to each case, which can be time demanding and is the opposite of an automated tool. An alternative can be the wavelet analysis. Furthermore, CSC has the disadvantage of being inefficient in terms of computer resources.

The wavelet analysis is yet another interpretation of time-frequency analysis as this technique decomposes the signal in terms of a family of *wavelets* that have a fixed shape, but can be shifted an dilated in time [1].

Mathematically, the complex continuous wavelet transform can be expressed as [1]:

$$W(a;b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\, \psi^* \left(\frac{t-b}{a} \right) \tag{3.9}$$

where ψ is the mother wavelet, translated, as in displacement, by b and dilated, as in expansion, by a. ψ^* is the complex function. In simpler terms, a, the scaling parameter measures the degree of compression while b determines the time location of the wavelet [74].

The dilation factor is known as scale and If $|a| < 1$, the wavelet W is a compressed version of the mother ψ wavelet and corresponds mainly to higher frequencies. In contrast, if $|a| > 1$ the wavelet W has a larger time width than the

mother wavelet ψ and corresponds to lower frequencies as illustrated in Figure 3.7. Accordingly, one can state that wavelets have time widths adapted to their frequencies. However, one may also observe that the resolution of wavelets at different values of the scaling parameter varies both in time and frequency domain, according to the Heisenberg uncertainty principle [74].

Given that equation 3.9 is a convolution, wavelets can be considered as a set of

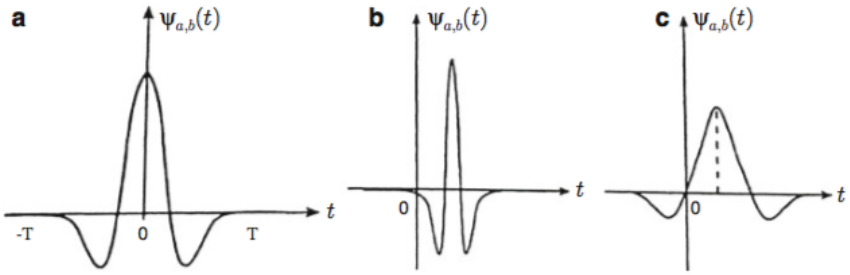


Figure 3.7: Effect of a and b. (a) Typicall mother wavelet. (b) Compressed and shifted wavelet: $|a| < 1$ and $b > 0$. (c) Magnified and shifted wavelet: $|a| > 1$ and $b > 0$ [74].

impulse responses of filters, which because of the dilation factor keep the bandwidth properties. Wavelets present the advantage of giving a better time localization at high frequencies, which can be useful for detecting local events in a signal. This means that this technique is mostly used to detect local faults in bearings and gears [1].

Wavelets can be divided into four main groups: orthogonal, non-orthogonal, continuous or discrete, and currently there are numerous families of wavelets available: Haar, Daubechies, Meyer, Gaussian, Mexican hat, Morlet, Shannon, Fejer-Korovkin, among others. The Daubechies dilation wavelets are compact and have irregular shapes in time domain, but are infinite in frequency domain. The complex wavelets, on the other hand, are compact in frequency domain, but infinite in time domain. These have the appearance of harmonic functions and are typically of one-octave bandwidth, even though at times they are narrower. The biggest advantage of complex wavelets is that the imaginary part of the wavelet is orthogonal to the real part, consequently, the overall result is not sensitive to the phasing of the event. Furthermore, the local sum of squares of the real and imaginary parts result in a smooth function [1].

On a more general form, orthogonal wavelets are the most efficient to use when analysis, or synthesis, has to be performed, or when the significant features of the signal have to be represented with a minimum number of parameters. For analytical purposes, non-orthogonal wavelets, Morlet wavelets for instance, are more convenient, and it is often generally preferable to use redundant 'lapped' transforms prioritizing visual interpretation.

The main applications of this technique in machine diagnostics is in denoising signals in both time and frequency domains simultaneously. There are two forms of thresholding to remove noise, by removing any components with less amplitude than a certain threshold value. One technique is *Hard thresholding* where the retained components (components not considered noise) are left unchanged. Oppositely, the technique *soft thresholding* estimates the noise using the threshold set, and these values are subtracted from the retained components [1].

The choice between families of wavelet to use must be done according to the similarities between the wavelets and the features in the signal that are to be extracted [1]. For bearing faults, which generate impulsive signals, using the impulse wavelet may return better results. However, it is common to choose the wavelet family by trial and error or even based on the knowledge of the analyst.

Morlet Wavelets

Morlet wavelets are non-orthogonal, Gaussian windowed sinusoids but suitable for analysing many machine signals because of their similarity in appearance to narrowband impulse responses, they can be tuned in order to correspond to different damping. The complex Morlet wavelet is defined in the time domain as a complex exponential wave multiplied by a Gaussian function and has the shape of a Gaussian window in the frequency domain. Originally, real Morlet wavelets were windowed cosines, but it is convenient to use the complex version, with a one-sided spectrum, so that the imaginary part is the Hilbert transform of the real part. Consequently, the complex morlet wavelet can be described as:

$$\Psi(f) = \Psi^*(f) = \exp\left[-\left(\frac{\pi^2}{\sigma^2}\right)(f - f_0)\right] \tag{3.10}$$

$$\psi(t) = \frac{\sigma}{\sqrt{\pi}} \exp(-\sigma^2 t^2) \exp(j2\pi f_0 t) \tag{3.11}$$

where f_0 is the window's central frequency and σ its width and * denotes the complex conjugate of $\Psi(f)$ that is the FT of $\psi(t)$, and because $\Psi(f)$ is real, its conjugate is equal to itself.

The advantage of complex Morlet wavelets, as opposed to the real version with windowed cosines, is that the imaginary parts have their maxima when the cosines hit values of zero, so that the squared amplitude of the wavelet coefficients is not sensitive to the phasing of local features in time signals [1].

3.4 Demodulation: Hilbert Transform

Contrary to other integral transforms, where there can be a change domains, such as the Fourier or Laplace transformations, the Hilbert Transform (HT) processes a function g into another one while maintaining the original domain. This operation, assigns a complementary imaginary part to a given real part or vice versa, by shifting each component of the signal by a quarter of a period [5]. To put in simpler terms, one can say that HT is the relationship between the real and imaginary parts of the FT of a one-sided function [1].

The Hilbert transform $\mathcal{H}[g(t)]$ of a signal $g(t)$ follows the eq. 3.12 [75]:

$$\mathcal{H}[g(t)] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{g(t-\tau)}{\tau}\, d\tau \tag{3.12}$$

The HT of $g(t)$ is the convolution of $g(t)$ with the signal $\frac{1}{\pi t}$. This is the response to $g(t)$ of a linear time-invariant filter called Hilbert transformer, with impulse response of $\frac{1}{\pi t}$ [5].

However, the integral translated in eq. 3.12 is improper: the integrand has a singularity and the limits of integration are infinite. In fact, the Hilbert transform is properly defined as the Cauchy principal value of the integral in 3.12, whenever this value exists. The Cauchy principal value is defined, for 3.12, as:

$$\mathcal{H}[g(t)] = \frac{1}{\pi} \lim_{\epsilon \to 0^+} \left(\int_{t-1/\epsilon}^{t-\epsilon} \frac{g(\tau)}{t-\tau}\, d\tau \; + \; \int_{t+\epsilon}^{t+1/\epsilon} \frac{g(\tau)}{t-\tau}\, d\tau \right) \tag{3.13}$$

One can see that the Cauchy principal value is obtained by considering a finite range of integration that is symmetric about the point of singularity, but which excludes a symmetric subinterval, taking the limit of the integral as the length of the interval approaches ∞ while, simultaneously, the length of the excluded interval approaches zero [75].

Some properties of the Hilbert transform are listed below [75].

1. **Linearity**

$$\mathcal{H}[a_1 g_1(t) + a_2 g_2(t)] = a_1 \mathcal{H}[g_1(t)] + a_2 \mathcal{H}[g_2(t)] \tag{3.14}$$

2. **Constant Signal** - For any constant c the transformation is 0.

$$g(t) = c \qquad \mathcal{H}[g(t)] = \mathcal{H}[c] = 0 \tag{3.15}$$

From the linearity property enumerated before it came that

$$\mathcal{H}[g(t) + c] = \mathcal{H}[g(t)] + \mathcal{H}[c] = \mathcal{H}[g(t)] \tag{3.16}$$

3. **Time Shifting** If $g(t)$ has Hilbert transform $\mathcal{H}[g(t)]$ then the transformation of $g(t - t_0)$ is given as follows

$$\mathcal{H}[g(t - t_0)] \tag{3.17}$$

4. **Convolution**

$$\mathcal{H}[g_1(t) \times g_2(t)] = \mathcal{H}[g_1(t)] \times g_2(t) = g_1(t) \times \mathcal{H}[g_2(t)] \tag{3.18}$$

On the other hand, Randall [1] suggests that the Hilbert transform can be simply the transformation of the signal into the frequency domain, shifting the phase of positive frequency components by $-\pi/2$ and of the negative frequency components by $+\pi/2$ and returning to time domain after the shifting.

An advantage of the HT is that the function $g(x) = f(x) + j\mathcal{H}[u(x)]$ is always representative of an analytic function, also known as regular function in the superior half of the complex domain. This means that the function is infinitely differentiable. Despite other useful applications, nowadays the biggest advantage of the HT is the ability to obtain the envelope by demodulation of a signal [1].

3.5 Fourier Transform and Power Spectral Density

Detecting waves, i.e. oscillatory behaviour, in the vibration signal provides useful information about the presence of faults. To perform this examination, time-frequency transformations are used. Time-frequency transformations are important mathematical signal processing techniques that process data to find the contribution made by each frequency during each point in time of the original signal.

One of the simplest transforms is the Fourier transform, an algorithm that works by decomposing the original signal into a linear sum of sinusoidal waves of different frequencies. The larger the amplitude of a given frequency, the more important that frequency is to the original signal. The output of the transform is this mapping of frequency to amplitude [1, 76, 10].

Discrete Fourier Transform (DFT)

Sampled time signals are in principle of infinite length, but when the record length is finite, this leads to the same situation as with the Fourier series in that the spectrum is discrete and the time record implicitly periodic. This leads to the time record and frequency spectrum being discretely sampled and periodic. The continuous infinite integrals of the Fourier transform become finite sums, usually expressed as the equation 3.19.

$$G(kf) = \frac{1}{N} \sum_{n=0}^{N-1} g(n) e^{\frac{-j2\pi kfn}{N}} \tag{3.19}$$

Where the angular frequency ω_n in rad/s has been replaced by the continuous frequency f expressed in Hz, k is an index and N is the length of the record. Dividing by this parameter, the Fourier series components are correctly scaled [1]. The forward DFT operation can be understood as the matrix multiplication shown in equation 3.20 [1]:

$$G_{kf} = \frac{1}{N} W_{kfn} g_n \tag{3.20}$$

Where G_kf represent the vector of N frequency components, g_n represents the N time samples $x(n)$. W_{kfn} represents a square matrix of unit vector $\exp(-j2\pi kfn/N)$ with angular orientation, depending the rows on the frequency index k and the columns on the time sample index n. As shown in equation 3.21 [1]:

$$\begin{bmatrix} X_0 \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \end{bmatrix} = \frac{1}{8} \begin{bmatrix} \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow \\ \uparrow & \nearrow & \rightarrow & \searrow & \downarrow & \swarrow & \leftarrow & \nwarrow \\ \uparrow & \rightarrow & \downarrow & \leftarrow & \uparrow & \rightarrow & \downarrow & \leftarrow \\ \uparrow & \searrow & \leftarrow & \nearrow & \downarrow & \nwarrow & \rightarrow & \swarrow \\ \uparrow & \downarrow & \uparrow & \downarrow & \uparrow & \downarrow & \uparrow & \downarrow \\ \uparrow & \swarrow & \rightarrow & \nwarrow & \downarrow & \nearrow & \leftarrow & \searrow \\ \uparrow & \leftarrow & \downarrow & \rightarrow & \uparrow & \leftarrow & \downarrow & \rightarrow \\ \uparrow & \nwarrow & \leftarrow & \swarrow & \downarrow & \searrow & \rightarrow & \nearrow \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} \tag{3.21}$$

For $k = 0$ the zero frequency value $G(0)$ is simply the mean value of the time samples $x(n)$. For $k = 1$ the unit vector rotate $-1/N$-th of revolution for each time

sample increment, which translates in a complete clockwise revolution after N samples. For higher values of k the rotation speed is proportionally higher. For $k = N/2$, also known as Nyquist frequency, or half the sampling frequency, the vector turns $-\pi$ for each time sample, however, it is not possible to assess in which direction it is turning. For $k > N/2$, the vector turns more than π in the negative direction, although it is easier to interpret this as turned in the opposite direction, which means a value less than π and that will transform the second half of G_{kf} in the negative frequency components ranging from minus 0,5 the Nyquist frequency to slight below zero. In order for this to be true, the time signal has to be low-pass filtered at half the sampling frequency.

Fast Fourier Transform (FFT)

The FFT is a simple and very efficient algorithm for calculating the DFT equations, requiring much less operations than the DFT. The only requirement for this new approach was that the number of points N, should be a power of 2, such as 128, 256, 512, etc..

Starting with the equation (3.20) in its simplest version, also known as radix 2 algorithm, because the FFT is based on N being a power of 2, it factorizes a modified version of the matrix W_{kfn} into $\log_2 N$ matrices. Accordingly, the number of mathematical operations is reduced, by a factor of more than 100 for a typical case where $N = 1024$ [1].

A modified version of the matrix W_{kfn} will be designated as matrix B, and to simplify the nomenclature, the original will be denoted matrix A. Matrix B has the rows arranged in bit-reversed order as shown in the equation (3.22), meaning that the most significant bit is indexed, rather than the least significant one, consequently incrementing the phase.

$$
\begin{array}{c|c|c}
\textit{RowB} & & \textit{RowA} \\
0\,0\,0\,(0) & \uparrow\ \uparrow\ \uparrow\ \uparrow\ \uparrow\ \uparrow\ \uparrow\ \uparrow & 0\,0\,0\,(0) \\
0\,0\,1\,(1) & \uparrow\ \downarrow\ \uparrow\ \downarrow\ \uparrow\ \downarrow\ \uparrow\ \downarrow & 1\,0\,0\,(4) \\
0\,1\,0\,(2) & \uparrow\ \rightarrow\ \downarrow\ \leftarrow\ \uparrow\ \rightarrow\ \downarrow\ \leftarrow & 0\,1\,0\,(2) \\
0\,1\,1\,(3) & \uparrow\ \leftarrow\ \downarrow\ \rightarrow\ \uparrow\ \leftarrow\ \downarrow\ \rightarrow & 1\,1\,0\,(6) \\
1\,0\,0\,(4) & \uparrow\ \nearrow\ \rightarrow\ \searrow\ \downarrow\ \swarrow\ \leftarrow\ \nwarrow & 0\,0\,1\,(1) \\
1\,0\,1\,(5) & \uparrow\ \swarrow\ \rightarrow\ \nwarrow\ \downarrow\ \nwarrow\ \leftarrow\ \searrow & 1\,0\,1\,(5) \\
1\,1\,0\,(6) & \uparrow\ \searrow\ \leftarrow\ \nearrow\ \downarrow\ \nwarrow\ \rightarrow\ \swarrow & 0\,1\,1\,(3) \\
1\,1\,1\,(7) & \uparrow\ \nwarrow\ \leftarrow\ \swarrow\ \downarrow\ \searrow\ \rightarrow\ \nearrow & 1\,1\,1\,(7)
\end{array}
\tag{3.22}
$$

The multiplication by B means that the results are also in bit-reversed order, but reorganize them to the correct place is a simple operation that can be done simultaneously. This operation also performs faster than the multiplications saved with this method. Equation (3.24) shows an example for a matrix B, with $N = 8$, factorized into three matrices X,Y and Z. For each one of those matrices there are only two non-zero elements for each row, the other one corresponds to 1 [1].

$$
\begin{bmatrix} I & I \\ I & -I \end{bmatrix} \tag{3.23}
$$

The factor matrices contain progressively finer rotations and the top left sub-matrix is always of the form given by equation (3.23). Factorization in powers different

than 2 is also possible, given that the properties of the FFT are the same as the DFT.

$$\begin{bmatrix} & & & & X & & & \\ \uparrow & \uparrow & 0 & 0 & 0 & 0 & 0 & 0 \\ \uparrow & \downarrow & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \uparrow & \rightarrow & 0 & 0 & 0 & 0 \\ 0 & 0 & \uparrow & \leftarrow & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \uparrow & \nearrow & 0 & 0 \\ 0 & 0 & 0 & 0 & \uparrow & \swarrow & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \uparrow & \searrow \\ 0 & 0 & 0 & 0 & 0 & 0 & \uparrow & \nwarrow \end{bmatrix} \begin{bmatrix} & & & & Y & & & \\ \uparrow & 0 & \uparrow & 0 & 0 & 0 & 0 & 0 \\ 0 & \uparrow & 0 & \uparrow & 0 & 0 & 0 & 0 \\ \uparrow & 0 & \downarrow & 0 & 0 & 0 & 0 & 0 \\ 0 & \uparrow & 0 & \downarrow & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \uparrow & 0 & \rightarrow & 0 \\ 0 & 0 & 0 & 0 & 0 & \uparrow & 0 & \rightarrow \\ 0 & 0 & 0 & 0 & \uparrow & 0 & \leftarrow & 0 \\ 0 & 0 & 0 & 0 & 0 & \uparrow & 0 & \leftarrow \end{bmatrix} \begin{bmatrix} & & & & Z & & & \\ \uparrow & 0 & 0 & 0 & \uparrow & 0 & 0 & 0 \\ 0 & \uparrow & 0 & 0 & 0 & \uparrow & 0 & 0 \\ 0 & 0 & \uparrow & 0 & 0 & 0 & \uparrow & 0 \\ 0 & 0 & 0 & \uparrow & 0 & 0 & 0 & \uparrow \\ \uparrow & 0 & 0 & 0 & \downarrow & 0 & 0 & 0 \\ 0 & \uparrow & 0 & 0 & 0 & \downarrow & 0 & 0 \\ 0 & 0 & \uparrow & 0 & 0 & 0 & \downarrow & 0 \\ 0 & 0 & 0 & \uparrow & 0 & 0 & 0 & \downarrow \end{bmatrix}$$

$$= \begin{bmatrix} & & & & B & & & \\ \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow \\ \uparrow & \downarrow & \uparrow & \downarrow & \uparrow & \downarrow & \uparrow & \downarrow \\ \uparrow & \rightarrow & \downarrow & \leftarrow & \uparrow & \rightarrow & \downarrow & \leftarrow \\ \uparrow & \leftarrow & \downarrow & \rightarrow & \uparrow & \leftarrow & \downarrow & \rightarrow \\ \uparrow & \nearrow & \rightarrow & \searrow & \downarrow & \swarrow & \leftarrow & \nwarrow \\ \uparrow & \swarrow & \rightarrow & \nwarrow & \downarrow & \nwarrow & \leftarrow & \searrow \\ \uparrow & \searrow & \leftarrow & \nearrow & \downarrow & \nwarrow & \rightarrow & \swarrow \\ \uparrow & \nwarrow & \leftarrow & \swarrow & \downarrow & \searrow & \rightarrow & \nearrow \end{bmatrix} \tag{3.24}$$

Power Spectrum Density Estimation: Welch Method

Power Spectrum estimation (PSE) is one of the most important areas of research and applications in Digital Signal Processing (DSP). The autocorrelation function of a random signal is the appropriate statistical average that will use for the characterizing random signals in the time domain, and the power density spectral is the Fourier transform of the autocorrelation function, provides the transformation from the time domain to the frequency domain. A power spectrum describes the energy distribution of a time series in the frequency domain. Energy is a real-valued quantity, so the power spectrum does not contain phase information. Because a time series may contain non-periodic or asynchronously-sampled periodic signal components, the power spectrum of a time series typically is considered to be a continuous function of frequency.

There are two main types of PSE methods: Parametric and nonparametric. Parametric, or non-classical methods, find the parameters for a mathematical model describing a signal. Non-parametric, or classical methods, do not take any assumption on the data generating process [77].

In this project, the method used for PSE was the welch method, a non-parametrical method that consists in dividing the time series data into overlapping segments, computing a modified periodogram of each segment, and then averaging the Power Spectrum Density (PSD) estimates. Even though, overlapping segments might tend to introduce redundant information, this effect is diminished by the use of a non-rectangular window, which reduces the importance given to the end samples of segments, the overlapping samples.

Mathematically, the Welch method follows simple expressions. The periodogram

$P_k(\upsilon)$ follows eq. 3.25

$$\Psi(f) = \Psi^*(f) = \exp\left[-\left(\frac{\pi^2}{\sigma^2}\right)(f - f_0)\right] \tag{3.25}$$

where W, using the window function $w[m]$, is:

$$W = \sum_{m=0}^{M} w^2[m] \tag{3.26}$$

the welch estimate of the PSD can obtained:

$$S_x(\upsilon) = \frac{1}{K}\sum_{k=1}^{K} P_k(\upsilon) \tag{3.27}$$

where K are the number of segments or batches.

3.6 Envelope Analysis: Window Selection & Filtering

Envelope analysis has become one of the prominent vibration signal processing techniques for detection and diagnosis of rolling element bearing incipient failure. Developed during the early 1970's, this technique has had many designations with the term *Envelope Analysis* being the most popular. The basis for this analysis is the already refered concept of impulse generation whenever a fault develops and the REB is running [78]. The impacts generated by gearbox and rolling element faults superimpose upon the random and noise components of the signal, resulting in amplitude modulation as shown in Figure 3.8, consequently causing sidebands in the spectrum around the frequency bins associated with the vibration signal. The sidebands then mix with the frequency components of the vibration signal so that it is hard to distinguish them in the spectrum.

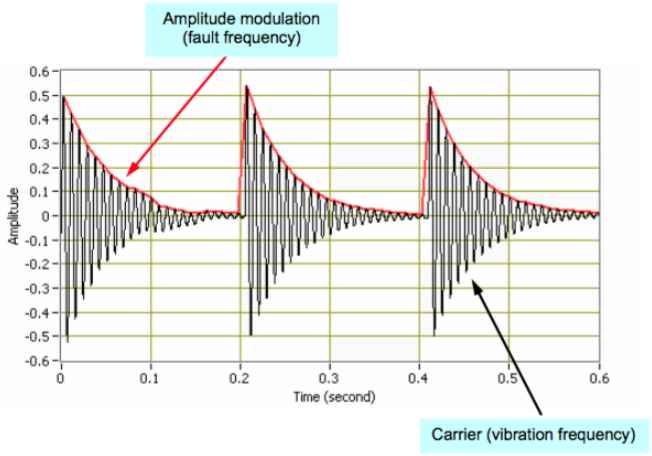


Figure 3.8: Vibration and fault frequencies caused by amplitude modulation.

Impacts in time domain generate many harmonics extending to very high frequency in frequency domain. Very often, some of these harmonics excite resonances in the structure. Exact location of the resonances is usually not known and cannot be determine easily. However, the resonance amplifies the modulating and carrier signals.

As was already stated, the spectrum of the pure signal, as in the immediate signal obtained from the transducer, contains poor information to assess the diagnosis of a bearing. Consequently, envelope analysis provides a mechanism for extracting out the periodic excitation or amplitude modulation of the resonance [78]. Such method consists of bandpass filtering the signal in a high frequency band where the fault impulses are amplified by structural resonances. The signal is then demodulated in amplitude resulting in the envelope signal, whose spectrum contains the desired diagnostic information in terms of both repetition frequency as well as modulation by the appropriate frequency at which the fault is passing through the load zone [13].

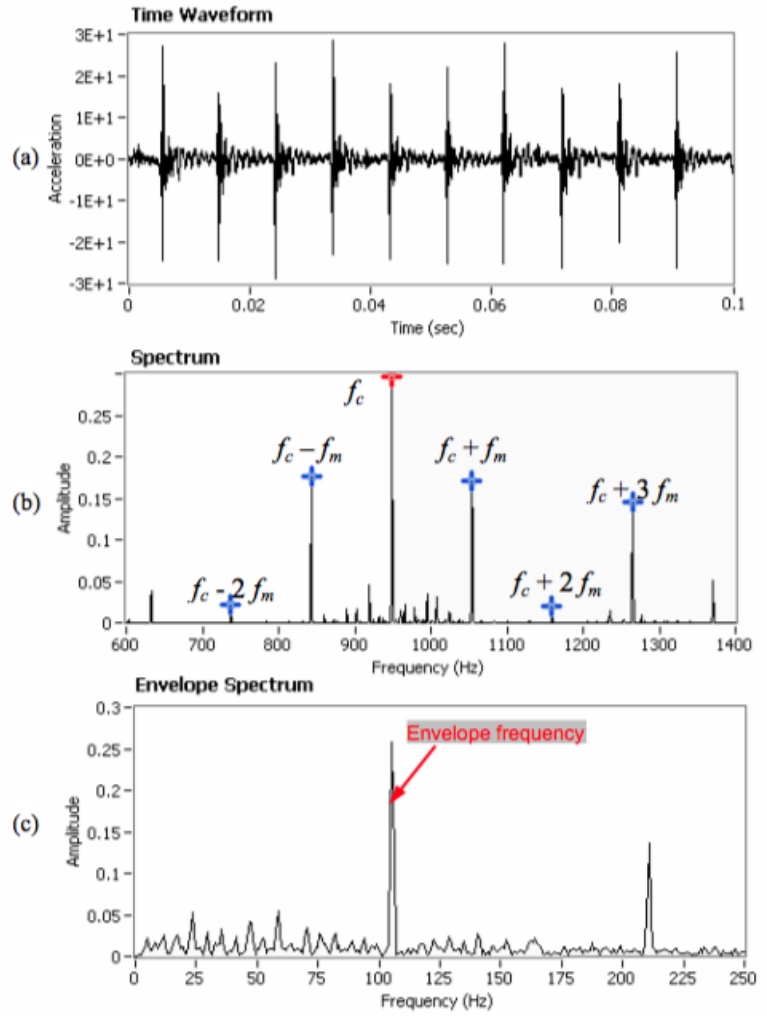


Figure 3.9: Envelope analysis applied to a signal. (a) Corresponds to the raw vibration signal, outer fault is present. (b) Carrier frequency and its sidebands. (c) Envelope frequency shows the BPFO.

However, one of the main setback of this analysis is how to determine the best frequency band, or *window*, around the theoretical fault frequency. Fixing a frequency band can be ineffective as fixed frequency bands may or may not encompass structural resonances, which are excited by the bearing damage. In this text, the cyclic spectral correlation was used to determine the central frequency to analyse and a procedure named power spectrum estimation to determine the band, further explained. The filtering of the signal is assured by the Butterworth bandpass filter of the 4^{th} order.

According to Randall [1] it is preferable to analyse the squared envelope signal rather than the envelope, due to the fact that mathematically, the envelope of a signal is the square root of the squared envelope and therefore a rectified signal is the square root of the squared signal. The square root operation introduces external components that were not in the original squared signal and that will cause masking of the desired information.

3.7 Proposed Semi-Automated Method

The aim of this project was to develop a semi-automated method for fault classification in rolling element bearings. Given that the classification methods run automatically, the only point that requires human input is for signal processing, more precisely, for the filtering operation.

In terms of signal processing techniques applied they are clarified in Figure 3.10.

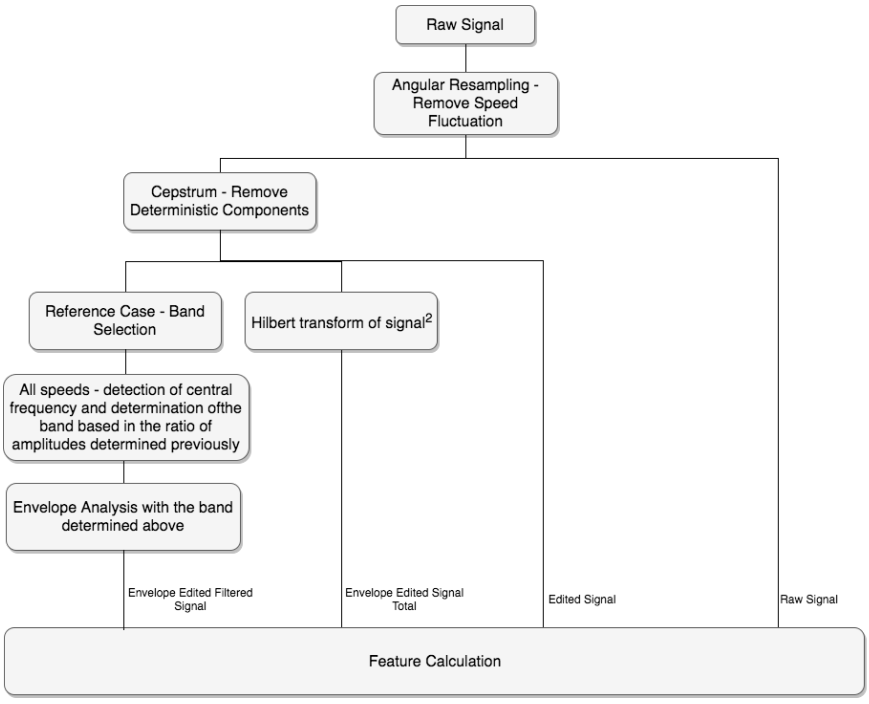


Figure 3.10: Proposed method for signal processing of vibration data.

The last process of this method is feature calculation.

Given that different frequency bands for filtering will result in different envelope

signals, the features calculated will be highly dependent on the envelope filtered signal and, consequently, of the filtering used. Further analysis on signals will focus on envelope filtered signal, from now on called Filtered Signal, and the Hilbert Demodulated signal without filtering, from now on called Non-Filtered Signal.

Because the filtering process is the most sensitive part of this process and the only one that needs human input, the method proposed tried to maintain the user input to a minimum, using methods and techniques proposed in the literature. The method proposed in this text is shown in Figure 3.11.

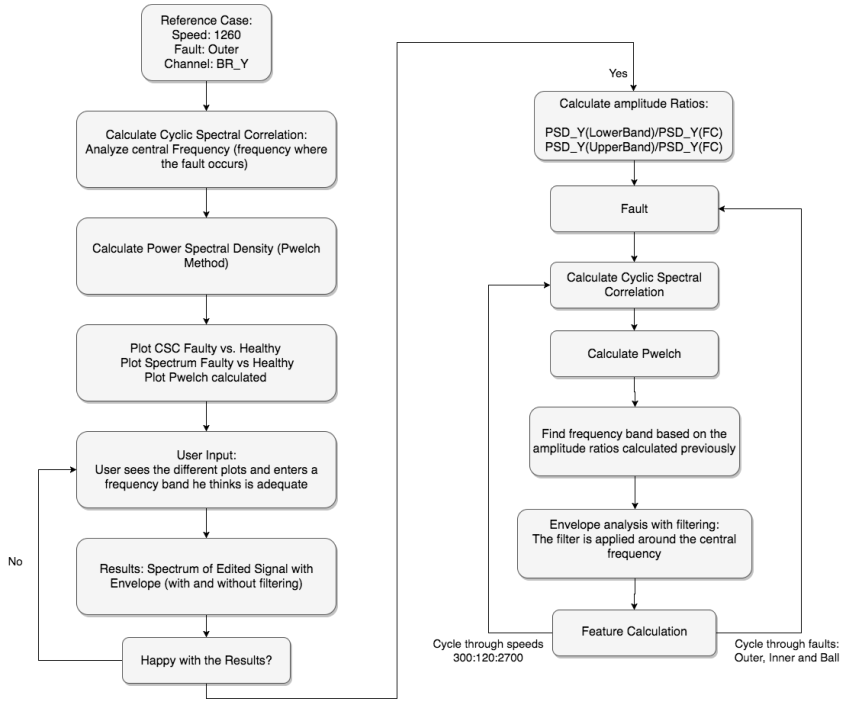


Figure 3.11: Proposed method for envelope analysis of vibration data.

As one can see from Figure 3.11 the CSC and PSD are calculated for the reference case, advised by SpectraQuest, of 1 260 rpm. After the CSC and PSD are calculated, these are plotted, along with a *healthy versus faulty* graph showing the spectrum of each raw vibrational signal. Using this information, the user can decide on an optimal frequency band, around the central frequency, to be used for filtering.

A new window will pop-up, showing the results of the filtered signal and if the user considers the results to be acceptable, the script will calculate amplitude ratios, on the PSD calculated curve of the reference case, and search these amplitude ratios on the PSD curves of all the other cases to be analysed.

The data used in this project was acquired using the SpectraQuest test-rig. Measured speeds ranged from 300 rpm to 2 700 rpm with an increment of 120 rpm. An example will be presented in chapter 6 (Figure 6.3).

CHAPTER 4

Feature Analysis

After the signal has been processed and enhanced, applying techniques that allows for diagnosis information to be extracted from it, it is then possible to develop an automated method for diagnosis of bearings. Rapid and reliable diagnosis is done through the analysis of indicators, or features, calculated on the signals resulting from the signal processing stage. Moreover, ultimate goal of pattern detection, important fo calculation of RUL, and its success will depend on the calculation of these features, whether they are indicative of a present fault or not. Furthermore, due to the large variations, direct comparison of vibration signatures is difficult, and thus statistical features provide an easier and automated form for application of pattern recognition and tracking technique. In theory, these features are more reliable, for this application than the analysis of the vibration signals themselves.

Moreover, the literature advises the calculation of several features such as the Kurtosis, Standard deviation, Crest factor and among others. However, no connection has been established on which features are best indicative of failure of a bearing. This text tries to do that, in a way, using feature selection methods that will be explained further in this section.

One finds different techniques for calculation of time-domain statistics. One of them calculates the features for the entire frequency range of the signal, while the other approach divides the signal into several frequency segments and performs feature calculation in each segment, resulting in various calculations for one single measured signal. This was the approach used for this project, where the vibrational signal was measured during 30 seconds. The length of the signal is then split into a total of 10 samples and the features are calculated on each of these samples. This procedure was repeated for all speeds (from 300 rpm to 2 700 rpm with increments of 120 rpm)

One can use the probability density function in order to calculate the statistical features of a signal. Each statistical feature will correspond to a moment order, which correspond to mechanical moments about the centroid of a plane [4]. Davies states in [4] that the odd moments are related to the position of the peak density, similar to the median value, while the even moments are informative of the spread in the distribution. One technique that is widely used is to normalize the moments higher than the second order, removing the mean and dividing by the standard deviation raised to the order of the moment. The same author, Davies in [4] adds that higher moments are sensitive to impulsiveness in the signal which make moments such as the

Skewness and Kurtosis, respectively third and fourth moments, relevant to the diagnosis analysis.

4.1 Feature Extraction

This subsection will be an overview of the statistical features calculated on each of the samples of the total signal. Every feature described will be explained relatively to the probability density function $p(x)$ and the overview will follow the character of increasing moment order.

4.1.1 Mean Value

The mean value, a moment of first order of the probability function, is defined as, for a symmetrical function as the line of symmetry of such function. Also known as the expected value of a random variable that follows the probability function, the mean translates as the arithmetic average value of the same random variable. The mean, follows the equation 4.1.

$$\mu = \int_{-\infty}^{\infty} x \cdot p(x)\ dx \tag{4.1}$$

4.1.2 Variance

The second-order moment of the probability density function is the variance. The strict definition is *"The average of the squared differences from the mean"*, in other words, the variance translates as the variability, or spread, of the data being studied. A value of 0 means that there is no spread on the data. Mathematically, following the strict the definition, this feature is expressed by equation 4.2.

$$\sigma^2 = \int_{-\infty}^{\infty} [x - \mu]^2\ p(x)dx \tag{4.2}$$

4.1.3 Skewness

The following feature, third-order, is the skewness. As its name states, it translated into how *skewed* or asymmetry the statistical distribution is relatively to the mean value. skewness can be quantified to define the extent to which a distribution differs from a normal distribution. In a normal distribution, the graph appears as a classical, symmetrical "bell-shaped curve." The mean, or average, and the mode, or maximum point on the curve, are equal. For this distribution the skewness returns a value of zero.

When a distribution is skewed to the left, the tail on the curve's left-hand side is longer than the tail on the right-hand side, and the mean is less than the mode. This situation is also called negative skewness. Oppositely, when a distribution is skewed to the right, the tail on the curve's right-hand side is longer than the tail on the left-hand side, and the mean is greater than the mode. This situation is also called positive skewness. This can be seen of figure 4.1.

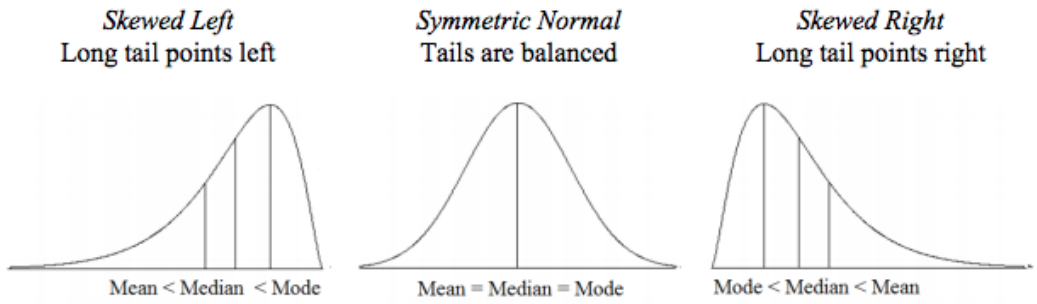


Figure 4.1: Types of Skewness [79]

According to [53], when related to vibrations, the higher asymmetry means higher irregularities in the signal and, consequently the presence of defects in the machine. Because it is a quantity of third-order moment of the probability density function, this power of three enhances the high values and suppresses the low values in the signal. Given this, this parameter is more useful for low-speed machinery. Skewness is expressed mathematically by the equation 4.3.

$$S = \frac{\int_{-\infty}^{\infty}[x-\mu]^3 \; p(x)dx}{\sigma^3} \tag{4.3}$$

4.1.4 Kurtosis

The kurtosis is a fourth moment of the probability distribution function, used to measure the impulsiveness of a signal. Similarly to the skewness, this parameter is a descriptor of the shape of a probability distribution function. Its value is very high for impulsive signals because the power of four in the equation gives more importance to local spikes, eq. 4.4.

$$K = \frac{\int_{-\infty}^{\infty}[x-\mu]^4 \; p(x)dx}{\sigma^4} \tag{4.4}$$

Depending on its value, the distribution can be defined as (figure 4.2):

- Mesokurtic: $K = 3$,
- Leptokurtic: $K > 3$,
- Platykurtic: $K < 3$

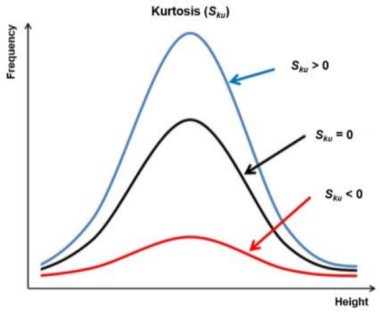


Figure 4.2: Types of distribution, according to values of kurtosis [80].

4.1.5 Root Mean Square

Strictly speaking, the RMS is the square root of the arithmetic mean of the squares of the values, or the square of the function that defines the continuous function. In other words, the RMS is a measure of the magnitude of a distribution, representing the overall magnitude. For a vector x with N elements, its value is given by equation (4.5).

$$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N} {x_n}^2} \tag{4.5}$$

Stamboliska recommends that isolated diagnosis based solely on the value of the RMS should not be performed, mainly because there is not a specific connection between deviations in the RMS value and the presence, or not, of a fault in a machine.

4.1.6 Peak-to-Peak

Peak-to-Peak (P2P) is the difference between the maximum positive and the maximum negative amplitudes of a waveform, or signal. Given that vibration signals are symmetric by nature, a higher value of P2P indicate higher values of amplitude in vibration signals, consequently meaning an increase in the impulsive level of the fault. This indicator follows a simple equation 4.6.

$$P_2P = x_{max} - x_{min} \tag{4.6}$$

4.1.7 75% Percentile

A percentile specifies the amount of dispersion in a specific interval, P_{75} and P_{25} refers to the number of points above 75% and bellow 25% of the mean, correspondingly.

$$P_{75} = p\ (x < x_{0,75}) \tag{4.7}$$

4.1.8 Crest Factor

Crest Factor (CF) a measure of a waveform, showing the ratio of peak values to the average value [53], eq. 4.8. In other words, CF indicates how extreme the peaks are in a waveform. It is a quick and useful calculation that gives an idea of how much impact is occurring in a time wave.

$$CF = \frac{|Peak|}{RMS} \tag{4.8}$$

The CF is an useful information that is lost if one is only viewing a spectrum, this because the FFT cannot differentiate between impacting and random noise. Impacting in a time waveform may indicate rolling element bearing wear, gear tooth wear, or cavitation. As for other machinery, also for low-speed machinery, the crest factor should be trended over time in order to see if the amount of impacting is increasing or not, just that the absolute values will be again of lesser amount comparing to higher speed machines [53].

4.1.9 Entropy and Wavelet Entropy

The entropy, eq. 4.9 is a quantitative measure of a system's randomness, or disorder. It is expected the value of the entropy to increase with the increase with the increase of the random character of a signal. The parameters used for calculation of entropy in this project were the Logarithmic Entropy ($\boldsymbol{Went_{Log}}$), eq. 4.11 and the Shannon Entropy ($\boldsymbol{Went_{Sha}}$), 4.10.

$$Entropy = -\sum_{n=1}^{N} p(x_n) \cdot \log_2 p(x_n) \tag{4.9}$$

$$Went_{Sha} = -\sum_{n=1}^{N} p(x_n)^2 \cdot \log p(x_n)^2 \tag{4.10}$$

$$Went_{Log} = -\sum_{n=1}^{N} \log p(x_n)^2 \tag{4.11}$$

4.1.10 Impulse, Margin and Shape Factors

According to [5] Chun Qing Li developed a new approach using other non-dimensional parameters. The parameters studied were the Impulse Factor (IF), Margin Factor (MF) or clearance factor and another one called the Shape Factor (SF). These three parameters were found to be useful under simulation conditions using a

Gaussian probability density function as a model of fatigue spalling on bearings.

$$IF = \frac{x_{Max}}{\frac{1}{N}\sum_{n=1}^{N}|x(n)|} \tag{4.12}$$

$$MF = \frac{x_{Max}}{\frac{1}{N}\left(\sum_{n=1}^{N}\sqrt{|x(n)|}\right)^2} \tag{4.13}$$

$$SF = \frac{RMS}{\frac{1}{N}\sum_{n=1}^{N}|x(n)|} \tag{4.14}$$

According to the same author, the margin factor is seen as the most sensitive and robust between the three to detection of incipient spalling.

4.1.11 Fourth Order Figure of Merit

Fourth Order Figure of Merit (FM4) was developed to detect changes in the vibration pattern resulting from damage from gears, for example on a limited number of gear teeth. [6] FM4 is calculated by applying the fourth normalized statistical moment to the difference signal, eq. 4.15.

$$FM4 = \frac{N\sum_{i=1}^{N}(d_i - \bar{d})^4}{(\sum_{i=1}^{N}(d_i - \bar{d})^2)^2} \tag{4.15}$$

4.1.12 M6A and M8A

These two indicators were developed to detect surface damage on machinery components, they are the 6th and 8th statistical moments of the signal, as opposed to the 4th statistical moment, as the Kurtosis or FM4. Given the higher order it is expected to be more sensitive to peaks.

$$\begin{aligned} M6A &= \frac{N^2\sum_{i=1}^{N}(x_i - \bar{x})^6}{(\sum_{i=1}^{N}(x_i - \bar{x})^2)^3} \\ M8A &= \frac{N^3\sum_{i=1}^{N}(x_i - \bar{x})^8}{(\sum_{i=1}^{N}(x_i - \bar{x})^2)^4} \end{aligned} \tag{4.16}$$

4.1.13 Harmonics

On frequency domain, the amplitudes of the first three harmonics of each fault were studied. Using the theoretical frequencies shown in chapter 2 and a searching threshold, the amplitude values were searched around the characteristic frequencies, within the threshold. The need for the threshold value is because the harmonics may shift from the theoretical values due to the variations in speed.

4.2 Feature Reduction: PCA and LDA

At this point, one important note has to be made. For this project, there were *two stages*, so to say. The first one regarded signal pre-processing and handling of data, which was performed in Matlab. The second stage, of feature selection and classification, was performed with the scripting language Python. The reason for this change was that the tools available for machine learning and feature selection are readily, and natively, available when downloading the tools needed for using Python.

4.2.1 Principal Component Analysis

To classify all the classes, or faults, and to obtain more definitive conclusions towards developing a diagnostic tool, it is important to present through specific plots, some form of distinction between the data-points belonging to each class.
PCA is a form of identifying patterns in data, and expressing the data in a way that highlights their similarities and their differences. Since patterns in data can be hard to find in high-dimensional data, where the luxury of graphical representation is not available, PCA is a powerful tool for analysing data.

The other main advantage of PCA is that once these patterns in the data are found, the data is compressed, by reducing the number of dimensions, while trying to minimize the loss of information.

In other words, PCA is a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called PC. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining uncorrelated data as possible. PCA is similar to another multivariate procedure called Factor Analysis.

PCA is a projection statistical method used for dimensionality reduction. This analysis produces a lower-dimensional representation that preserves the correlation structure between the variables [81]. Assuming a set of n observations or samples and m variables, stacked into a matrix X, whose variance-covariance matrix has eigenvalues λ and eigenvectors p verify condition (4.17)

$$(\lambda_1, p_1), (\lambda_2, p_2), \cdots, (\lambda_m, p_m) \qquad \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_m \geq 0 \tag{4.17}$$

Then the principal decomposition component of X can be expressed as:

$$X = TP^T + E \; = \; \sum_{i=1}^{l} t_i p_i^T + E \tag{4.18}$$

Where $T = [t_1, t_2, \cdots, t_l]$ is the matrix of principal component scores, $P = [p_1, p_2, \cdots, p_l]$ is the matrix of principal component loadings and E is the residual matrix in the sense of the minimum Euclidean norm and l is the index of the PC [82]. The identification of the PCA model consists in estimating its parameter by an eigenvalue-eigenvector decomposition and determining the number of components l to retain. In order to keep enough information about the data, an approach to select the number of components to retain is the experimental method, which judges the

cumulative sum contribution of the anterior l component is higher than 0,85 as stated by equation (4.19) [82].

$$100 \cdot \frac{\sum_{i=1}^{l} \lambda_i}{\sum_{i=1}^{m} \lambda_i} > 85\% \tag{4.19}$$

However, it must be noted that performing data normalization before applying PCA is important. Since PCA is a variance maximizing exercise that projects the original data onto directions that maximize the variance. Given that this project uses variables whose numbers are not in the same scale, then bigger numbers will correspond to bigger variances. Hence, normalizing the dataset, as in restructuring the dataset so that all variables vary between the same extreme values, is of utmost performance so that no features have more importance than others and PCA runs correctly.

4.2.2 Linear Discriminant Analysis

Suppose we are given a learning set L of multivariate observations (i.e., input values in R^r), and suppose each observation is known to have come from one of K predefined classes having similar characteristics. These classes may be identified, for example, as species of plants, different types of tumours, or condition of a given machine component. To distinguish the known classes from each other, we associate a unique class label (or output value) with each class; the observations are then described as labeled observations [83]. In the case of this project, classes were *Outer, Inner, Ball, Healthy* as in faulty (Outer race fault, Inner race fault, ball fault) or the healthy case. Izenman defines [83] discrimination analysis as *"the use of information in a learning set of labeled observations to construct a classifier, or classification rule, that will separate the predefined classes as much as possible"*.

Initially proposed by R. Fisher for discriminating between different types of flowers, hence the designation Fisher linear discriminant , LDA has aim of determining a subspace of lower dimension, compared to the original data sample dimension, in which the data points of the original problem are separable [84].

According to [85] LDA uses the within-class scatter matrix S_w, to evaluate the compactness within each class and the between-class scatter matrix, i.e, S_b, to evaluate the separability of different classes [85]. Let $X = x_1, x_2, \cdots, x_l \in R^{D \times l}$ be the training set, where each x_i belongs to a class $c_i = \{1, 2, \cdots, c\}$. l_i is the number of data points in the ith class and l corresponds to the number of data points in all classes.

Then, the between-class scatter matrix, i.e., S_b, the within-class scatter matrix,

i.e., S_w, and the total-class scatter matrix, i.e., S_t, are defined as

$$S_b = \sum_{i=1}^{c} l_i \left(\mu_i - \mu\right)\left(\mu_i - \mu\right)^T \tag{4.20}$$

$$S_w = \sum_{i=1}^{c} \sum_{x_i \in c_i} \left(x_i - \mu_i\right)\left(x_i - \mu_i\right)^T \tag{4.21}$$

$$S_t = \sum_{i=1}^{l} \left(x_i - \mu\right)\left(x_i - \mu\right)^T \tag{4.22}$$

$$\tag{4.23}$$

where $\mu_i = 1/l_i \sum_{x_i \in c_i} x_i$ is the mean of the data points in the ith class, and $\mu = 1/l \sum_{i=1}^{l} x_i$ is the mean of the data points in all classes.

One can easily see that this method performs this linear decomposition using labeled information, making it a supervised method, whereas PCA is an un-supervised method since it doesn't require labeled data. Figure 4.3 shows the difference between PCA and LDA.

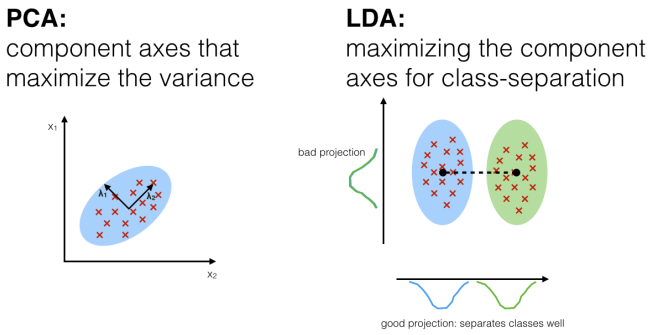


Figure 4.3: Difference between PCA and LDA [86].

4.3 Feature Selection

As already stated in chapter 1, feature selection can help in increasing the classifier's performance and eliminate redundant features. Isabelle states in [87] that feature selection can have other motivations, such as:

- General data reduction, to limit storage requirements and increase algorithm speed;
- Feature set reduction, to save resources in the next round of data collection or during utilization;
- Performance improvement, to gain in predictive accuracy;
- Data understanding, to gain knowledge about the process that generated the data or simply visualize the data.

Moreover, Amarnath proved in his analysis [88] that using all of the features considered in his study, may not be significant for the classification purpose. Using more

irrelevant features actually may reduce the performance of the classification algorithm and increase the computational resources required [88].

Given that the literature advises to perform feature selection methods prior to classification, two feature methods are proposed and clarified in the following chapter.

4.3.1 mRMR: minimum-Redundancy Maximum-Relevance

Feature selection is an important problem for pattern classification systems. However, in feature selection, it has been recognized that the combinations of individually good features do not necessarily lead to good classification performance. In other words, "the m best features are not the best m features" [89].

In many pattern recognition applications, identifying the most characterizing features of the observed data, i.e., feature selection, is critical to minimize the classification error [89]. The optimal characterization condition often means the minimal classification error. In an unsupervised situation where the classifiers are not specified, minimal error usually requires the maximal statistical dependency of the target class c on the data distribution in the subspace R^m, and vice versa. This scheme is nominated maximal dependency (Max-Dependency).

One of the most popular approaches to realize Max-Dependency is maximal relevance (Max-Relevance) feature selection: selecting the features with the highest relevance to the target class c. Relevance is usually characterized in terms of correlation or mutual information, of which the latter is one of the widely used measures to define dependency of variables [88].

An important concept, that serves as a basis to this concept is the mutual information concept. Assuming two random variables x and y, their mutual information is defined in terms of their probabilistic density functions $p(x)$, $p(y)$ and $p(x, y)$. The mutual information between these two variables follows the expression 4.24:

$$I(x;y) = \int\int p(x,y) log \frac{p\,(x,y)}{p(x)p(y)} \mathrm{d}x\mathrm{d}y \tag{4.24}$$

Max-Dependency constraint

In terms of mutual information, the purpose of feature selection according to this criteria, is to find a feature set S with m features x_i, which jointly have the largest dependency on the target class c. This scheme, called Max-Dependency, follows the following expression:

$$\max D(S,c), \quad D = I(x_i, i = 1, \cdots, m; c) \tag{4.25}$$

Obviously, when $m = 1$, the solution is the feature that maximizes $I(x_j; c)(1 \leqslant j \leqslant M)$. When $m > 1$, a simple incremental search scheme is to add one feature at one time. However, in spite of its theoretical value, Max-Dependency is often hard to get an accurate estimation for multivariate density, due to two main difficulties:

1. the number of samples is often insufficient;

2. the multivariate density estimation often involves computing the inverse of the high-dimensional covariance matrix.

Consequently, even though Max-Dependency feature selection might be useful to select a very small number of features when the number of samples is high, it is not appropriate for applications where the aim is to achieve high classification accuracy with a reasonably compact set of features [88].

Max-Relevance and Min-Redundancy

As seen previously, the Max-Dependecy criteria is hard to implement. An alternative is to apply the criteria of Max-Relevance, where feature are selected based on maximum relevance. This criteria searches features that satisfy 4.26 which aproximates $D(S;c)$ in 4.25 with the mean value of all mutual information values between individual feature x_i and class c. One can rewrite 4.25 as 4.26:

$$\max D(S,c), \quad D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i;c) \tag{4.26}$$

It is likely that features selected according to 4.26 might be highly dependant between each other. When two features highly depend on each other, the respective class-discriminative power would not change much if one of them were removed. Therefore, the following minimal redundancy (Min-Redundancy) condition can be added to select mutually exclusive features [88]:

$$\min R(S), \quad R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i, x_j) \tag{4.27}$$

The criterion that combines both 4.27 and 4.26 is called "minimal-redundancy maximal-relevance" (mRMR) [88]. The operator $\Phi(D;R)$ combines D and R and the method follows the eq. 4.28

$$\max \Phi(D;R), \quad \Phi = D - R \tag{4.28}$$

However, the most important part of this method is how it calculates the mutual information between the variables:

$$I(S_m;c) = H(c) \; + \; H(S_{m-1}, x_m) \; - H(S_{m-1}, x_m, c) \tag{4.29}$$

where $H(\cdot)$ is the entropy (see Information Gain Criteria in section 4.3.2) of the respective multivariate variables, S_{m-1} is the set of $m-1$ features, x_m is the mth feature being evaluated and c is the class.

Data Binning

Further in the analysis performed, it was noted that data discretisation was necessary given that the method was not working because the entropy function implemented within the method wasn't working well with the continuous values available in our data. Accordingly, the technique for data discretisation adopted was the binning technique.

Generally, the distribution of the observations changes with time. Binning the data, i.e., grouping observations into time intervals, leads to an approximation of this distribution by a piecewise-constant distribution, constant in each time interval.

The choice of the set of bins is crucial, as binning will always lead to a certain distortion between the true and estimated distributions. A binning strategy should be "good", in the following senses:

- For a given number of bins, the locations of the bin edges must be chosen so as to minimize heterogeneity of the data in each bin.
- The number of bins must be carefully chosen, i.e., require sa good tradeoff between a large number of bins and a large number of observations in each bin. The true distribution can be accurately approximated by a piecewise-constant distribution with a large number of bins, while a large number of observations in each bin is required to accurately estimate this true distribution.

There are various ways to implement binning. The two simplest are:

- Equal-width binning: K bins of length $(tmax - t_{min})/K$.
- Equal-size binning: K bins, each with n/K data points. If n is not a multiple of K, we can correct so that each bin has either $[n/K]$ or $[n/K] + 1$ data points.

4.3.2 Decision Tree

Generally, the DT is a classification method. A standard tree consists of a number of branches, one root, a number of nodes and a number of leaves However, given that it is possible to export the DT and visually interpret the model, one can use the classification method as a feature selection method.
Because of the easy interpretability of this method, some authors have used this classifier as a feature selection method using the results as input for other classifiers such as SVM, Proximal Support Vector Machines (PSVM), Artificial Neural Networks (ANN) and others [90, 91].

The use of the DT as a feature selection method is due to the fact that this classification method has a feature selection algorithm built-in. As such, using this classification method, one can train the model without performing feature selection before-hand, because the method will perform feature selection for itself. However, one must consider that using classification methods such as SVM, PSVM or ANN one is using black-box models that cannot be interpreted afterwards.

This section will only describe how feature selection is performed within the Decision Tree itself. The classification method will be clarified in chapter 5.
Using the tools available within the scripting language Python, DT uses two criterias for spliting the data, or choosing features: The Gini index, or the entropy information gain.

The Gini index (Gini)

Gini impurity is a measure of misclassification, which applies in a multiclass classifier context. Developed by Breiman et al. in 1984, the Gini function measures the impurity of an attribute with respect to the classes [92].

$$i(t) = 1 - \sum_{i=1}^{k} \left(p(c_i|t)\right)^2 \tag{4.30}$$

In a node t, an impurity function based on the Gini Index criterion assigns a training example to a class c_i with the probability $p(c_i|t)$. The estimated probability

that the item is actually in class j is $p(c_j|t)$. Therefore, the estimated probability of misclassification under this rule is 4.31

$$i(t) = 1 - \sum_{j=1}^{k} (p(c_j|t))^2 \tag{4.31}$$

The Gini Index criterion selects a test that maximizes the function 4.32:

$$gini(T) = 1 - \sum_{i=1}^{k} (p(c_i))^2 - \sum_{i=1}^{n} p(t_i) \sum_{j=1}^{k} (p(c_j|t_i)(1 - p(c_j|t_i)) \tag{4.32}$$

A Gini score gives an idea of how good a split is by how mixed the classes are in the two groups created by the split. A perfect separation results in a Gini score of 0, whereas the worst case split that results in 50/50 classes in each group result in a Gini score of 0,5 (for a 2 class problem).

Information Gain Criteria

The Information Gain function has its origin in information theory. It is based on the notion of entropy, which characterizes the impurity of an arbitrary set of examples [92]. If one emits a message stating that an example was chosen from a set, belonging to the class c_i, then the probability of this message is equal to $p(c_i)$ and the amount of information it conveys is $log_2(p(c_i))$. The expected information provided by a message with respect to the class membership can be expressed as:

$$info(\mathcal{L}) = -\sum_{i=1}^{k} p(c_i) log_2(p(c_i)) \tag{4.33}$$

where the quantity $info(\mathcal{L})$ measures the average amount of information needed to identify the class of an example in $\mathcal{L}$. This quantity is also kown as the entropy of the set $\mathcal{L}$ relative to the k-wise classification. The reason for the logarithm in base 2 is because the entropy is a measure of the expected encoding length measured in bits. Within a test T, the expected information requirement is the weighted sum over the subsets:

$$info_T(\mathcal{L}) = -\sum_{i=1}^{k} p(t_i) info(T_i) \tag{4.34}$$

The information gained by partitioning $\mathcal{L}$ in accordance to the test T is measured by the quantity 4.35

$$gain(t) = info(\mathcal{L}) - info_T(\mathcal{L}) \tag{4.35}$$

The information Gain criterion selects a test that maximizes the Information Gain function, or eq. 4.35.

Figure 4.4 compares compares the values of the impurity measures for binary classification problems. On this image, p refers to the fraction of records that belong to one of the two classes. One can see that all three measures attain their maximum value when the class distribution is uniform (i.e., when $p = 0,5$). The minimum values

for the measures are attained when all the records belong to the same class (i.e., when p equals 0 or 1).

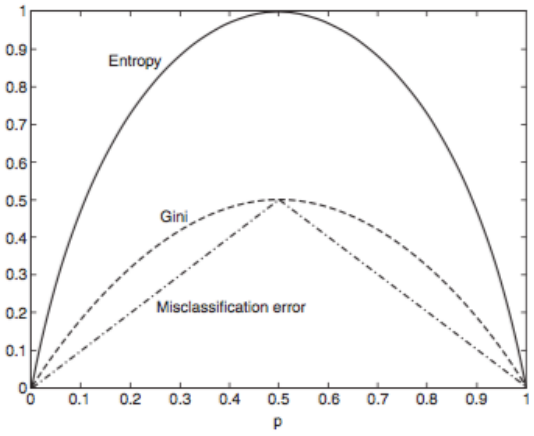


Figure 4.4: Comparison of impurity measures for binary classification problems [93].

CHAPTER 5

Classification Methods

In machine learning and statistics, classification is the problem of assigning a set of categories to a certain class of observation, on the basis of a training set of data containing observations whose category membership is known. An example would be assigning a given email into "spam" or "non-spam" classes or assessing the condition of a bearing based on labeled data (faulty or healthy). Classification is an example of pattern recognition.

Before going further into the classification methods used, common concepts must be clarified:

- **Classification**: The task of learning a target function f that maps each attribute set x to one of the predefined class labels y. The target function is also informally denoted as a classification model.
- **Descriptive Modelling**: A classification model can serve as an explanatory tool to distinguish between objects of different classes.
- **Predictive Modelling**: A classification model can also be used to predict the class label of unknown records

A key characteristic that distinguishes classification from regression is the class label. If a class label y is a continuous attribute then the problem is concerned about regression, on the other hand if y is discrete then the learning problem is a classification one [93].

The general approach to a classification problem is to use a classifier to build classification models from an input data set. These classifiers can be DT, rule-based, Neural Networks, Support Vector Machines or Naïve Bayes classifiers, among others. Each technique employs a learning algorithm to identify a model that best fits the relationship between the attribute set and class label of the input data. The model generated by a learning algorithm should both fit the input data well and correctly predict the class labels of records it has never seen before. Therefore, a key objective of the learning algorithm is to build models with good generalization capability: models that accurately predict the class labels of previously unknown records [93].

The errors committed by a classification model are generally divided into two types: training errors and generalization errors. Training error, also known as re-substitution error or apparent error, is the number of misclassification errors committed

on training records, whereas generalization error is the expected error of the model on previously unseen records. As already stated, a good classification model must not only fit the training data well, it must also accurately classify records it has never seen before. In other words, a good model must have low training error as well as low generalization error. This is important because a model that fits the training data too well can have a poorer generalization error than a model with a higher training error. Such a situation is known as model **overfitting** and should be avoided otherwise it can render the results as not reliable [93].

5.1 Decision Tree Analysis

A DT, a *tree-based* knowledge representation methodology, is used to represent classification rules. A standard tree usually consists of a number of *branches*, one *root*, a number of *nodes* and a number of *leaves*. One branch is a chain of nodes from root to a leaf; and each node involves one attribute [93]. The occurrence of an attribute in a tree provides the information about the importance of the associated attribute. In other words, one can say that a Decision Tree has three types of nodes:

- A **root node**, the beginning of the Decision Tree itself or its *root*. A root node has no incoming ramifications or *edges*.
- **Internal nodes**, each of which has exactly one incoming edge and two or more outgoing edges.
- **Leaf** or **terminal nodes**, each of which has exactly one incoming edge and no outgoing edges.

In a DT, each leaf node corresponds to a class label. The non-terminal nodes, which include the root and other internal nodes, contain attribute test conditions to separate records that have different characteristics. An illustration of this concept is shown in Figure 5.1

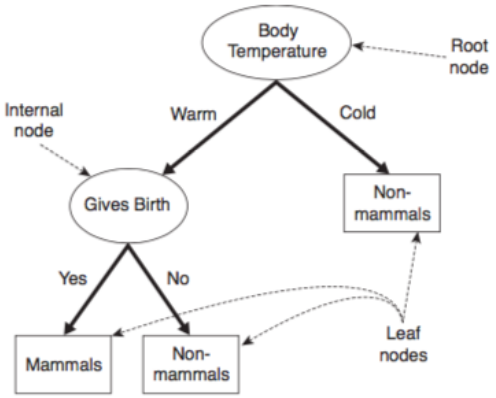


Figure 5.1: A DT modelling the concept of mammal classification [93].

Classifying a specific sample is simple once a Decision Tree has been constructed. Starting from the root node, we apply the test condition to the record and follow the

appropriate branch based on the outcome of the test. This will lead either to another internal node, for which a new test condition is applied, or to a leaf node. The class label associated with the leaf node is then assigned to the sample, classifying the sample in question.

Different algorithms exist to construct decision trees: CART, C4.5, C5.0 or ID3. According to the documentation available online[1] of the DT function available on the *scikit-learn* library for machine learning in Python, there are differences between the algorithms.

ID3, or Iterative Dichotomiser 3, was developed in 1986 by Ross Quinlan. The algorithm creates a multiway tree, finding for each node, through a greedy manner, the categorical feature that will yield the largest information gain for categorical targets. Trees are grown to their maximum size and then *pruned*[2] to improve the ability of the tree to generalise to unseen data.

C4.5 is the successor to ID3 and removed the restriction that features must be categorical by dynamically defining a discrete attribute, based on numerical variables, that partitions the continuous attribute value into a discrete set of intervals. C4.5 converts the trained trees, i.e. the output of the ID3 algorithm, into sets of if-then rules. The accuracy of each rule is then evaluated to determine the order in which they should be applied. Pruning is done by removing a rule's precondition if the accuracy of the rule improves without it.

C5.0 is Ross Quinlan's latest version released under a proprietary license. It uses less memory and builds smaller rulesets than C4.5 while being more accurate.

CART (Classification and Regression Trees) is very similar to C4.5, but it differs in that it supports numerical target variables (regression) and does not compute rule sets. CART constructs binary trees using the feature and threshold that yield the largest information gain at each node. The algorithm used in *scikit-learn* is an optimised version of the CART algorithm.

According to the same source, DTs have advantages and disadvantages. The main advantages are the following:

- Simple to understand and to interpret, as trees can be visualised. Making it a white box model (contrary to a black-box model).
- Requires little data preparation. Other techniques often require data normalisation, dummy variables need to be created and blank values to be removed.
- The cost of using the tree, i.e., predicting data, is logarithmic in the number of data points used to train the tree.
- Able to handle both numerical and categorical data.
- Able to handle multi-output problems.
- Possible to validate a model using statistical tests. That makes it possible to account for the reliability of the model.

[1] http://scikit-learn.org/stable/modules/tree.html

[2] pruning is to trim a tree, shrub, or bush by cutting away dead or overgrown branches or stems, especially to encourage growth. In a DT, this term has the meaning of trimming the tree in order to avoid overfitting

- Performs well even if its assumptions are somewhat violated by the true model from which the data were generated.

However, it also presents the following disadvantages:

- Decision-tree learners can create over-complex trees that do not generalise the data well (recall overfitting concept). Mechanisms such as pruning, setting the minimum number of samples required at a leaf node or setting the maximum depth of the tree are necessary to avoid this.
- Decision trees can be unstable because small variations in the data might result in a completely different tree being generated. This problem is mitigated by using a group of decision trees (also known as Random Forest method, to be explained later).
- The problem of learning an optimal Decision Tree is known to be NP-complete under several aspects of optimality and even for simple concepts. Consequently, practical decision-tree learning algorithms are based on heuristic algorithms such as the greedy algorithm where locally optimal decisions are made at each node. Such algorithms cannot guarantee to return the globally optimal Decision Tree. This can be mitigated by training multiple trees in an ensemble learner, where the features and samples are randomly sampled with replacement.
- Decision tree learners create biased trees if some classes dominate. It is therefore recommended to balance the dataset prior to fitting with the Decision Tree. Also known as normalizing the data.
- Decision Trees can only do linear splits on the data, while other methods allow non linear splits.

Given that the Decision Tree classifier is prone to overfitting, *pruning* was performed on the model created for each signal. This step consists of dividing the total data into two sets: one for learning and creating the model and verify with which tree depth the classification was best, and another for effective test and validation.

5.2 Random Forest

As seen in the previous section, DTs have their disabilities and setbacks. Breiman, in [94], came up with the idea of using a large group of DTs as a method in order to overcome the disadvantages of single DTs. The technique called Bootstrap Aggregating, or *bagging*, which is a common machine learning ensemble meta-algorithm designed to improve the stability and accuracy of machine learning algorithms used in statistical classification and regression. It also reduces variance and helps to avoid overfitting. Hence, using a large group of DTs is essentially the same as DT *bagging*.The technique is named as Random Forest and consists, basically, as using DTs as a group.

This method's algorithm creates multiple CART-like trees each trained on a bootstrapped sample of the original training data, and searches only across a randomly selected subset of the input variables to determine a split, for each node. For classification, each tree in the Random Forest casts a unit vote for the most

popular class at input x. The output of the classifier is determined by a majority vote of the trees. Additionally, the trees in Random Forests are not pruned, further reducing the computational load

As a result, the Random Forest algorithm can handle high dimensional data and use a large number of trees in the ensemble. This combined with the fact that the random selection of variables for a split seeks to minimize the correlation between the trees in the ensemble, results in error rates that have been compared to much more complex methods, SVM for instance, while being computationally much lighter. As each tree is only using a portion of the input variables in a Random Forest, the algorithm is considerably lighter than conventional bagging with a comparable tree-type classifier [95]. This method presents multiple advantages:

- Can be used when there are many more variables than observations.
- Can be used both for two-class and multi-class problems of more than two classes.
- Has good predictive performance even when most predictive variables are noise, and therefore it does not require a pre-selection of features
- Does not overfit.
- Can handle a mixture of categorical and continuous predictors.
- Incorporates interactions among predictor variables.
- The output is invariant to monotone transformations of the predictors.
- There is little need to fine-tune parameters to achieve excellent performance.

However, relatively to our project, as stated by [94, 95], the number of trees generated in the group can be quite large as the authors refer that no less than 100 trees are used, rendering this method to be similar to a black-box method. For bearing classification, and for sakes of critical thinking, one expects to see a *map* that leads to why such classification was performed, or in other words, such decision was made (faulty versus non-faulty). Hence, analysing a wide range of *maps* can render the tool as not indicative, or interpretable, as to why such bearing has been classified as faulty or as healthy.

In spite of this, it is expected that this method outperforms the conventional DT [94] and it was still used in this project allowing further conclusions. This method was also available in the *scikit-learn* library of Python scripting language, allowing the use of Python for applying this method.

5.3 Support Vector Machines

An extensive explanation and mathematical formulation can be found in data science literature or in speciality books on SVM [96]. However, a concise and brief explanation on SVM will be given through this section. A first look on a Web Page is recommended [3] in order to have a first idea on how SVM work.

SVM is an excellent algorithm used for classification and regression. This

[3]Link to Reddit: `https://goo.gl/MckZDC`

computational learning method is based on the statistical learning theory developed by Vapnik. SVMs are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Originally a binary classification method, where given a set of training examples, each marked as belonging to class a or b, a SVM training algorithm builds a model that assigns new examples to one class or the other, making it a non-probabilistic binary linear classifier.

SVM's formulation embodies the Structural Risk Minimisation (SRM) principle, which has been shown to be superior to traditional principle, that are used on conventional neural networks (ANN and CNN) [97]. SRM minimises an upper bound on the expected risk, as opposed to Empirical Risk Minimisation (ERM) that minimises the error on the training data. This difference gives SVM with a greater ability to generalise, which is the goal in statistical learning. Today, SVM is a very popular technique for classification problems.

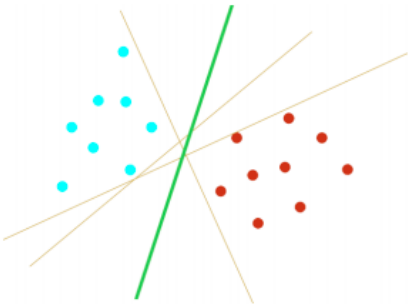


Figure 5.2: An example of linear classifiers [97].

The best way of explaining SVM is to consider a two-class problem. This problem concerns the separation of two classes using a *line*, also known as classifier. The goal is to produce a classifier that will work well on unseen examples, i.e. it generalises well. With the help of Figure 5.2 one can see that there are many possible linear classifiers, that can separate the data, but there is only one that maximises the margin (maximises the distance between it and the nearest data point of each class). This linear classifier is termed the **optimal separating hyperplane** [97].

This optimal linear classifier is obtained by maximizing the *vectors* (hence the name of support vector), normal to the classifier, between the closest datapoint and said line.

However the separation is not always possible using a linear classifier and sometimes it is needed a non-linear classifier, however non-linear classifiers are very sensible to parameter tuning. Figure 5.3 illustrates this concept clearly.

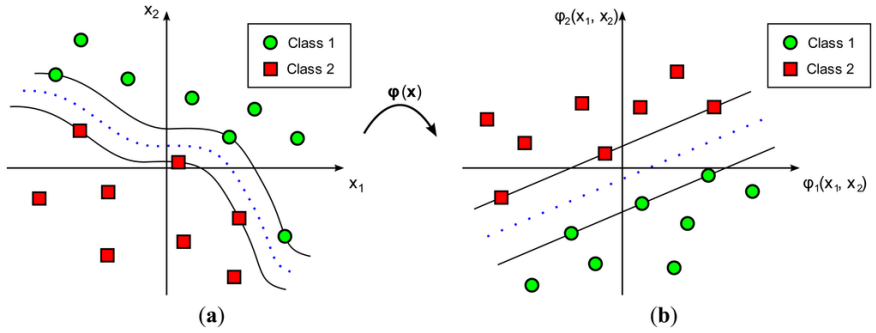


Figure 5.3: (a) non linear classifier and (b) linear classifier [98].

However, linear or non-linear classifiers may not be possible or reasonably easy to compute. A solution to this is to project the data to a n-dimensional hyperplane, try to compute a classifier that correctly separates the classes, Figure 5.4, and then transform said classifier back afterwards. This hyperplane is called the kernel and this projection technique is the idea behind the "kernel-trick" in SVM.

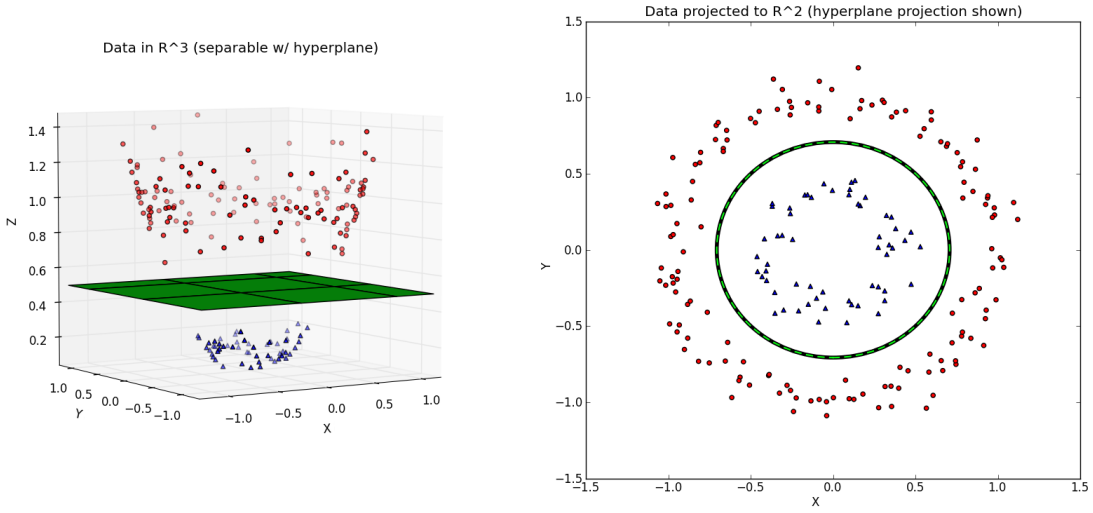


Figure 5.4: The "Kernel-trick" in SVM [99].

SVM have severall advantages according to the *scikit-learn* library web page:

- Effective in high dimensional spaces.
- Still effective in cases where number of dimensions is greater than the number of samples.
- Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.
- Versatile: different Kernel functions can be specified for the decision function. Common kernels are provided, but it is also possible to specify custom kernels.

However, the disadvantages of support vector machines include:

- If the number of features is much greater than the number of samples, avoid over-fitting in choosing Kernel functions and regularization term is crucial.
- SVMs do not directly provide probability estimates, these are calculated using an expensive five-fold cross-validation.

5.3.1 Kernels available

In Python the kernel functions available are the following:

- **linear**: $\langle x, x' \rangle$.
- **polynomial**: $(\gamma \langle x, x' \rangle + r)^d$.
- **rbf**: $\exp(-\gamma \|x - x'\|^2)$.

A comparison between the kernels can be seen in Figure 5.5

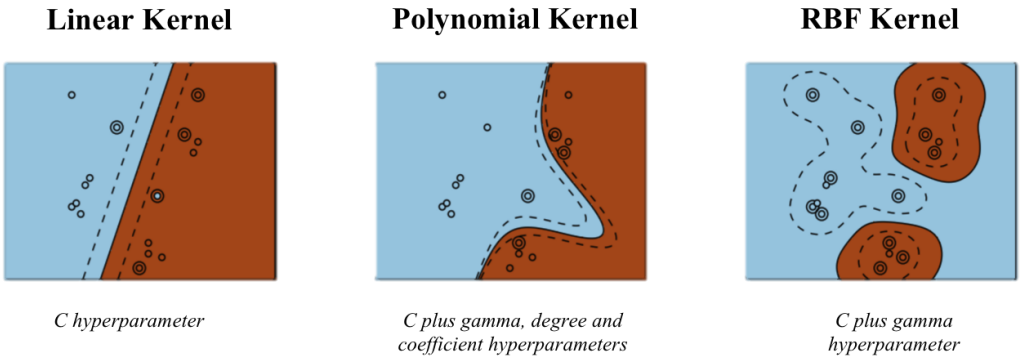


Figure 5.5: Comparison of the kernels available [100].

The kernel used for this project was the one called 'rbf'.

5.3.2 Python Implementations

There are several implementations and algorithms for applying SVM. Given that the *scikit-learn* library was being used, the possible implementations for SVM were the **SVM.SVC**, **SVM.NuSVC** and the **SVM.LinearSVC** which are capable of performing multi-class classification on a given dataset.

SVC and **NuSVC**, implement the *one-against-one* approach for multi-class classification, but accept slightly different sets of parameters and have different mathematical formulations. On the other hand, **LinearSVC**, a method that uses the *one-vs-the-rest* approach for multi-class classification, is another implementation of Support Vector Classification for the case of a linear kernel, this implementation does not accept the parameter *kernel* because it is, by nature, a linear classifier. This is shown in Figure 5.6.

According to the same source, *one-vs-rest* classification is usually preferred,

since the results are mostly similar, but the runtime is significantly less.

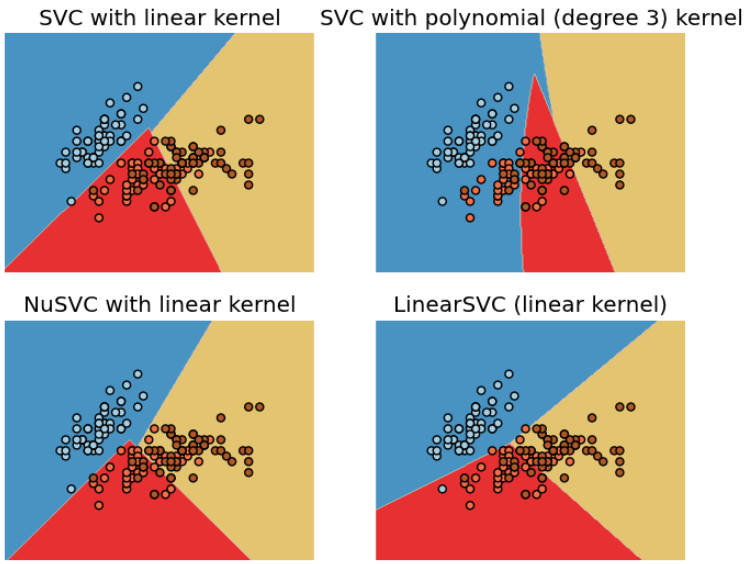


Figure 5.6: Comparison of the different SVM implementations, according to *scikit-learn* sourcefourge webpage.

5.4 Convolutional Neural Networks

Deep learning is a branch of machine learning based on algorithms that attempt to model high level abstractions of data. Convolutional Neural Networks (CNN) is a deep learning algorithm with hierarchical neural networks whose convolutional layers alternate with sub-sampling layers, followed with a full connection layer. A CNN primarily mimics the human visual system, which can efficiently recognize the parterns and structures in a visual scenery [101].

As a result, nowadays CNNs are successfully applied in many areas relating to image processing such as face recognition, object recognition, hand written recognition, video analysis, among other problems.

As propposed by [101] CNN can be applied to bearing fault classification through the analysis of a 2D gray level image generated by time-domain vibration signals. Given this, it is possible to use CNN for bearing analysis.

Typical CNN consists of four types of layers: convolution layer, sub-sampling layer, full connection layer and output layer. The network layers are arranged in a feed-forward structure: each convolution layer is followed by a sub-sampling layer. The last subsampling layer is followed by a full connection layer, which finally followed by the output layer. At convolutional layer, the previous layer feature maps are convolved with learnable kernels and put through the activation function to form the output feature map. Each kernel is used at every position of the input. CNN exploit sparse connectivity by making the kernel smaller than the input and enforcing a local

connectivity pattern among neurons of adjacent layers. Each output map may combine convolution with multiple inputs maps. However, for each output map, the input maps are convolved with distinct kernels. Each kernel is used at every position of the input. The parameter sharing used by the convolution operation means that rather than learning a separate set of parameters for every location, we learn only one set.

Each convolution layer is followed by a sub-sampling layer. A sub-sampling layer produces down-sampled versions of the input maps, progressively reduces the spatial size of the representation. That helps to decrease the number of parameters and computation in the network. Moreover, sub-sampling layer makes the representation become invariant with a small translation of the input. If there are N input maps, there will be exactly N output maps, although the output maps will be smaller.

The full connection layer is a traditional feed-forward neural network, neurons in this layer have full connections to all activations in the previous layer. The purpose of the full connection layer is to use the features from previous layer for classifying the input image into various classes. The final layer in a CNN is output layer, using the softmax as the activate function.

With three architectural ideals: local receptive fields, weight sharing and subsampling, CNN has many strengths: First, feature extraction and classification are integrated into one structure and fully adaptive. Second, the network extracts 2-D image features at increasing dyadic scales. Third, it is relatively invariant to geometric, local distortions in the image.

Even though this method is briefly explained in this section, it was not directly used for this project. OMNEO, a subsidiary company of Siemens PLM Software, applied this deep learning method to the same data used in this thesis. Consequently, the results obtained will be used as a comparison.

However, one must bear in mind that this method is a much more complex method than the ones mentioned before. Additionally it is even stated at the deep learning web page that, in order to run their implementation of the method, a computer must meet a certain number of requirements rendering this as a very powerful, though very complex and computer-dedicated, method for classification problems.

The Neural Networks techniques are part of Deep Learning (DL), a broader family of techniques and algorithms of machine learning. Even though Deep Learning is receiving more attention lately, it has a few disadvantages:

- Training a DL requires more data than other methods.
- DL and Neural Networks (NN) models are not very interpretable.
- Even though it works very well in some cases, it doesn't replace other methods.

The results obtained by OMNEO will be presented in chapter 6 both as a reference, validating or not the method developed in this text, and as an alternative approach.

5.5 Cross-Validation

An important concept that must be made clear is how cross-validation was performed across the Decision Tree, Support Vector Machines and Random Forest methods.

A well designed model should avoid learning the parameters of a prediction function and testing it on the same data. This is a mistake since a model that would just repeats the labels of the samples that it has just seen would have a perfect score but would fail to predict anything useful on yet-unseen data. This situation is called overfitting.

In order to avoid this, it is common practice when performing a (supervised) machine learning experiment to divide the data into two sets: one for training and another for effective validation.

Cross validation was performed using the function *StratifiedKFold* which, as one can see from image 5.7, randomly splits the data into a set of rounds (10 rounds were used for this project) and, randomly selects one dataset as validation data and the rest as training data. This is important since validation must be performed on sets of data that the model has never seen before.

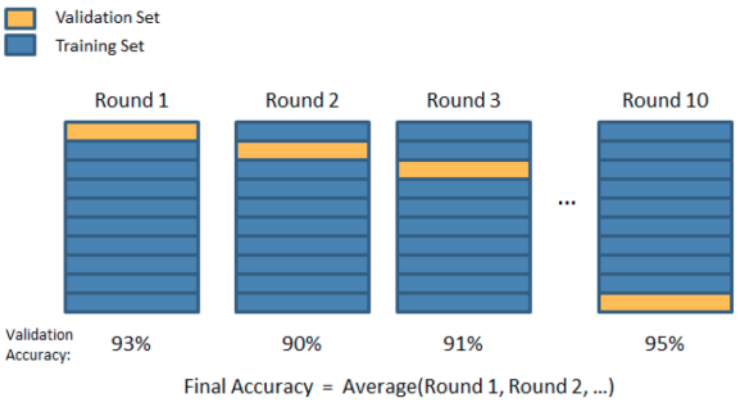


Figure 5.7: Illustration of cross-validation through the *StratifiedKFold* function [102].

CHAPTER 6

Results

6.1 Test-rig Description and Sensors

In order to simulate running conditions a machine fault simulator was used. The machine in question, from SQ is one shown in Figure 6.1. Using this equipment one is able to mount two bearings on a shaft. The shaft is powered by an electric motor which is controlled by a simple variable frequency drive.

Table 6.1: SQ operating conditions.

SQ operating conditions [5]	
Cases	Healthy, Inner and Outer
Loads	LM
Speeds	300:120:2700

The engine was a tree-phased electric motor with a power rating of 0,5 HP, nominal speed of 3 450 rpm and rated frequency of 60 Hz. As clarified in Table 6.1 the range of speeds measured was from 300 up to 2 700 rpm with increments of 120 rpm, resulting in a total of 21 cases. In addition, a static load of 5 kg was added to the center of the shaft, denoted as Load Middle (LM) in order to increase the amplitude of the faults in the bearings. The cases studied in these tests were the Healthy, Inner race fault, Outer race fault and Ball fault. The bearings were placed in the right housing while the left housing kept a healthy bearing for all the measurements. The bearings used were ER-16K whose geometric properties are detailed in the table 6.2. For the SQ machine, the sensors used were 2 triaxial accelerometer, 2 acoustic emission sensors and 2 microphones placed in the healthy, to record the validation results, and in the faulty side. For this test bench, the sensors which returned the best results was the one located close to the housing where was located the defected bearing.

Table 6.2: Tested bearing geometric properties [5].

ER-16K	
Ball Diameter	7,94 mm
Pitch Diameter	39,32 mm
Number of Balls	9
Contact Angle	0°

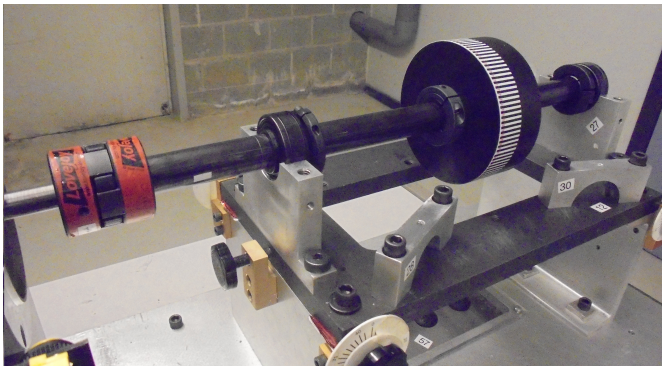


Figure 6.1: Fault machine simulator from SpectraQuest [5].

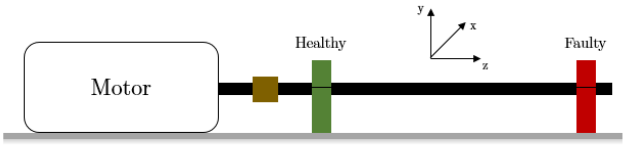


Figure 6.2: SQ Envelope - Axis definition [5].

6.2 Pre-processing results and automated processing results

Given that the aim of this project was to develop a semi-automatic tool, the first result is the input window where the user chooses the frequency band to filter the signal, Figure 6.3, corresponding to the reference case, and consequently, generate amplitude ratios to find the frequency bands to filter the signals that belong to other speeds (remember Figure 3.11 showing how the frequency band for filtering was chosen).

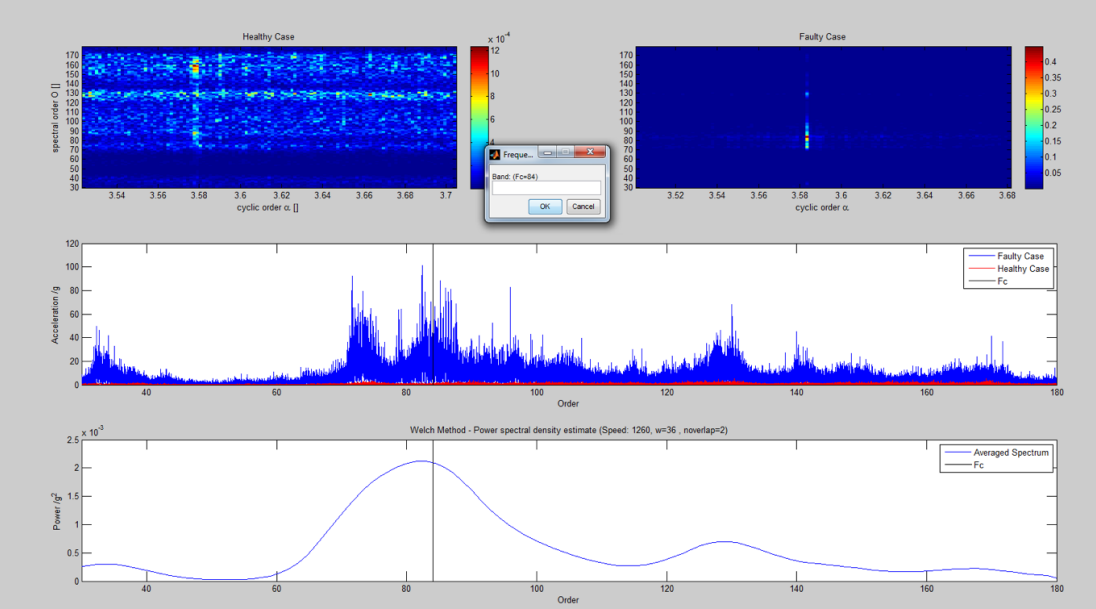
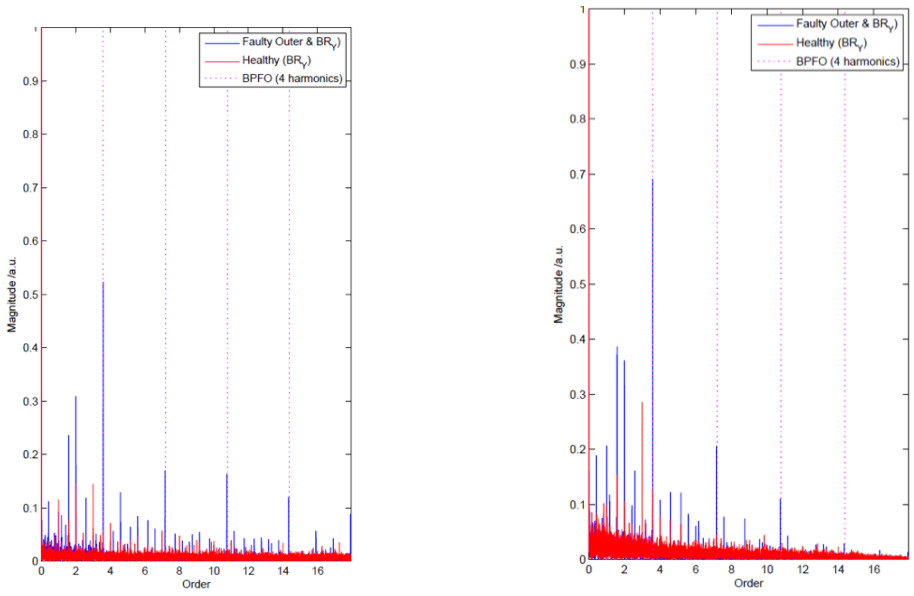


Figure 6.3: Input Window. User can enter an interval of frequencies for filtering around the FC.

Comment 1. Figure 6.3 shows, at the top, the Cyclic Spectral Correlation calculated for the Healthy signal and for the Faulty signal. One can see that, for the faulty signal, a *point* is isolated in the middle of the image, a situation very different to the healthy case. In addition the colormap scales are completely different. If one needs to analyse these images on a real-world situation, this would automatically lead to the decision of a faulty bearing. Below this is a comparison of the spectrum of healthy and faulty signals, showing a clear peak around the FC highlighted on the CSC map. As a guide for the user, the spectral density of the faulty vibration signal is displayed on the last plot providing the user the idea of where the highest density is, hence which area to chose for filtering.

Next are presented the filtering results for two speeds: 1 260 rpm (Figures 6.4 and 6.5) and 1 500 rpm (Figures 6.6 and 6.7).



Non-filtered signal. Filtered Signal.

Figure 6.4: Spectrum comparison for 1 260 rpm.

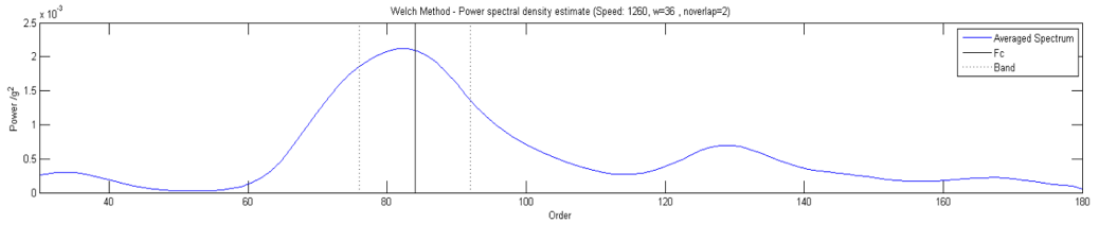
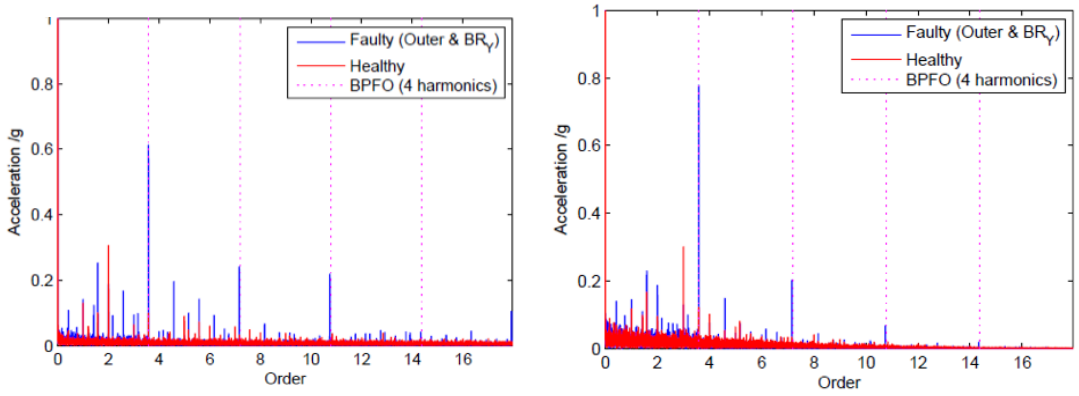


Figure 6.5: PSD curve for 1 260 rpm showing FC and the frequency window chosen.



Non-filtered signal. Filtered Signal.

Figure 6.6: Spectrum comparison for 1 500 rpm.

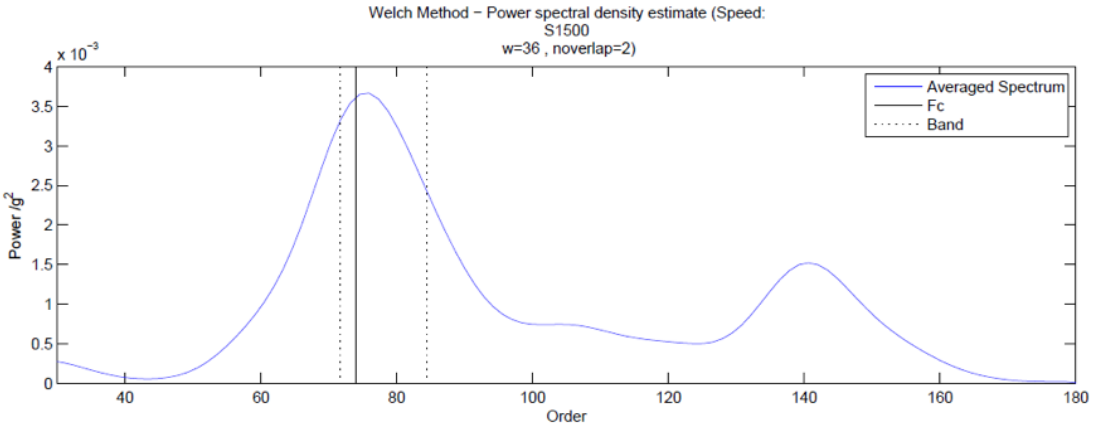


Figure 6.7: PSD curve for 1 500 rpm showing FC and the frequency window chosen.

Comment 2. Using the filtering bands set on Figures 6.5 and 6.7 one arrived at the results shown in Figures 6.4 and 6.6. One can clearly see an increase in character in the first and second harmonics for the case of 1 260 rpm, whereas in the case of 1 500 rpm a clear increase in character for the first harmonic is seen. This proves the efficacy of the filtering process.

The method proposed, filters the faulty data while simultaneously filtering the healthy data as well. At the end, trying to assemble results from each class (outer, inner and ball fault and each correspondent healthy data) rendered as an incorrect analysis given that it was not possible to define a refference case that was common to all classes.

At the same time, analysing the Figures 6.8, 6.9, 6.10 and 6.11 one can conclude that it is not possible to define a common area for filtering, because the frequencies excited vary by speed regime and sensor and are not consistent to a common place.

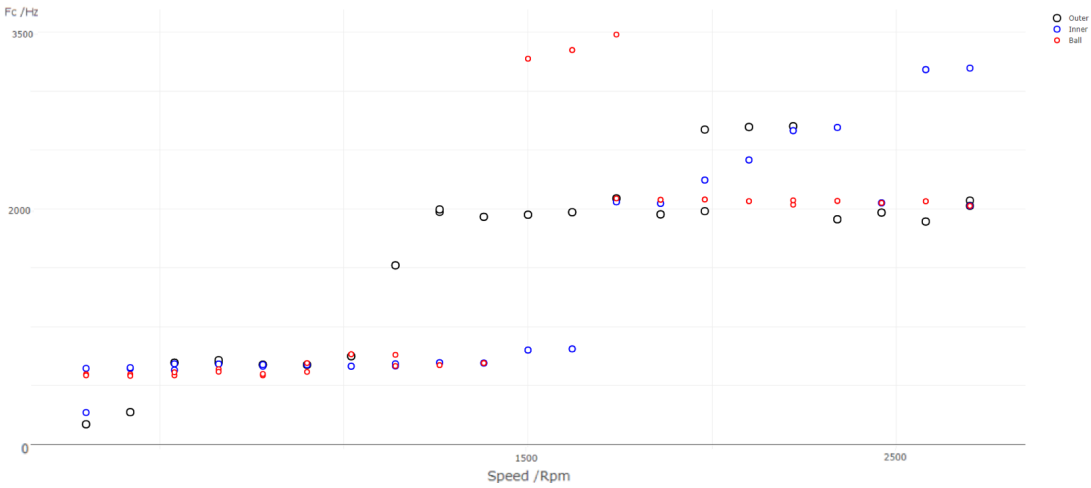


Figure 6.8: Frequencies excited by each fault measured by the Mic sensor.

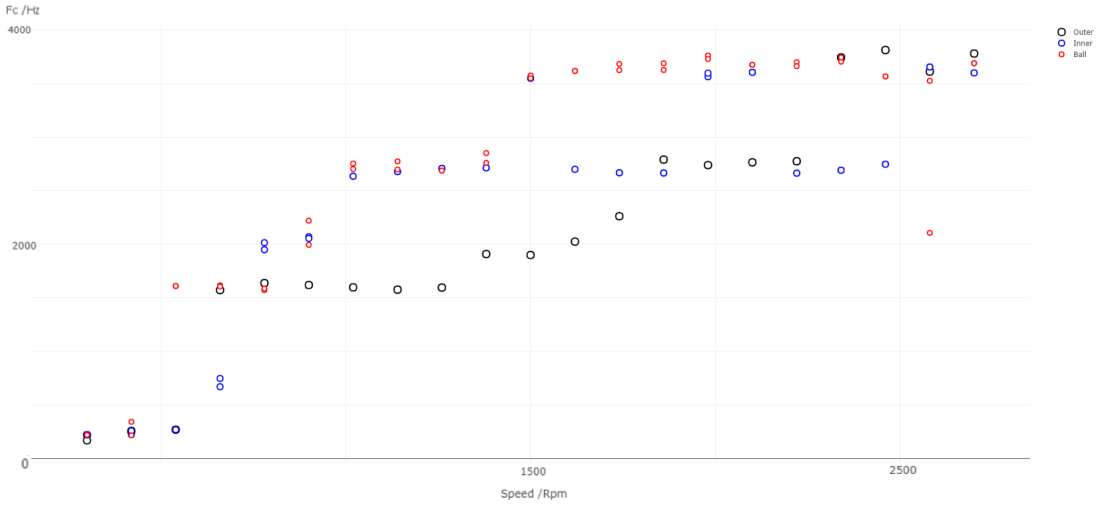


Figure 6.9: Frequencies excited by each fault measured by the $\boldsymbol{BR_x}$ sensor.

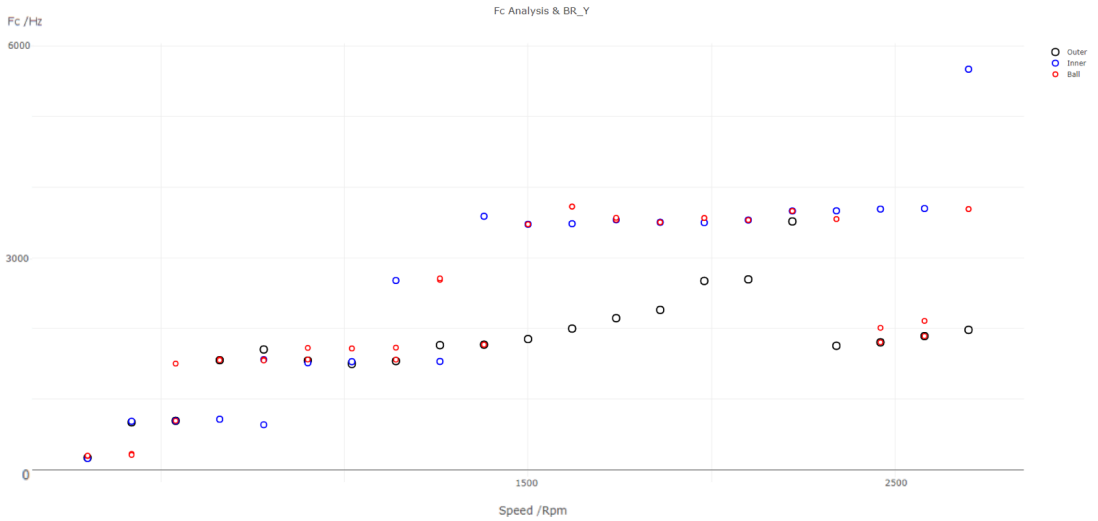


Figure 6.10: Frequencies excited by each fault measured by the $\boldsymbol{BR_y}$ sensor.

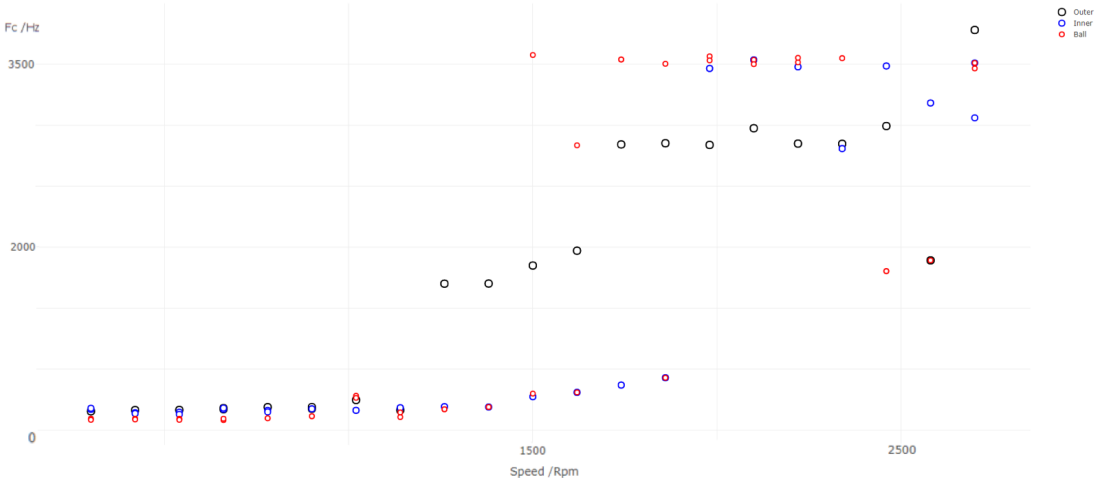


Figure 6.11: Frequencies excited by each fault measured by the $\boldsymbol{BR_z}$ sensor.

Given this, the method applied was to use the information calculated from each fault and speed (central frequencies and filtering bands) and perform a cross-filtering process while simultaneously filtering the healthy case. For instance, the signal corresponding to the outer fault was filtered with the information calculated to the inner and ball fault. After this, the features calculated for each filtering band were assembled on a single matrix and this data was used for the preceding analyses. A nomenclature was added to differentiate with which band the feature was calculated: **IB** for Inner Band, **OB** for Inner Band and **BB** for Ball Band.

6.3 Feature Reduction

This section will present the results for feature reduction for each sensor. First for PCA and then for LDA analysis. Following the explanation given in section 4, PCA performs an orthogonal transformation to convert a dataset of possibly correlated values into one composed of linearly uncorrelated variables called PC, each component holding a given percentage of data compared to the original, all of this without taking into account the class labels. On the other hand, LDA finds a linear combination of features that characterizes or separates two or more classes of objects or events, using the information given by the class labels, hence a supervised method.

6.3.1 PCA Results

Figure 6.12 shows different perspectives around the *3d* scatter plot resulted from applying PCA to this sensor's signal.This signal had the signal processing techniques embedded into its hardware, thus, the signal obtained was not treated with the tool developed in section 3.

AE sensor

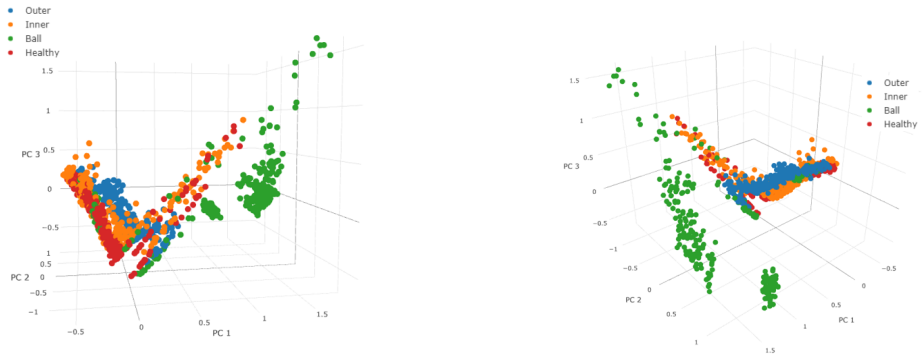


Figure 6.12: Diferent perspectives of PCA plot for AE sensor.

Comment 1. Regarding PCA analysis of the AE signal, one can only state, from analysing figure 6.12, that class separation is particularly difficult for this sensor. Furthermore, using the first three components amounts to a representation of 91,09% of the total data. Moreover, the features with more importance for this sensor are the Speed, P2P, P2R, 75% and the entropy.

Mic sensor

Oppositely to the AE signal, the subsequent sensors' signals where treated and enhanced using the tool developed. Figures 6.13 and 6.14 show the plots for PCA analysis applied on non-filtered signal and filtered Signal, respectively.

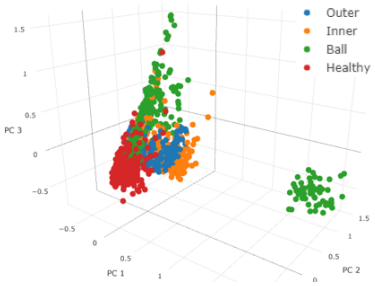


Figure 6.13: Non-filtered Signal.

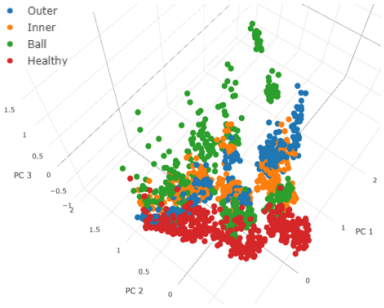


Figure 6.14: Filtered Signal.

Comment 2. Regarding PCA analysis of the Mic signal, and comparing the non-filtered and filtered signals, it can be stated that filtering benefited class separation once some clustering can be seen on the PCA plots of the filtered signal. Furthermore, for the non-filtered signal, using the first three PC a representation of 75,87% of the total data is accomplished. Oppositely, for the filtered signal and using the same number of components, the value drops down to 52,98%. The drop in representation percentage between signals and improvement in class separation is due to the fact that the enhancement techniques removed most of the *noise* that was refraining PCA from achieving better separation on the non-filtering signal.

For this analysis, only three PC were used for visualisation purposes (if more than three were used, one could not see such representation on a 3D plot). For further analysis however, for both cases, the number of components to use should be sufficient to represent a considerable percentage of the total data, for example 90% or more. For this sensor, the most relevant features were the speed, the Entropy, Went_Sha, Went_Log , 75%, P2R, RMS, Mean, Imp and Mar factors.

BR_y sensor

Regarding the accelerometers, only the results from the BR_y sensor will be presented because this was the direction (vertical direction) that best captured the fault signature. One can observe on Figures 6.15 and 6.16 PCA analysis applied on non-filtered signal and filtered Signal, for the BR_y sensor. The results from other directions can be found in Appendix A.1.

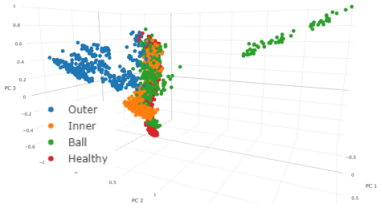


Figure 6.15: Non-filtered Signal.

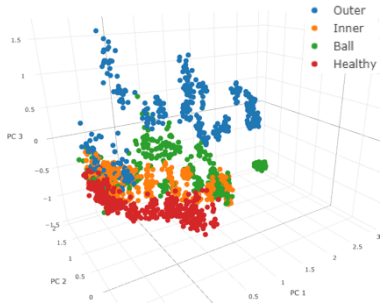


Figure 6.16: Filtered Signal.

Comment 3. The BR_y sensor was the one that best benefited from the filtering tool developed. Class separation on this sensor using PCA analysis outperforms PCA analysis applied to other sensors. Furthermore, for this sensor, and using the first three PC, one can obtain 73,17% and 59,31% representation of the total data, respectively for the non-filtered signal and filtered signal. Again, here one concludes that the signal enhancing (signal processing) techniques developed in this thesis helped to increase the fault's character which consequently improves class separation, even though the amount of information kept by the first three components is lower for the filtered signal.

Further analysis using PC should use a number of said components that represents a considerable percentage of the total data, for instance 90% or more.

6.3.2 LDA Results

Following the same pattern, the following figures show different perspectives around the scatter plot resulting from applying LDA to different sensor's data.

AE sensor

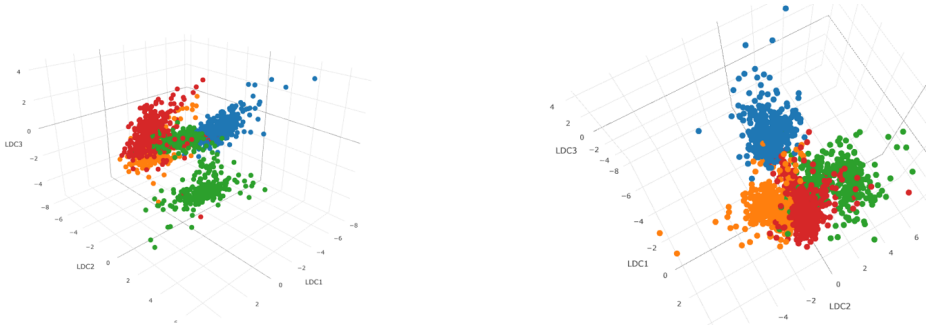


Figure 6.17: Diferent perspectives of LDA plots for AE sensor.

Comment 1. Because LDA is a labeled method, it can separate the classes in a considerable manner. For the AE sensor, and using this method, class separation is acceptable taking into consideration that this sensor's signal was not treated with the method developed in this thesis. Furthermore, this sensor had applied some processing at hardware level, the signal was demodulated and filtered at approximately 100 kHz.

Mic sensor

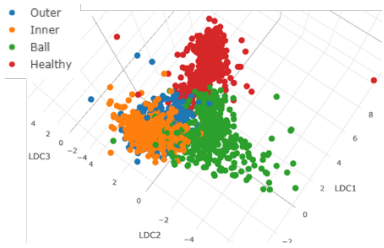


Figure 6.18: Non-filtered signal.

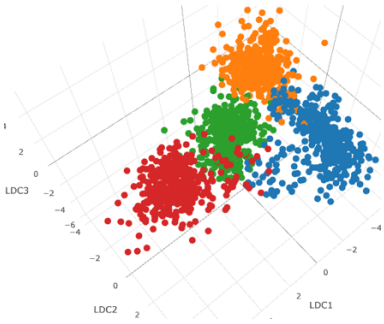


Figure 6.19: Filtered signal.

Comment 2. For the Mic sensor, one can conclude that class separation is better after applying the signal processing method developed in this project.

BR_y sensor

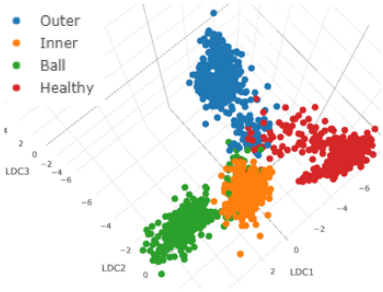


Figure 6.20: Non-filtered signal.

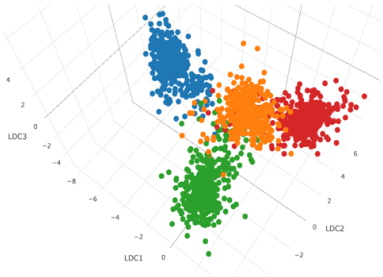


Figure 6.21: Filtered signal.

Comment 3. Comparing Figures 6.20 and 3 one can say that, the classes have come closer for the filtered signal than for the non-filtered signal. However, class separation is still possible and acceptable. Additionally, one can conclude that for the non-filtered signal (Figure 6.20), an intersection area between the ball and inner faults is seen. For the filtered signal (Figure), this intersection is no longer seen.

In a general manner, the filtering process shows benefits for the BR_y and Mic sensors and shows that class separation is possible. For the AE signal, because it was processed at hardware level, one can conclude that there is good separation. After applying LDA one sees that for all sensors class separation does not get worse after filtering. Furthermore, PCA is an unsupervised whereas LDA is a supervised method, which explains the differences in separating the classes. Additionally, it was found that using less features for LDA, the classes come closer together, which is expected because it reduces the dimensional space where the transformations occur (as can be seen in Figures 6.22 and 6.23 below).

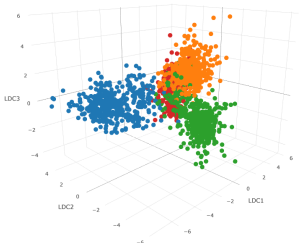


Figure 6.22: LDA using 43 manually selected features, BR_y sensor.

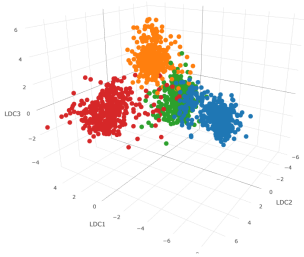


Figure 6.23: LDA with a total of 82 features (the entire set of features), BR_y sensor.

6.4 Feature Selection and Classification

It is generally accepted that the best sensor to detect REB failures is the one placed in the vertical direction. Hence, the results hereon will only concern thy $\boldsymbol{BR_y}$ sensor and the AE and Mic sensors due to the advantages in using their technology for fault detection.

6.4.1 No Feature Selection

Without selecting individual features, two tests were performed to assess feature classification. The first one consisted in using the PC from PCA and directly try to classify classes with the PC as input. The second test concerned in running the feature classification methods immediately after the step of feature extraction.

Using PC after PCA

For each of the sensors studied, the number of PC used corresponded to the maximum available for each sensor. Table 6.3 shows the results obtained.

Table 6.3: Best accuracy scores /% for non-filtered and filtered signal using PC.

	Non-Filtered					Filtered				
Sensors	PC	PC %	DT	RF	SVM	PC	PC %	DT	RF	SVM
AE	-	-	-	-	-	3	91,09	70,11	75,83	65,30
Mic	10	95,70	74,28	84,58	81,67	20	94,58	81,55	90,36	83,39
$\boldsymbol{BR_y}$	10	96,70	91,43	94,94	94,83	20	95,62	84,54	94,10	92,56

After studying Table 6.3 one can see that for the different sensors, the accuracies depend on the classification method used. Another important thing to note is that, for the filtering signal, because of what was explained previously regarding the bands, a total of 82 features were used against a total of 28 features for the non-filtered signal. It is also possible to conclude that, for the Mic sensor, filtering benefitted the accuracies for all the methods.

Immediately after feature extraction

The disadvantage of using classification with PC is that the meaning of the features is lost. Using the features directly to the Decision Tree method will result in a classification based on the features selected by the method itself, which can be interpreted by analysing the image created by the method. This topic will be explored in the next section. For now, the accuracies obtained without using the PC as input to the classification methods nor feature selection, will be shown.

As an example, table 6.4 shows a confusion matrix of the classification performed, no feature selection and filtered signal.

Table 6.4: Confusion Matrix for DT, $\boldsymbol{BR_y}$, no feature selection (Filtered Signal)/%.

	Outer	Inner	Ball	Healthy
Outer	93,27	1,42	4,74	0,48
Inner	1,44	94,31	3,79	0,48
Ball	0,96	2,84	94,79	1,43
Healthy	1,92	0,00	1,42	96,67

Comment 1. Using a confusion matrix like on Table 6.4, one can match the method and signal processing technique to the client. For example, for the ball defect case, 1,43% of the faults were identified as healthy and will go unnoticed. This parameter depends on the type of machine and the client's request. For instance, if this tool is applied to important machinery, the customer may not be able to afford to miss those 1,43% of unidentified ball faults.

Table 6.5 shows the cross-validation accuracy scores, obtained using the *pruned* Decision Tree without feature selection.

Table 6.5: Classification scores with Decision Tree for $\boldsymbol{BR_y}$ and no feature selection.

Accuracy Scores /% for $\boldsymbol{BR_y}$ sensor in rounds $R_{1...10}$									
98,84	96,43	94,05	95,24	97,62	92,86	96,43	95,24	85,54	95,18

Table 6.6 shows the best accuracies obtained for each method without feature selection.

Table 6.6: Best accuracy scores /% for each method. Non-filtered vs. filtered without feature selection.

	Non-Filtered			Filtered		
Sensors	DT	RF	SVM	DT	RF	SVM
AE	-	-	-	93,33	97,26	81,61
Mic	87,02	91,96	81,43	90,83	98,51	93,81
$\boldsymbol{BR_y}$	93,44	97,44	93,21	96,90	98,87	96,67

Comment 2. Comparing Tables 6.6 and 6.3 one can see that using the PC for classification leads to accuracies slightly lower. In addition, one can see that for table 6.6 the filtering tool developed on this project leads to higher accuracies across the sensors $\boldsymbol{BR_y}$ and Mic. It can also be observed that, independently of the classification algorithm applied, the filtering tool developed showed an increase of the global accuracies.

6.4.2 Manual Method

The first method performed for selecting features was to analyse the correlation between the several features. In order to do this, a diagonal matrix showing the correlation between the several features was studied for the $\boldsymbol{BR_y}$, AE and Mic sensors. This correlation matrix was calculated using a function available within the *pandas* library, a library for data analysis. The argument used in this function was the pearson argument, a standard correlation coefficient.

$\boldsymbol{BR_y}$ and Mic Sensors

Figures 6.24 and 6.25 show the correlations between features before and after feature elimination, respectively.

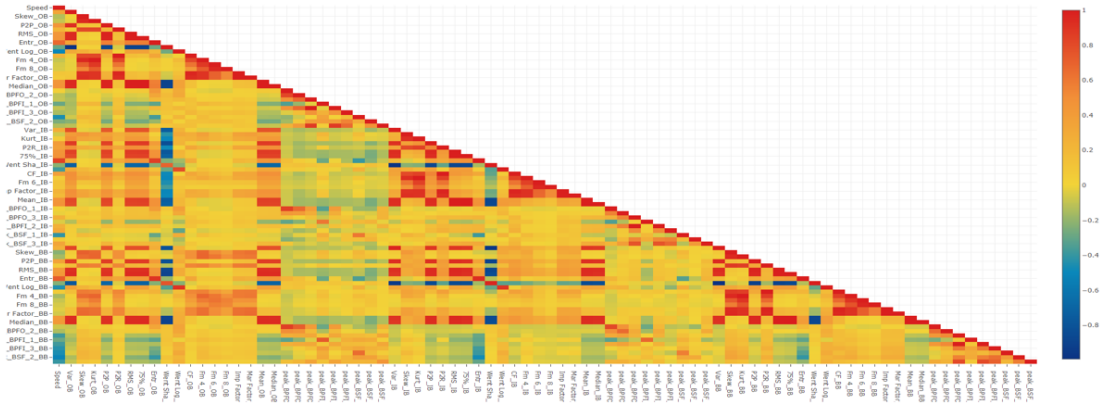


Figure 6.24: Correlation matrix of $\boldsymbol{BR_y}$ and Mic Sensors, before.

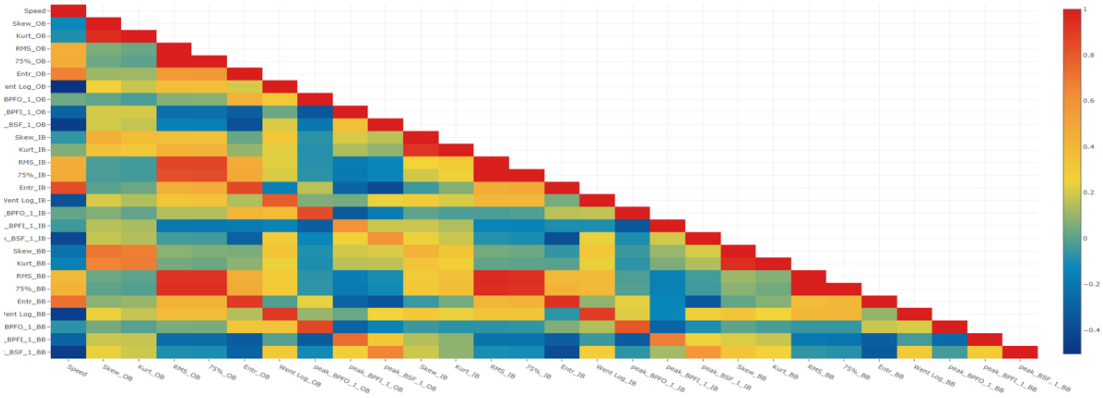


Figure 6.25: Correlation matrix of $\boldsymbol{BR_y}$ and Mic Sensors, after.

Comment 1. The process for eliminating features through this method was to analyse the most correlated features and remove one between two. For example, one can see in Figure 6.24 that the feature Skew_OB is highly correlated with the feature FM4_OB.

In Figure 6.25 it can be seen that only the FM_4 feature was removed. These features were common to the $\boldsymbol{BR_y}$ and Mic sensors given that the correlation matrices were very similar.

Manual selection based on the correlation matrix results in the following features table 6.7:

Table 6.7: Manual feature selection: $\boldsymbol{BR_y}$ and Mic.

BR_Y and BR_Mic Sensors		
Speed	Skew_OB	Kurt_OB
RMS_OB	75%_OB	Went Log_OB
Entr_OB	Peak_BPFO_1_OB	Peak_BPFI_1_OB
Peak_BSF_1_OB	Skew_IB	Kurt_IB
RMS_IB	75%_IB	Went Log_IB
Entr_IB	Peak_BPFO_1_IB	Peak_BPFI_1_IB
Peak_BSF_1_IB	Skew_BB	Kurt_BB
RMS_BB	75%_BB	WentLog_BB
Entr_BB	Peak_BPFO_1_BB	Peak_BPFI_1_BB
	Peak_BSF_1_BB	

AE Sensor

The same manual process for eliminating features that was applied for the $\boldsymbol{BR_y}$ and Mic sensors was applied for the AE sensor.

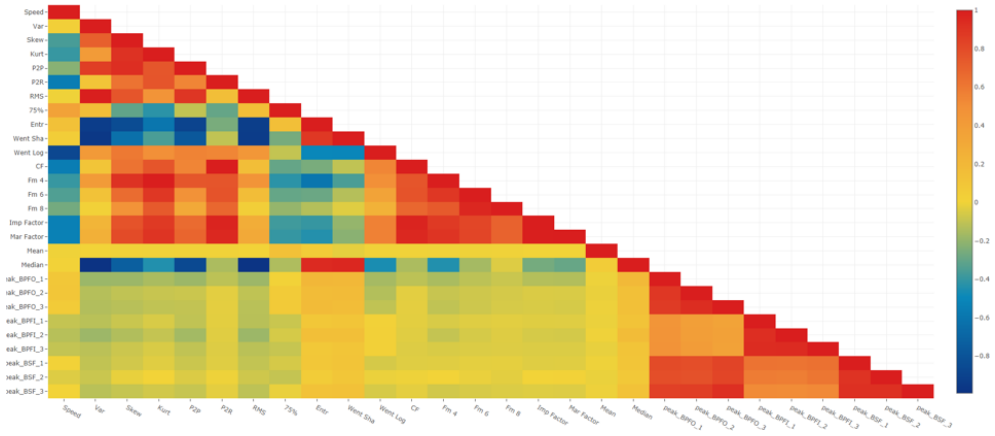


Figure 6.26: Correlation matrix of AE Sensor, before.

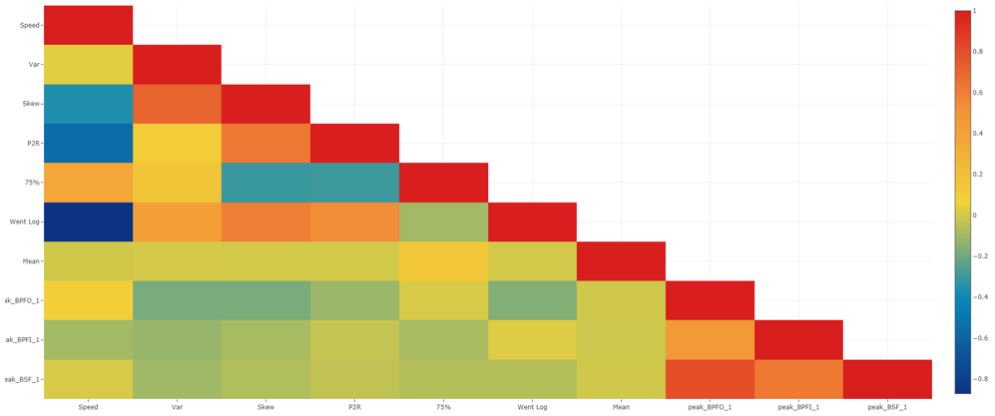


Figure 6.27: Correlation matrix of AE Sensor, after.

Comment 2. As one can see the number of *red dots* on the correlation matrix dropped considerably. The correlation levels on the negative side were not removed because, as can be seen from the scales, the biggest correlation between features is located on the positive side of the scale. Additionally, the negative side of the scale was not explored due to time constraints and because manual feature selection tends to be a subjective method. Two users might choose different features to remove and keep.

Based on the correlation matrix plotted in Figures 6.26 and 6.27, the features selected are shown on table 6.8:

Table 6.8: Manual feature selection: AE sensor.

AE Sensor
Speed
Var
Skew
P2R
75%
Went_Log
Mean
Peak_BPFO_1
Peak_BPFI_1
Peak_BSF_1

Using the manual feature selection method, the values shown in Tables 6.9 and 6.10 were obtained. Table 6.9 shows the confusion matrix obtained for the $\boldsymbol{BR_y}$ sensor and Table 6.10 shows the best accuracies obtained for each method, comparing non-filtered and filtered signals.

For further details, Figures 6.24, 6.25, 6.26 and 6.27 are presented in Appendix B on a larger scale.

Table 6.9: Confusion Matrix for DT, $\boldsymbol{BR_y}$, manual feature selection (Filtered Signal) /%.

	Outer	Inner	Ball	Healthy
Outer	95,65	0,96	3,41	0,00
Inner	3,38	93,78	2,44	0,46
Ball	0,97	1,44	97,07	0,46
Healthy	0,00	0,00	0,00	100,00

Table 6.10: Best accuracy scores /% for each method. Non-filtered vs. filtered without feature selection.

	Non-Filtered			Filtered		
Sensors	DT	RF	SVM	DT	RF	SVM
AE	-	-	-	94,15	97,80	77,38
Mic	87,61	93,45	73,33	91,68	97,92	88,27
$\boldsymbol{BR_y}$	95,36	97,5	93,63	96,67	98,57	93,33

Comment 3. Analysing Tables 6.10 and 6.9 one can see that the accuracies obtained after performing manual feature selection are close to the ones obtained without performing any form of feature selection. Due to the similarities in accuracies

and because this method tends to be subjective (depends on the operator) this method for feature selection is not recommended, or if used, should be used with caution.

6.4.3 mRMR method

The second approach to feature selection was done using the mRMR method. As explained in chapter 4 this method selects features according to the *minimum-Redundancy Maximum-Relevance* criteria, were the features are analysed, essentially, by their entropy level. It's natural form of selecting features is by selecting all features with $entropy > 0$, however, if one wishes to select a specific number of features, the method uses the specified number and selects that specific number of features.

In terms of comparing the features selected by the mRMR method, the features selected by the Decision Tree algorithm, without pre-selecting features are shown in Tables 6.11, 6.12 and 6.13 for the filtered signal for each sensor. These features are obtained by analysing the decision tree images that created by the function in *Python*, Appendix C shows several decision trees created for this classification problem.

Table 6.11: Feature Selection with DT for $\boldsymbol{BR_y}$ sensor.

BR_Y Sensor
Went Log_OB
Went Log_IB
Peak_BPFO_1_IB
Peak_BPFO_1_BB
Peak_BPFI_1_OB
Peak_BPFI_1_IB
75%_BB
Peak_BPFI_1_BB
Skew_IB
Skew_BB

Table 6.12: DT Feature Selection with DT for Mic sensor.

BR_Mic Sensor
Var_IB
Went Log_OB
75%_BB
Peak_BPFO_1_OB
Skew_IB
Peak_BSF_1_OB
Peak_BPFO_3_BB
Went Sha_BB
Went Log_BB
RMS_OB
Mar Factor_IB
75%_IB
Peak_BPFI_2_OB
Peak_BPFI_2_IB
Peak_BPFO_3_OB
Peak_BSF_2_IB
Peak_BPFO_2_IB
Mean_IB
Peak_BPFI_1_OB
Var_BB
Peak_BPFO_1_BB
Peak_BSF_1_BB
Median_OB
Peak_BSF_3_IB

Table 6.13: Feature Selection with DT for AE sensor.

BR_AE Sensor
RMS
Var
Mar Factor
Peak_BPFO_3
Median
Skew
Peak_BPFI_1
Went Log
Peak_BPFO_2
Fm 6
Peak_BPFI_2
Fm 8
75%
CF

Comment 1. Looking at the previous tables, and without specifying any criteria to the Decision Tree algorithm, one can see that the features from the frequency domain (the harmonics' amplitude) are presented as relevant features. Additionally, the features that are selected depend on the sensor being studied.

The first approach to using this method for a pre-selection of features was through its native form of selecting features. According to the explanation given in section 4 this method makes use of the *entropy* quantity to calculate the information gain given by each feature. Thus, the features selected for the $\boldsymbol{BR_y}$, AE and Mic sensors, using the filtered signals, are presented in Tables 6.14, 6.15 and 6.16.

Table 6.14: Feature Selection with mRMR for $\boldsymbol{BR_y}$ Sensor.

BR_Y Sensor
Went Log_OB
Went Log_IB
Peak_BPFO_1_IB
Peak_BPFO_1_BB
Peak_BPFO_1_OB
Peak_BPFI_1_IB
75%_BB
Entr_OB
Peak_BPFI_1_OB
Skew_IB
Skew_BB
Peak_BPFO_2_OB
Fm 8_BB

Table 6.15: Feature Selection with mRMR for Mic Sensor.

BR_Mic Sensor
Entr_BB
Fm_8_IB

Table 6.16: Feature Selection with mRMR for AE Sensor.

BR_AE Sensor
RMS
P2P
Peak_BPFO_2
75%

Using the features shown in Tables 6.14, 6.15 and 6.16, class identification was performed and the results are shown in tables 6.17 and 6.18.

Table 6.17: Confusion Matrix for DT, $\boldsymbol{BR_y}$, mRMR feature selection (Filtered Signal) /%.

	Outer	Inner	Ball	Healthy
Outer	96,62	0,00	3,41	0,00
Inner	1,45	95,22	2,93	0,46
Ball	0,48	0,00	99,02	0,46
Healthy	0,00	0,48	0,49	99,09

Table 6.18: Best accuracy scores /% for each method. Non-filtered vs. filtered with mRMR feature selection.

	Non-Filtered			Filtered		
Sensors	DT	RF	SVM	DT	RF	SVM
AE	-	-	-	88,92	91,13	74,35
Mic	74,96	73,04	62,56	59,04	59,23	38,45
$\boldsymbol{BR_y}$	89,83	93,16	85,77	97,50	98,33	90,83

Comment 2. Observing Tables 6.17 and 6.18 one can see that the accuracies for the Mic signal are lower than in table 6.6. This is due to the fact that as seen in table 6.15 the method selected only four features for the Mic sensor. Furthermore, for the values in table 6.18, when considering the non-filtered signal, the method selected seven features for the $\boldsymbol{BR_y}$ sensor and only two features for the Mic sensor.

Since the mRMR method, natively, selected only two features for the Mic sensor and four features for the AE sensor, a new criteria was used. The new criteria consisted in selecting a specific number of features, in this case ten, features per sensor, resulting in the features shown in tables 6.19, 6.20 and 6.21.

Table 6.19: New Features with mRMR for $\boldsymbol{BR_y}$ Sensor.

BR_Y Sensor
Went Log_OB
Went Log_IB
Peak_BPFO_1_IB
Peak_BPFO_1_BB
Peak_BPFO_1_OB
Peak_BPFI_1_IB
75%_BB
Peak_BPFI_1_OB
Skew_IB
Skew_BB

Table 6.20: New Features with mRMR for Mic Sensor.

BR_Mic Sensor
Entr_BB
Fm 8_IB
Peak_BPFO_1_OB
Went Log_BB
Entr_IB
Went Sha_OB
Var_IB
Skew_IB
Peak_BPFO_1_BB
Fm 8_OB

Table 6.21: New Features with mRMR for AE Sensor.

BR_AE Sensor
RMS
P2P
Peak_BPFO_2
75% Var
Median_BR_AE
Peak_BPFI_1 Peak_BPFO_3
Skew
Fm 8

Table 6.22: Confusion Matrix for DT, $\boldsymbol{BR_y}$, mRMR feature selection (10 features and Filtered Signal) /%.

	Outer	Inner	Ball	Healthy
Outer	96,14	1,91	1,95	0,00
Inner	1,45	93,78	4,39	0,46
Ball	1,45	5,74	92,20	0,46
Healthy	0,00	0,00	0,00	100,00

Table 6.23: Best accuracy scores /% for each method. Non-filtered vs. filtered, mRMR feature selection (10 features and Filtered Signal).

	Non-Filtered			Filtered		
Sensors	DT	RF	SVM	DT	RF	SVM
AE	-	-	-	93,44	96,25	77,44
Mic	89,03	90,77	71,79	92,27	97,14	76,85
$\boldsymbol{BR_y}$	94,16	95,95	88,04	95,59	97,92	86,91

Comment 3. Comparing tables 6.22 and 6.23 with 6.17 and 6.18 one can say that the overall best accuracies decreased for the Mic and $\boldsymbol{BR_y}$ sensors, by setting the number of features selected to ten. However, using these new features, the Decision Tree method was able to identify the healthy case without miss-identifications with other cases.

Furthermore, the features selected by the mRMR (Tables 6.14 and 6.15, 6.16, 6.19, 6.20 and 6.21) and the Decision Tree methods (Tables 6.11, 6.12 and 6.13) are very similar, both selecting features in common and presenting features from frequency domain as relevant. Additionally, the Decision Tree method selected for the $\boldsymbol{BR_y}$ sensor, a considerable amount of features as relevant (Table 6.11). In general, both methods selected features in common, specially after the criteria used on the mRMR method was changed.

However, it must be noted that the optimal number of features to use is, by itself, object of study which was not performed in this thesis due to time constraints.

6.5 Feature Classification: Combination of Sensors

As stated previously, OMNEO a subsidiary company from Siemens PLM Software, used convolutional neural networks in order to classify the classes using the same dataset used in this thesis. Convolutional Neural Networks are a much more dedicated and complex method for feature classification. In this thesis the same groups of sensors were used and the results were directly compared with the results obtained by OMNEO. Table 6.24 shows which sensors are in each group and the best accuracy obtained for each classification method. This table was obtained using the filtered signal and the mRMR feature selection, selecting ten features per sensor.

Table 6.24: Comparison between classification methods for each sensor group /%.

Classes	Speeds	Sensors	OMNEO	DT	RF	SVM
all	all	1: all	99,20	98,33	99,88	99,17
all	all	2: AE	70,96	92,37	96,49	77,74
all	all	3: Mic	99,29	92,27	97,14	78,33
all	all	4: BR_x	97,64	91,90	94,52	81,37
all	all	5: BR_y	99,32	95,59	97,92	86,91
all	all	6: BR_z	99,57	89,41	94,46	85,42
all	all	7: AE+Mic	99,11	95,11	98,93	93,75
all	all	8: AE+BR_y	99,14	96,31	98,93	92,86
all	all	9: AE+Mic+BR_y	99,53	97,03	99,46	98,04
all	all	10: Mic+BR_y	99,78	96,90	99,29	95,56

Table 6.25 compares the best accuracies obtained for each sensor and groups of sensors with and without signal processing techniques. For this, mRMR feature selection was used, selecting ten features per sensor.

Table 6.25: Comparison between groups, before and after signal processing /%.

Classes	Speeds	Sensors	DT	RF	SVM	DT	RF	SVM
			Non-filtered Signal			Filtered Signal		
all	all	1: all	96,54	99,70	98,45	98,33	99,88	99,17
all	all	2: AE	-	-	-	92,37	96,49	77,74
all	all	3: Mic	89,03	90,77	71,79	92,27	97,14	78,33
all	all	4: BR_x	90,58	94,94	82,68	91,90	94,52	81,37
all	all	5: BR_y	94,16	95,95	88,04	95,59	97,92	86,91
all	all	6: BR_z	89,53	95,06	89,05	89,41	94,46	85,42
all	all	7: AE+Mic	96,55	99,05	84,05	95,11	98,93	93,75
all	all	8: AE+BR_y	97,98	98,93	94,35	96,31	98,93	92,86
all	all	9: AE+Mic+BR_y	97,26	99,46	96,91	97,03	99,46	98,04
all	all	10: Mic+BR_y	94,51	97,26	94,17	96,90	99,29	95,60

Comment 1. Comparing tables 6.24 and 6.25 one concludes that using the methods displayed on this project one can obtain accuracy values close to the ones obtained using CNN, a method that requires more data and is not interpretable.

Furthermore, the signal processing method developed in this thesis benefited feature classification for the Mic and $\boldsymbol{BR_y}$ sensors in overall. The group of sensors that obtained the best classification was the group with all sensors, whose accuracies were superior for all three classification methods used, however, using group 9 (AE+Mic+BR_y) one obtains accuracies similar to the ones obtained using all sensors. Individually, $\boldsymbol{BR_y}$ the sensor that resulted in the best accuracies using the methods developed in this thesis.

At the end one must think about the ratio of accuracy vs. price, since using a

lot of sensors becomes expensive, especially if the increase in accuracy does not improve considerably to justify the investment.

CHAPTER 7

Conclusion and Future Work

7.1 Conclusions

After literature study, programming and data testing, it was possible to form some conclusions.

According to various papers, condition monitoring through vibration analysis is still the most reliable and used method for condition monitoring. Additionally, a wide range of signal processing techniques that can be applied to the vibration signal are available to use and, basically, which is best to use depends on the author, since various techniques are suggested and advised. However one common technique is to apply envelope analysis with filtering. Furthermore, deterministic component removal is successfully applied using the CEP.

A common difficulty to applying envelope analysis is how to set the window to envelope. This project developed a semi-automated method that performs this filtering requiring one input from the user. This method also performs successfully as is proven by PCA and LDA results showing improved class separability after filtering, for some sensors. Additionally, classification accuracies benefited as well from the envelope filtering step, proving this step as successful.

The feature selection tool selects features both on time domain and frequency domain for the data from the $\boldsymbol{BR_y}$ sensor, using the native parameter of choosing features while $entropy > 0$. However, for other sensors the parameter to select the optimal number of features needs to be further investigated. The results obtained by the mRmR algorithm showed that the features selected are comparable with the first layers of the ones selected by the decision tree. Thus, the decision tree method should be explored as a feature selection method as well, and the results used as input for the other feature classification algorithms. Due to time limitation, this technique was not further implemented.

Using the PC after applying PCA, as suggested by the literature, leads similar classification accuracies. Here is important to understand how many PCs should be taken into account. In this thesis this selection was not optimal. Additionally, the meaning of the features and their interpretability is lost, because the features are translated into PC.

Classification results are similar to the ones that OMNEO obtained where a

much more dedicated and complex method was used . This proves that the classifiers used on this thesis were valid, even though the SVM implementations could benefit from parameter optimization. SVM Functions are highly affected from the parameters given, hence, for better results, these parameters must be optimized.

Additionally, one can see that the Decision Tree, a fully interpretable and white-box method, returns acceptable results, no less than 92% of accuracy (looking at the $\boldsymbol{BR_y}$, Mic and AE sensors individually) and no less than 97% using groups of sensors. The Random Forest method outperformed the Decision Tree method across all groups of sensors and sensors individually, however, as already stated, is similar to a black-box method. The SVM method, through the implementations possible returns the worst results, proving that this method is very sensible to parameter optimization.

With this, one can establish a guideline:

- If the goal is to acquire classification with great accuracy, without caring for model interpretability, then one should use the Random Forest method or the SVM methods, for example, as shown in this document. An alternative, more complex and resource-greedy, is to use Neural Networks, here through the Convolutional Neural Networks method.
- if the goal is to use a *white-box*, easy to interpret model, then the Decision Tree method gives useful and meaningful insights.

Analyzing the groups of sensors and their accuracies (tables 6.24 and 6.25) one can conclude that using the Decision Tree method, the group of sensors that performed best was group number 1, using all sensors. However, similar accuracy is also obtained using either the $\boldsymbol{BR_y}$ sensor or group of sensors number 9. Sensors analyzed during the last chapter ($\boldsymbol{BR_y}$ and Mic) pretty much all benefited from signal processing, the Mic sensor proving this with a clear difference between the results.

The method developed here also presents the importance of applying the step of feature selection, for the results if not better, did not get worse. In on-line situations of permanent monitoring, where space availability can be a problem, the best strategy is to acquire useful information immediately from the measurement, instead of using the entire dataset, even redundant information, using valuable cloud or disk space unnecessarily. Although some characteristics of this step need to be studied, so it should be used carefully.

The last point, the perspective of an individual with strong background in mechanical engineering is important, given that interpreting sensor and signals information, relating the features extracted to how and why they where used to characterise the fault and how they relate to the presence of a fault in the bearing cannot be performed by someone without such knowledge.

7.2 Future Work

Dispite the conclusions and techniques available, there is still room for improvement in this subject.

Further work should be performed in order to apply techniques for pattern recognition so that RUL calculation can be performed. Furthermore, and according to [88] the optimal number of features to use is, in itself, object of further studies. Using fewer features can lead to misrepresentation of the dataset, resulting in bad classification accuracies. However, selecting more features, might end up as using a bigger number of more redundant features, resulting in low accuracies. Hence, the optimal number of features to use, for each method, is a topic for future development. Additionally, the decision tree algorithm should be explored as a feature selection algorithm in addition to studying another method for feature selection.

Frequency domain features can also be studied slightly further. This project used individually the harmonics' amplitude and another way of using these features would be to sum the first three harmonics' amplitude for each fault and assess whether or not it would constitute a good feature.

On the other hand, this data was acquired from a test-rig. Due to time-constraints it was not possible to acquire data that belonged to a *real-world* situation. Hence, applying this method to such a system would be useful: a crane or a wind turbine for instance.

Additionally, it is found in the literature, [24, 25], that PCA was used as a tool for feature dimension reduction and its results, the PC, are used as input for classification methods. Moreover, the optimal number of PC to use (% of information retained) should also be studied.

Furthermore, at the end, a problem of optimisation is proposed. Using a wide range of sensors can be an expensive measure without classification accuracy increase to prove the investment. Given this, the ratio between sensors' cost and classification accuracy should be optimised in order to obtain the best ratio.

However, this method is promising, being a step-further to RUL estimation, and, consequently, a better maintenance plan can be designed for rotating machines

Bibliography

[1] R. B. Randall, *Vibration-based Condition Monitoring*. United Kingdom: John Wiley & Sons, Ltd, 2011.

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

[3] S. Sharma and D. Mahto, "Condition monitoring of wind turbines: A review," *International Journal of Scientific & Engineering Research*, vol. 4, pp. 35–50, 2013.

[4] A. Davies, *Handbook of Condition Monitoring*. United Kingdom: Chapman & Hall, 1998.

[5] P. J. Morais, "Condition monitoring of bearings under low and medium speed rotation," Master's thesis, Faculdade de Engenharia da Universidade do Porto, July 2016.

[6] A. Maurício, "Condition monitoring of gearbox faults with acoustic and vibration signals," Master's thesis, Faculdade de Engenharia da Universidade do Porto, January 2017.

[7] S. Mostafa, S.-H. Lee, J. Dumrak, N. Chileshe, and H. Soltan, "Lean thinking for a maintenance process," *Production & Manufacturing Research*, vol. 3, no. 1, pp. 236–272, 2015.

[8] B. K. N. Rao, "Advances in diagnostic and prognostic strategies and technologies for failure-free maintenance of industrial assets," in *22nd International congress on condition monitoring and diagnostic engineering management (COMADEM)*, 2009.

[9] B. Al-Najjar and I. Alsyouf, "Enhancing a company's profitability and competitiveness using integrated vibration-based maintenance: a case study," *Journal of European Operation Research*, vol. 157, pp. 643–657, 2004.

[10] B. Al-Najjar, "A maintenance model for identification, quantification and elimination of losses in companies profitability: an application example," in *22nd International congress on condition monitoring and diagnostic engineering management (COMADEM)*, 2009.

[11] F. Besnard, J. Nilsson, and L. Bertling, "On the economic benefits of using condition monitoring systems for maintenance management of wind power

systems," in *2010 IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems*, pp. 160–165, 2010.

[12] Y. Wang, "Cost benefit analysis of condition monitoring systems for optimal maintenance decision making," Master's thesis, Wichita State University, May 2013.

[13] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics - a tutorial," *Mechanical Systems and Signal Processing*, vol. 25, pp. 485 – 520, 2010.

[14] D. Ho and R. B. Randall, "Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals," *Mechanical Systems and Signal Processing*, vol. 14, pp. 763 – 788, 2000.

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

[16] R. B. Randall, "A history of cepstrum analysis and its application to mechanical problems," *Mechanical Systems and Signal Processing*, vol. 97, pp. 3 – 19, 2016.

[17] B. Zhang, G. Georgoulas, M. Orchard, A. Saxena, D. Brown, G. Vachtsevanos, and S. Liang, "Rolling element bearing feature extraction and anomaly detection based on vibration monitoring," in *Proceedings of MED'08, 2008 Mediterranean Conference on Control and Automation*, pp. 1792 – 1797, IEEE, 2008.

[18] F. de Assis Boldt, T. W. Rauber, and F. M. Varejão, "A fast feature selection algorithm applied to automatic faults diagnosis of rotating machinery," *Journal of applied computing research*, vol. 3, no. 2, pp. 78 – 86, 2013.

[19] P. Stepanic, I. V. Latinovic, and Z. Djurovic, "A new approach to detection of defects in rolling element bearings based on statistical pattern recognition," *The International Journal of Advanced Manufacturing Technology*, vol. 45, no. 1-2, pp. 91–100, 2009.

[20] E. Bechhoefer and P. Menon, "Bearing envelope analysis window selection," in *Annual Conference of the Prognostics and Health Management Society, 2009*, 2009.

[21] A. Bellini, F. Filippetti, C. Tassoni, and G.-A. Capolino, "Advances in diagnostic techniques for induction machines," *EEE Transactions on Industrial Electronics*, vol. 55, pp. 4109–4126, 2008.

[22] J. B. Ali, L. Saidi, A. Mouelhi, B. Chebel-Morello, and F. Fnaiech, "Application of feature reduction techniques for automatic bearing degradation assessment," in *Proceedings of 2014 International Conference on Electrical Sciences and Technologies in Maghreb (CISTEM)*, pp. 1–6, IEEE, 2014.

[23] N. Sakthivel, B. B. Nair, M. Elangovan, V. Sugumaran, and S. Saravanmurugan, "Comparison of dimensionality reduction techniques for the fault diagnosis of mono block centrifugal pump using vibration signals," *Engineering Science and Technology, an International Journal*, vol. 17, no. 1, pp. 30 – 38, 2014.

[24] F. Wang, J. Sun, D. Yan, S. Zhang, L. Cui, and Y. Xu, "A feature extraction method for fault classification of rolling bearing based on pca," *Journal of Physics: Conference Series*, vol. 628, no. 1, 2015.

[25] M. D. Farrell and R. M. Mersereau, "On the impact of pca dimension reduction for hyperspectral detection of difficult targets," *IEEE Geoscience and Remote Sensing Letters*, vol. 2, no. 2, pp. 192–195, 2005.

[26] T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechak, "Remaining useful life estimation based on nonlinear feature reduction and support vector regression," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 7, pp. 1751–1760, 2013.

[27] A. Sharma, M. Amarnath, and P. Kankar, "Feature extraction and fault severity classification in ball bearings," *Journal of Vibration and Control*, vol. 22, pp. 176 – 192, 2014.

[28] J. Miao and L. Niu, "A survey on feature selection," *Procedia Computer Science*, vol. 91, pp. 919 – 926, 2016.

[29] R. J. Urbanowicz, M. Meeker, W. LaCava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: Introduction and review," *Journal of Machine Learning Research*, 2017.

[30] C. Sobie, C. Freitas, and M. Nicolai, "Simulation-driven machine learning: Bearing fault classification," *Mechanical Systems and Signal Processing*, vol. 99, no. Supplement C, pp. 403 – 419, 2018.

[31] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Journal of Machine Learning*, vol. 46, no. 1, pp. 389–422, 2002.

[32] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufman, fourth ed., 2017.

[33] R. Berk, *Criminal Justice Forecasts of Risk: A Machine Learning Approach*. Springer, 2012.

[34] W. Knight, "Intelligent machines the financial world wants to open ai's black boxes intelligent machines: The financial world wants to open ai's black boxes," April 2017 (Acessed January 2018). goo.gl/pNphNU.

[35] C. Lewis and D. Monett, "Ai & machine learning black boxes: The need for transparency and accountability," April, 2017 (Acessed: January 2018). goo.gl/WERJ1s.

[36] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[37] J. Conyers, "Maintenance bits: A blog on maintenance best practices for industry," September, 2012 (Acessed: January 2018). goo.gl/v9RvXi.

[38] B. J. Hamrock and W. J. Anderson, "Rolling-element bearings," tech. rep., NASA: National Aeronautics and Space Administration, June 1983.

[39] A. R. Mohanty, *Machinery Condition Monitoring: Principles and Practices*. Taylor & Francis Group, LLC, 2015.

[40] H. Ocak and K. A. Loparo, "Estimation of the running speed and bearing defect frequencies of an induction motor from vibration data," *Mechanical Systems and Signal Processing*, vol. 18, no. 3, pp. 515–533, 2004.

[41] S. Orhan, N. Aktürk, and V. çelik, "Vibration monitoring for defect diagnosis of rolling element bearings as a predictive maintenance tool: Comprehensive case studies," *NDT&E International*, vol. 39, no. 4, pp. 293–298, 2006.

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

[43] J. Halme and P. Andersson, "Rolling contact fatigue and wear fundamentals for rolling bearing diagnostics - state of the art," *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, vol. 224, no. 4, pp. 377–393, 2010.

[44] A. Gabelli, J. Lai, T. Lund, K. Rydén, I. Strandell, and G. E. Morales-Espejel, "The fatigue limit of bearing steels – part ii: Characterization for life rating standards," *International journal of fatigue*, vol. 38, no. Supplement C, pp. 169 – 180, 2012.

[45] SKF, "Bearing damage chart," 2012 (Acessed: January 2018). goo.gl/2Csr3m.

[46] FAG, SCHAEFFLER Group, *Rolling Bearing Damage: Recognition of damage and bearing inspection*, 2001 (Acessed: January 2018). goo.gl/XFK54o.

[47] Barden Precision Bearings, *Bearing Failure: Causes And Cures*, Acessed: January 2018. goo.gl/6GhgkX.

[48] T. Tallian, *Failure atlas for Hertz contact machine elements*. ASME Press, 1992.

[49] F. P. G. Márquez, A. M. Tobias, J. M. P. Pérez, and M. Papaelias, "Condition monitoring of wind turbines: Techniques and methods," *Renewable Energy*, vol. 46, pp. 169 – 178, 2012.

[50] W. Zhou, T. G. Habetler, and R. G. Harley, "Bearing condition monitoring methods for electric machines: A general review," in *Proceedings of SDEMPED 2007, IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*, pp. 3–6, IEEE, 2007.

[51] "Condition monitoring and diagnostics of machines – vibration condition monitoring – part 1: General procedures," standard, International Organization for Standardization, Geneva, CH, February 2002.

[52] J. D. Rodrigues, *Apontamentos de Vibrações de Sistemas Mecânicos*. Porto, Portugal: FEUP, 2017.

[53] Z. Stamboliska, E. Rusiński, and P. Moczko, *Proactive condition monitoring of low-speed machines*. Springer, 2015.

[54] C. Aszkler, "Chapter 5 - acceleration, shock and vibration sensors," in *Sensor Technology Handbook* (J. S. Wilson, ed.), pp. 137 – 159, Burlington: Newnes, 2005.

[55] D. Mba and R. B. K. N. Rao, "Development of acoustic emission technology for condition monitoring and diagnosis of rotating machines: Bearings, pumps, gearboxes, engines, and rotating structures," *The Shock and Vibration digest*, vol. 38, pp. 3–16, 2006.

[56] A. Choudhury and N. Tandon, "Application of acoustic emission technique for the detection of defects in rolling element bearings," *Tribology international*, vol. 33, no. 1, pp. 39–45, 2000.

[57] C. Tan, "Application of acoustic emission to the detection of bearing failures," in *Proceedings of International Tribology Conference 1990: Putting Tribology to Work; Reliability and Maintainability through Lubrication and Wear Technology*, Institution of Engineers, Australia, 1990.

[58] W. Zhou, T. G. Habetler, and R. G. Harley, "Bearing condition monitoring methods for electric machines: A general review," in *Proceedings of 2007 IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*, pp. 3–6, IEEE, 2007.

[59] University of St. Andrews and Jim Lesurf, "The sampling theorem: Signal reconstruction," 2000 (Acessed: January 2018). goo.gl/MwHm2P.

[60] P. Burk, L. Polansky, D. Repetto, M. Roberts, and D. Rockmore, "Music and computers: A theoretical and historical approach. chapter 2: The digital representation of sound. section 2.3: Sampling theory," 2011 (Accessed: January, 2018). http://sites.music.columbia.edu/cmc/MusicAndComputers/chapter2/02_03.php.

[61] K. Fyfe and E. Munck, "Analysis of computed order tracking," *Mechanical Systems and Signal Processing*, vol. 11, no. 2, pp. 187 – 205, 1997.

[62] T. Wang, W. Cheng, J. Li, and F. Chu, "A new angular resampling algorithm for the bearing fault diagnosis under the time-varying rotational speed," *Journal of Physics: Conference Series*, vol. 628, no. 1, pp. 12–84, 2015.

[63] M. Zhao, J. Lin, Y. Miao, and X. Xu, "Feature mining and health assessment for gearboxes using run-up/coast-down signals," *Journal of Sensors*, vol. 16, no. 11, 2016.

[64] C. Freitas, J. Cuenca, P. Morais, A. Ompusunggu, M. Sarrazin, and K. Janssens, "Comparison of vibration and acoustic measurements for detection of bearing defects," in *Proceedings of ISMA 2016, Noise and Vibration Engineering Conference*, ISMA, 2016.

[65] A. Ompusunggu and T. A Bartic, "Automated cepstral editing procedure (acep) for removing discrete components from vibration signals," *International Journal of Condition Monitoring*, vol. 6, pp. 56–61, 09 2016.

[66] P. Borghesani, P. Pennacchi, R. Randall, N. Sawalhi, and R. Ricci, "Application of cepstrum pre-whitening for the diagnosis of bearing faults under variable speed conditions," *Mechanical Systems and Signal Processing*, vol. 36, no. 2, pp. 370 – 384, 2013.

[67] R. Randall and N. Sawalhi, "A new method for separating discrete components from a signal," *Sound and Vibration*, vol. 45, pp. 6–9, 05 2011.

[68] R. Randall, N. Sawalhi, and M. Coats, "A comparison of methods for separation of deterministic and random signals," *International Journal of Condition Monitoring*, vol. 1, pp. 11–19, 06 2011.

[69] R. B. Randall and N. Sawalhi, "Cepstral removal of periodic spectral components from time signals," in *Advances in Condition Monitoring of Machinery in Non-Stationary Operations*, pp. 313–324, Springer, 2014.

[70] J. Antoni, "Cyclic spectral analysis of rolling-element bearing signals: facts and fictions," *Journal of Sound and vibration*, vol. 304, no. 3, pp. 497–529, 2007.

[71] R. Randall, J. Antoni, and S. Chobsaard, "The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals," *Mechanical Systems and Signal Processing*, vol. 15, no. 5, pp. 945 – 962, 2001.

[72] G. L. McDonald, "Vibration signal-based fault detection for rotating machines," Master's thesis, Department of Electrical and Computer Engineering , University of Alberta, 2011.

[73] J. Antoni, "Cyclic spectral analysis in practice," *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 597–630, 2007.

[74] L. Debnath and F. A. Shah, "Lecture notes on wavelet transforms," 2017.

[75] F. R. Kschischang, "The hilbert transform," tech. rep., The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Faculty of Applied Science & Engineering, 2006. goo.gl/wgVhd2.

[76] R. Wald, T. Khoshgoftaar, and J. C. Sloan, "Fourier transforms for vibration analysis: A review and case study," in *Proceedings of IEEE International Conference on Information Reuse and Integration (IRI), 2011*, pp. 366–371, IEEE, 2011.

[77] P. K. Rahi and R. Mehra, "Analysis of power spectrum estimation using welch method for various window techniques," *International Journal of Emerging Technologies and Engineering*, vol. 2, no. 6, pp. 106–109, 2014.

[78] I. Howard, "A review of rolling element bearing vibration: detection, diagnosis and prognosis," , Defence Science and Technology Organization, Melbourne (Australia), 1994.

[79] D. P. Doane and L. E. Seward, "Measuring skewness: a forgotten statistic," *Journal of Statistics Education*, vol. 19, no. 2, pp. 1–18, 2011.

[80] A. Bonyár, "Application of localization factor for the detection of tin oxidation with afm," in *Proceedings of IEEE 21st International Symposium for Design and Technology in Electronic Packaging (SIITME), 2015*, pp. 25–30, IEEE, 2015.

[81] H. Bendjama and M. S. Boucherit, "Wavelets and principal component analysis method for vibration monitoring of rotating machinery," *Journal of Theoretical and Applied Mechanics*, vol. 54, no. 2, pp. 659–670, 2016.

[82] L. I. Smith, "A tutorial on principal components analysis," , Department of Computer Science, University of Otago, 2002. goo.gl/fw94bH.

[83] A. J. Izenman, *Linear Discriminant Analysis*. New York, NY: Springer New York, 2008.

[84] P. Xanthopoulos, P. M. Pardalos, and T. B. Trafalis, *Robust data mining*. Springer Science & Business Media, 2012.

[85] X. Jin, M. Zhao, T. W. S. Chow, and M. Pecht, "Motor bearing fault diagnosis using trace ratio linear discriminant analysis," *IEEE Transactions on Industrial Electronics*, vol. 61, pp. 2441–2451, May 2014.

[86] S. Raschka, "Machine learning and data science blog: Linear discriminant analysis - bit by bit," 2014 (Accessed: January, 2018). goo.gl/BsVCvB.

[87] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, *Feature extraction: foundations and applications*, vol. 207. Springer, 2008.

[88] M. Amarnath, V. Sugumaran, and H. Kumar, "Exploiting sound signals for fault diagnosis of bearings using decision tree," *Measurement*, vol. 46, no. 3, pp. 1250–1256, 2013.

[89] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[90] V. Sugumaran, V. Muralidharan, and K. Ramachandran, "Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing," *Mechanical systems and signal processing*, vol. 21, no. 2, pp. 930–942, 2007.

[91] M. Saimurugan, K. Ramachandran, V. Sugumaran, and N. Sakthivel, "Multi component fault diagnosis of rotational mechanical system based on decision tree and support vector machine," *Expert Systems with Applications*, vol. 38, no. 4, pp. 3819–3826, 2011.

[92] L. E. Raileanu and K. Stoffel, "Theoretical comparison between the gini index and information gain criteria," *Annals of Mathematics and Artificial Intelligence*, vol. 41, no. 1, pp. 77–93, 2004.

[93] M. Tan, Pang-Ning and Steinbach and V. Kumar, "Classification: basic concepts, decision trees, and model evaluation," *Introduction to data mining*, vol. 1, pp. 145–205, 2006.

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

[95] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Journal of Pattern Recognition Letters*, vol. 27, no. 4, pp. 294–300, 2006.

[96] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

[97] S. R. Gunn, "Support vector machines for classification and regression," , Image Speech and Intelligent Systems Research Group, University of Southampton, 1997. goo.gl/nnjk5B.

[98] R. Ruiz-Gonzalez, J. Gomez-Gil, F. J. Gomez-Gil, and V. Martínez-Martínez, "An svm-based classifier for estimating the state of various rotating components in agro-industrial machinery with a vibration signal acquired from a single point on the machine chassis," *Journal of Sensors*, vol. 14, no. 11, pp. 20713–20735, 2014.

[99] E. Kim, "Everything you wanted to know about the kernel trick (but were too afraid to ask)," 2013 (Accessed: January, 2018). goo.gl/WQMCiS.

[100] Cambridge Spark - Data Science online courses, "Module 5: Support vector machines," 2016 (Accessed: January, 2018). goo.gl/DNknBM.

[101] D.-T. Hoang and H.-J. Kang, "Convolutional neural network based bearing fault diagnosis," in *Proceedings of International Conference on Intelligent Computing, 2017*, pp. 105–111, Springer, 2017.

[102] A. Bronshtein, "Towards data science: Train/test split and cross validation in python," 2017 (Accessed: January, 2018). goo.gl/vDfzot.

APPENDIX A

Feature Reduction

A.1 PCA

BR_x sensor

Again, Figures A.1 and A.2 show PCA analysis applied on non-filtered signal and Filtered Signal, respectively, for this sensor.

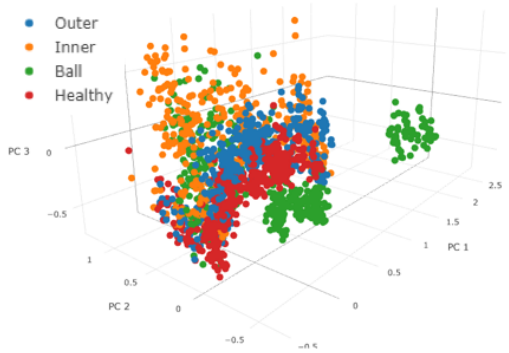


Figure A.1: Non-filtered Signal.

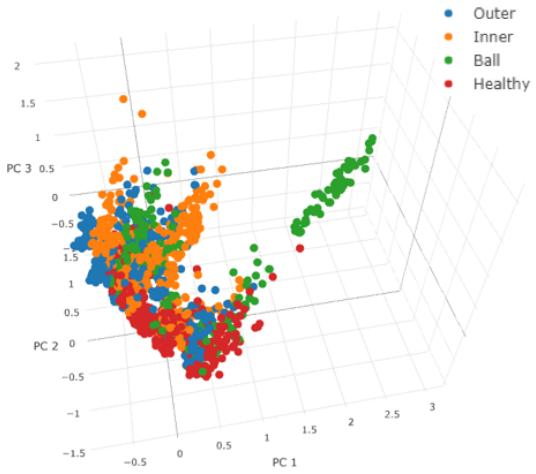


Figure A.2: Filtered Signal.

Comment 1. Regarding PCA analysis of the BR_x signal, it can be concluded that this sensor did not benefit from filtering, given that class separation did not improve. Additionally, using the first three principal components amounts to a representation of the total data of 65,92% and 55,02%, respectively for non-filtered signal and for filtered signal.

BR_z sensor

Figures A.3 and A.4 show PCA analysis applied on non-filtered signal and filtered Signal, respectively.

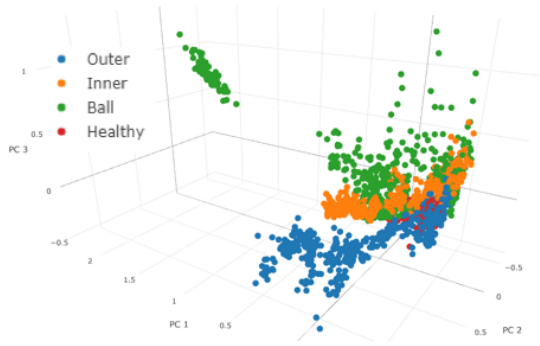


Figure A.3: Non-filtered Signal.

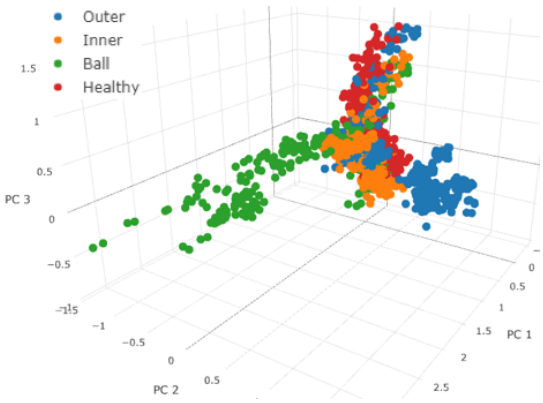


Figure A.4: Filtered Signal.

Comment 2. Filtering on BR_z sensor did not benefit class separation as well. Additionally, for this sensor it was found that using the first three principal components amounts to 74,74% and 51,27% of the total data, respectively for the non-filtered signal and filtered signal. Again, analyses using the principal components for this sensor should use more than the first four principal components.

A.2 LDA

BR_x sensor

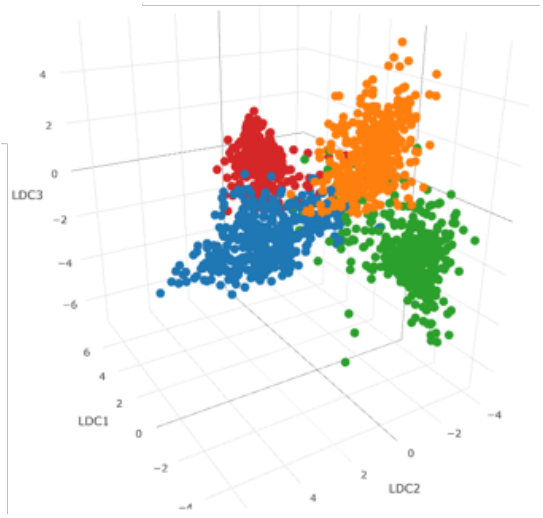


Figure A.5: Non-filtered signal.

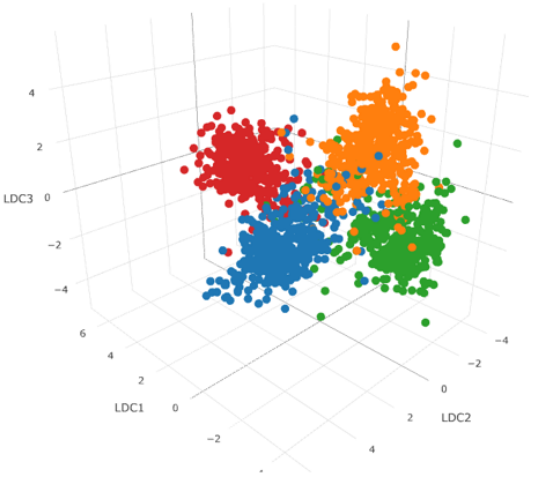


Figure A.6: Filtered signal.

Comment 1. For the BR_x sensor, comparing both signals, one can say that class separation after filtering remains acceptable.

BR_z sensor

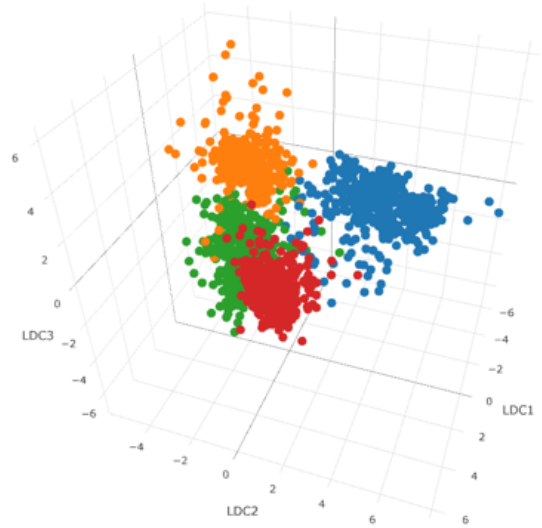


Figure A.7: Non-filtered signal.

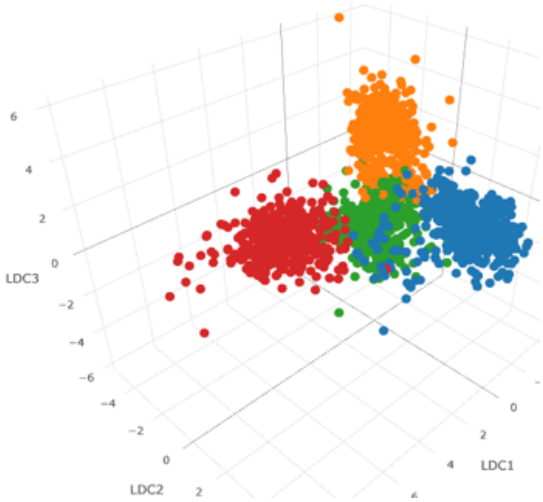


Figure A.8: Filtered signal.

Comment 2. For the $\boldsymbol{BR_z}$ signal, the groups are closer together on the filtered signal than on the non-filtered signal. However, class separation is still possible.

APPENDIX B

Correlation Matrices

Figure B.1: Correlation matrix of $\boldsymbol{BR_y}$ and Mic Sensors, before.

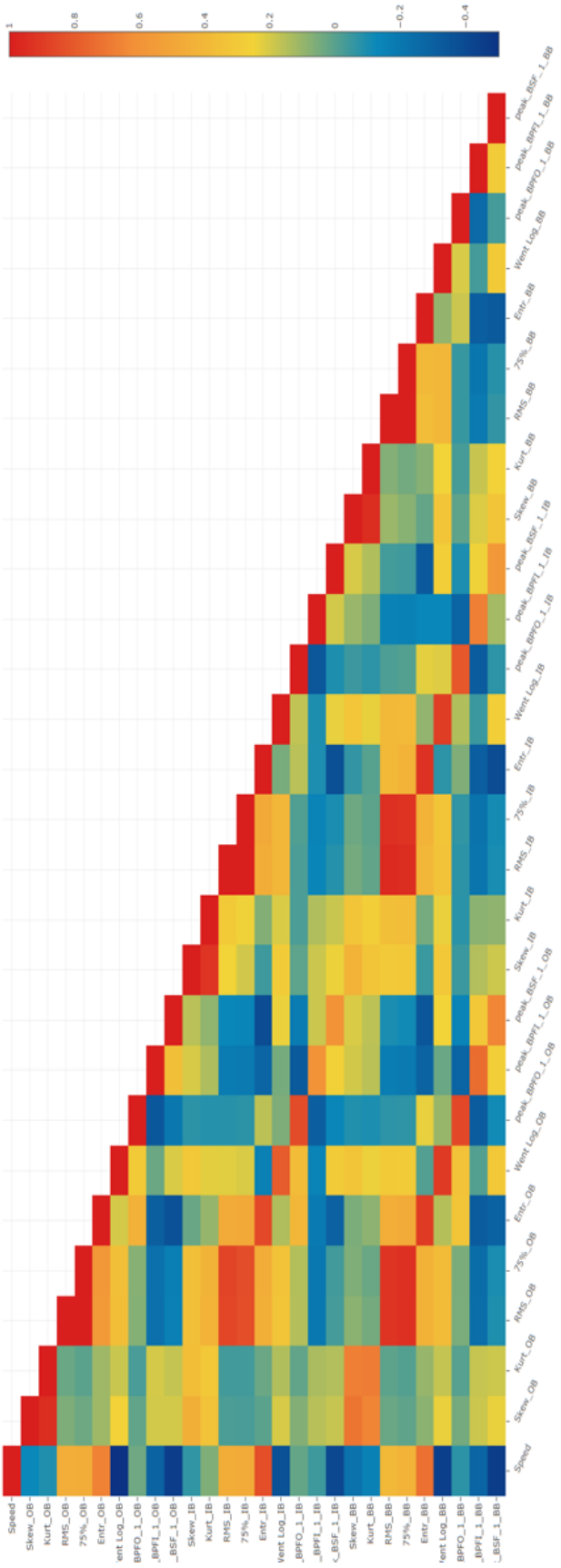


Figure B.2: Correlation matrix of $\boldsymbol{BR_y}$ and Mic Sensors, after.

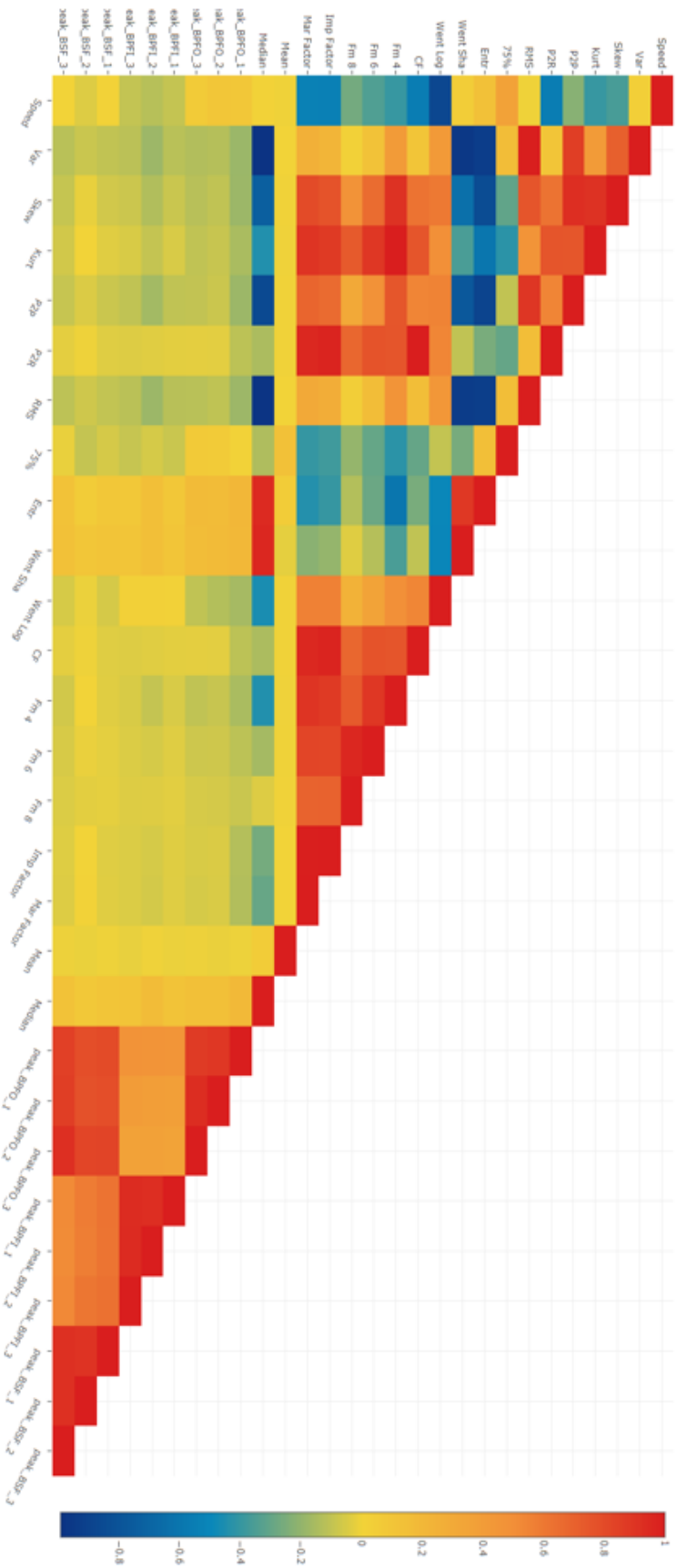


Figure B.3: Correlation matrix of AE Sensor, before.

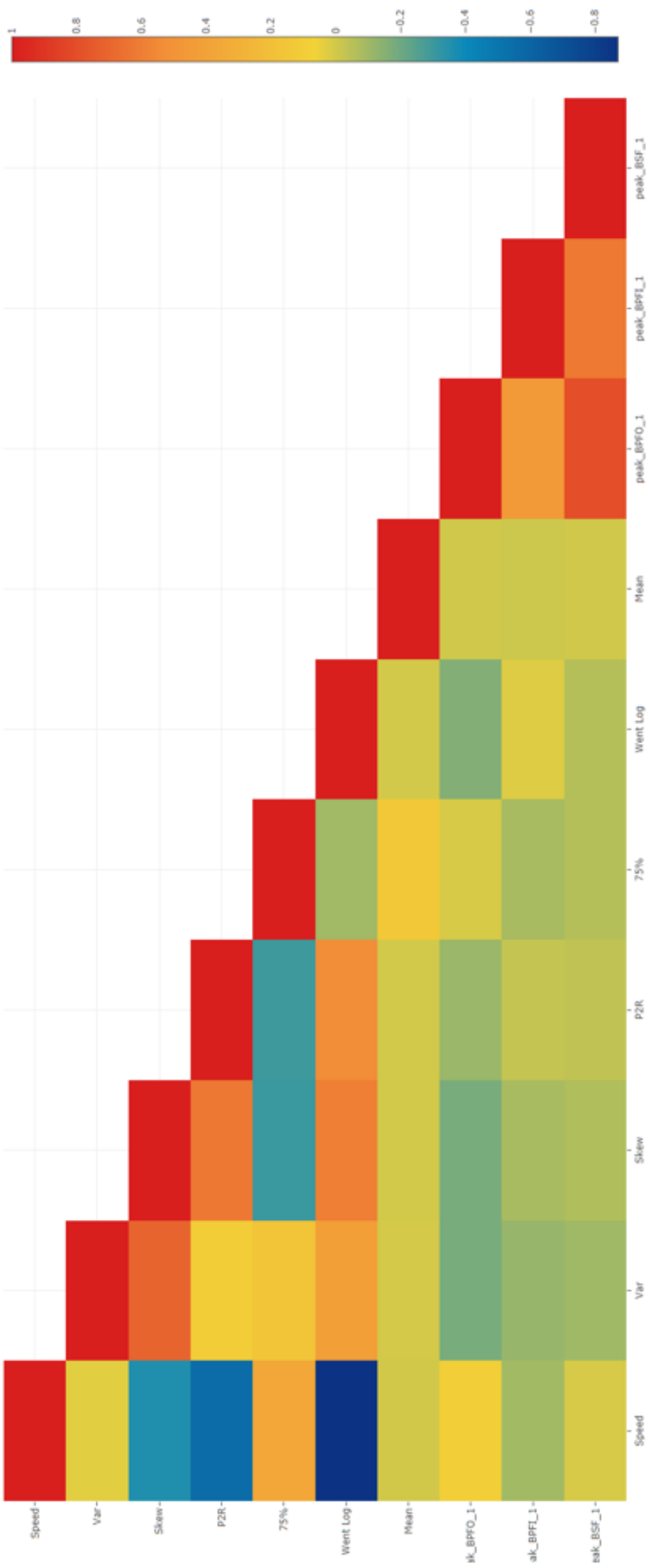


Figure B.4: Correlation matrix of AE Sensor, after.

APPENDIX C

Decision Trees modelling the Classification problem

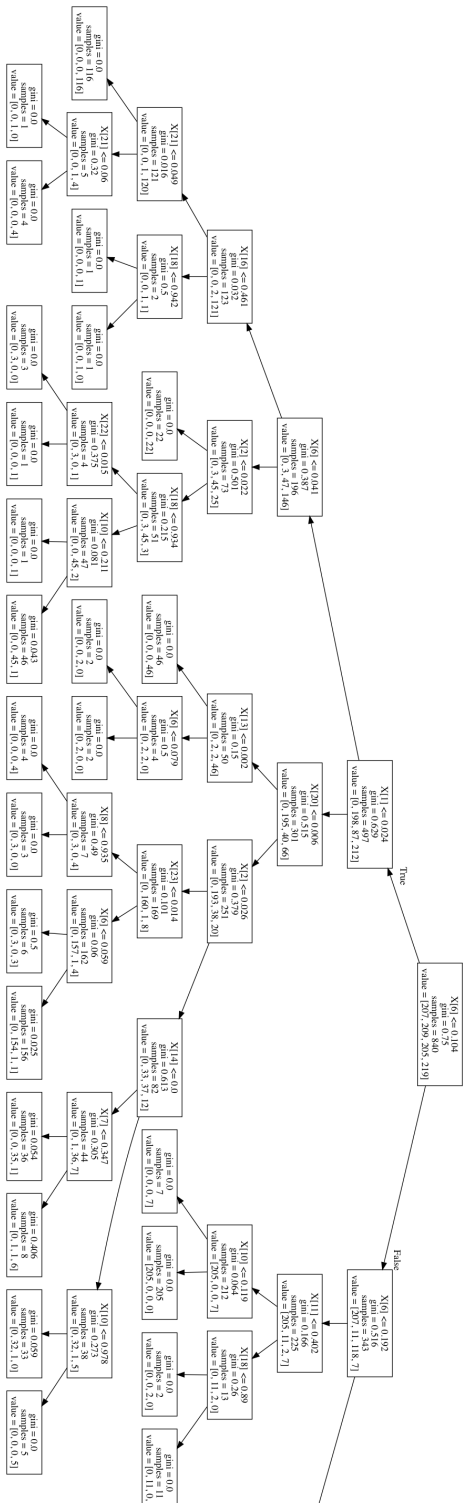


Figure C.1: Decision Tree modelling REB fault classification without feature selection, AE sensor.

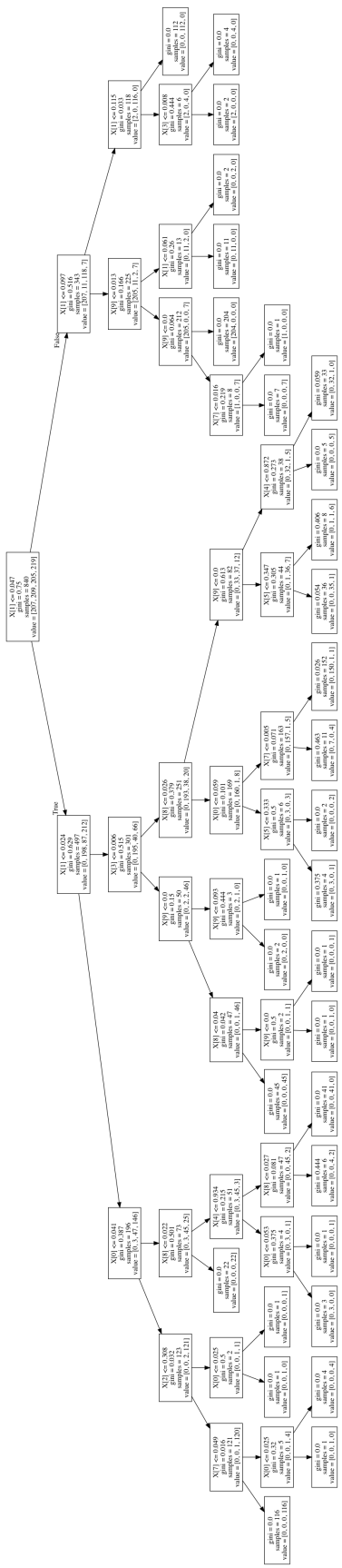


Figure C.2: Decision Tree modelling REB fault classification with feature selection, AE sensor.

Figure C.3: Decision Tree modelling REB fault classification without feature selection and non-filtered signal, Mic sensor.

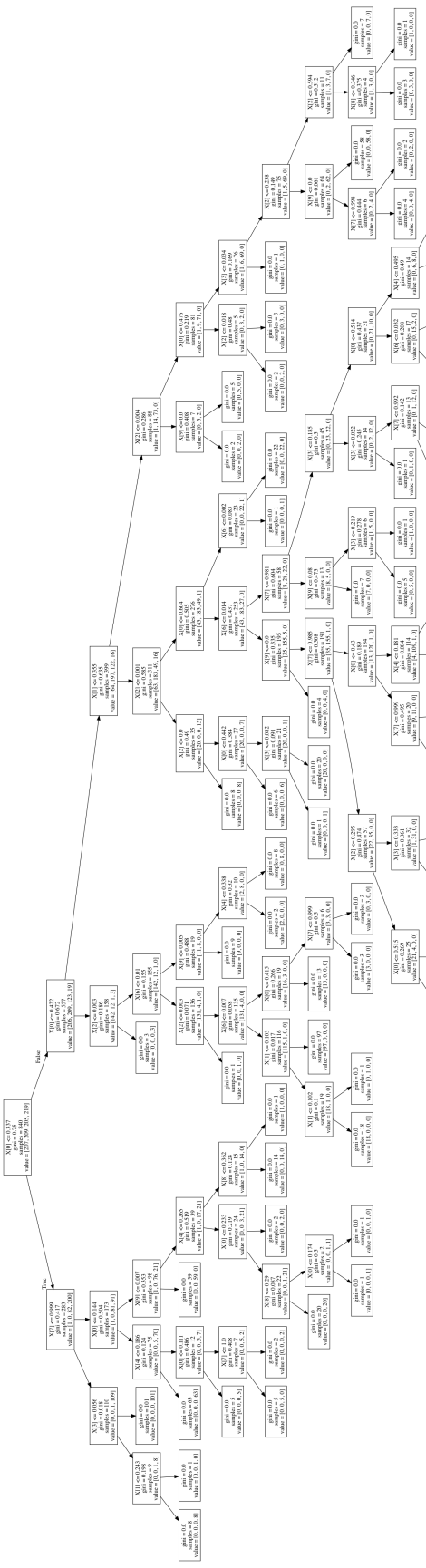


Figure C.4: Decision Tree modelling REB fault classification with feature selection and non-filtered signal, Mic sensor.

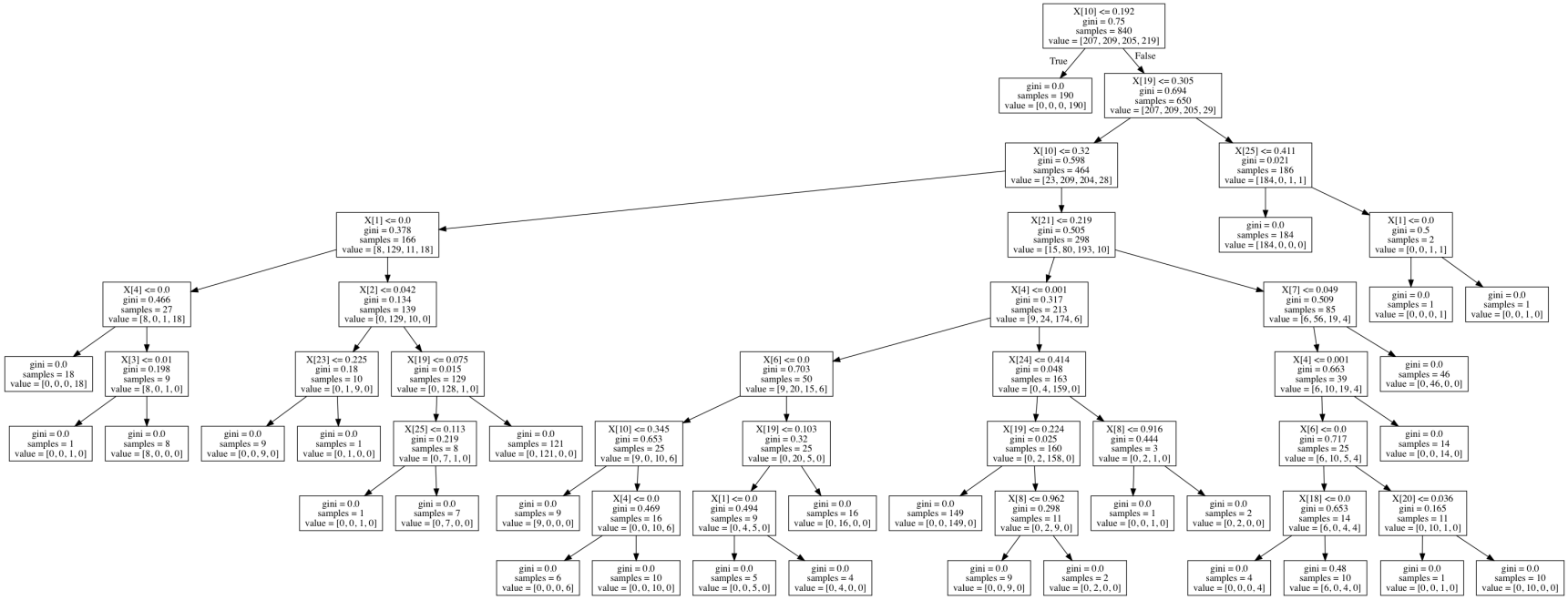


Figure C.5: Decision Tree modelling REB fault classification without feature selection and non-filtered signal, $\boldsymbol{BR_y}$ sensor.

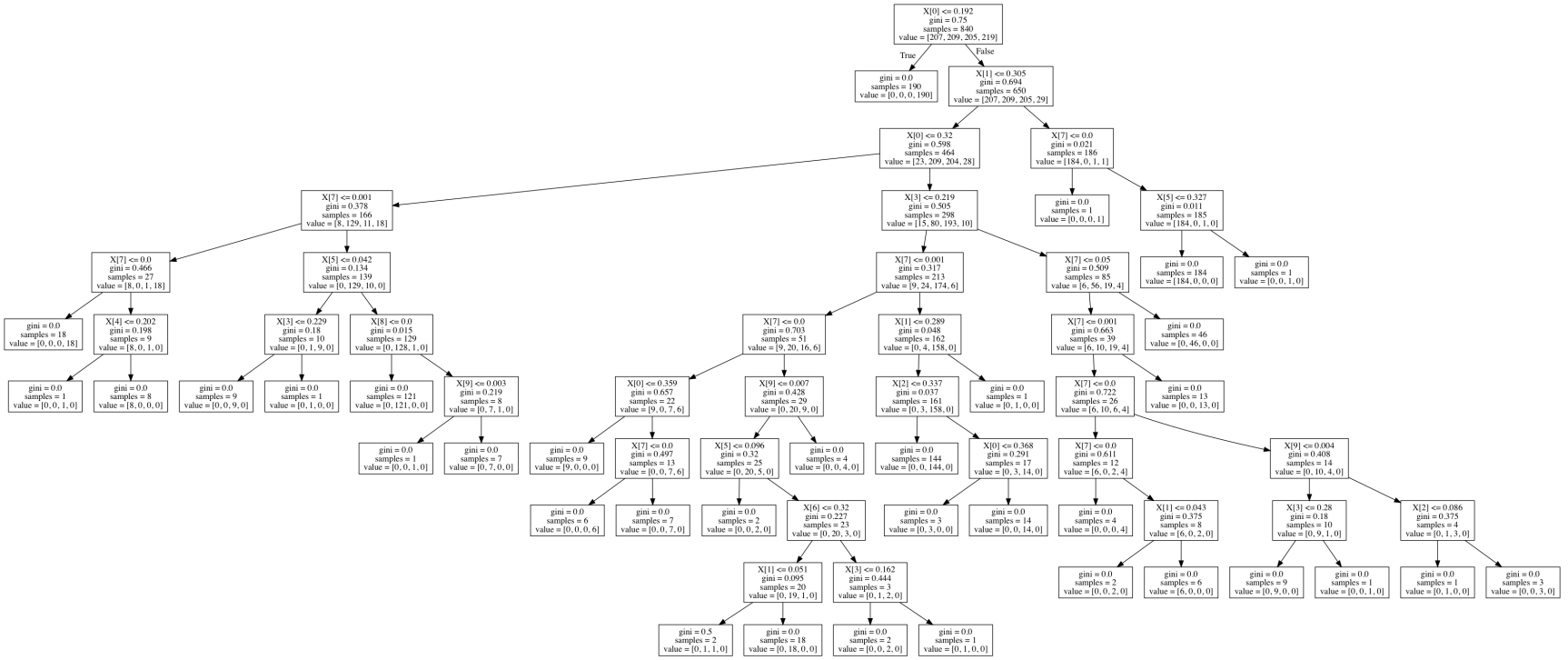


Figure C.6: Decision Tree modelling REB fault classification with feature selection and non-filtered signal, $\boldsymbol{BR_y}$ sensor.

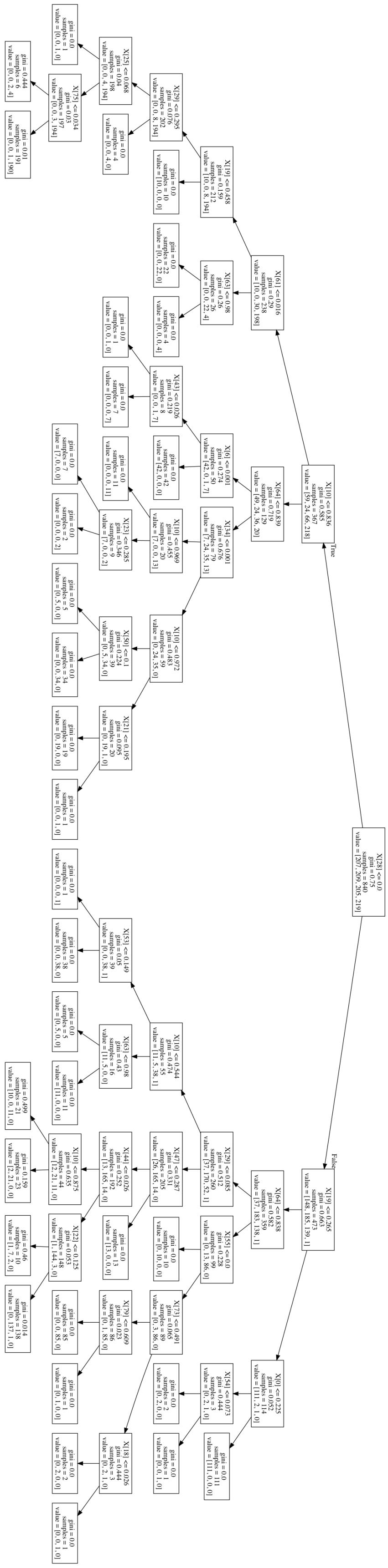


Figure C.7: Decision Tree modelling REB fault classification without feature selection and filtered signal, Mic sensor.

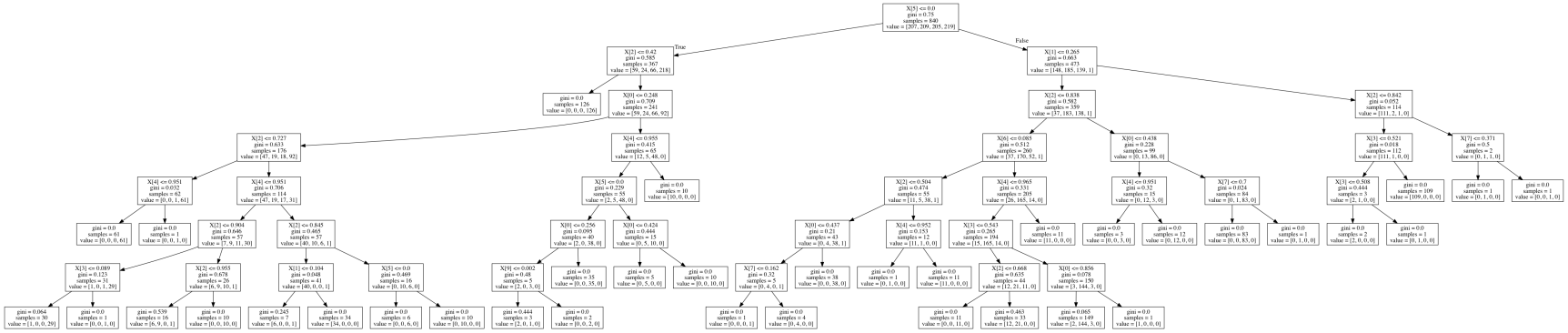


Figure C.8: Decision Tree modelling REB fault classification with feature selection and filtered signal, Mic sensor.

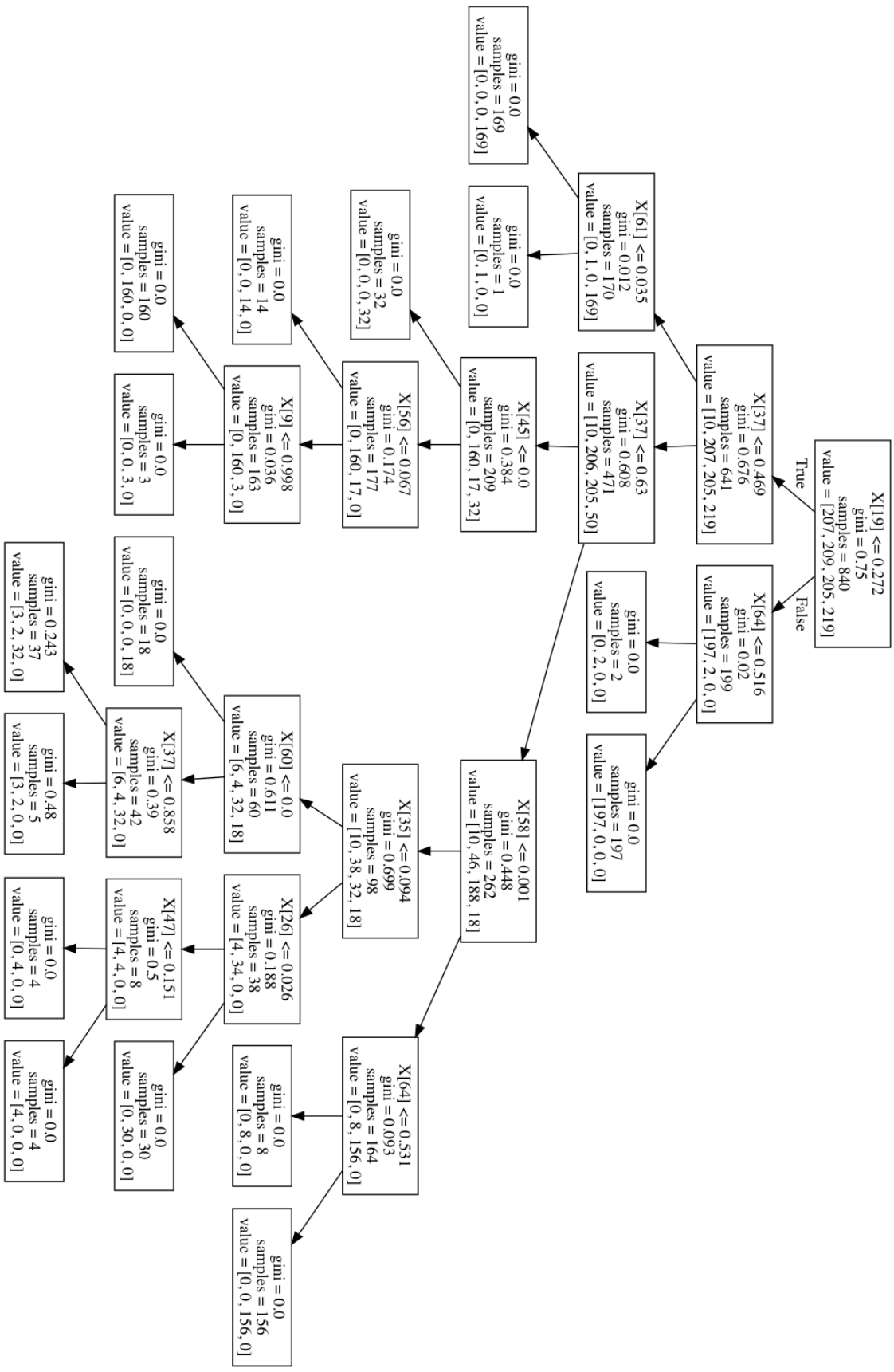


Figure C.9: Decision Tree modelling REB fault classification without feature selection and filtered signal, $\boldsymbol{BR_y}$ sensor.

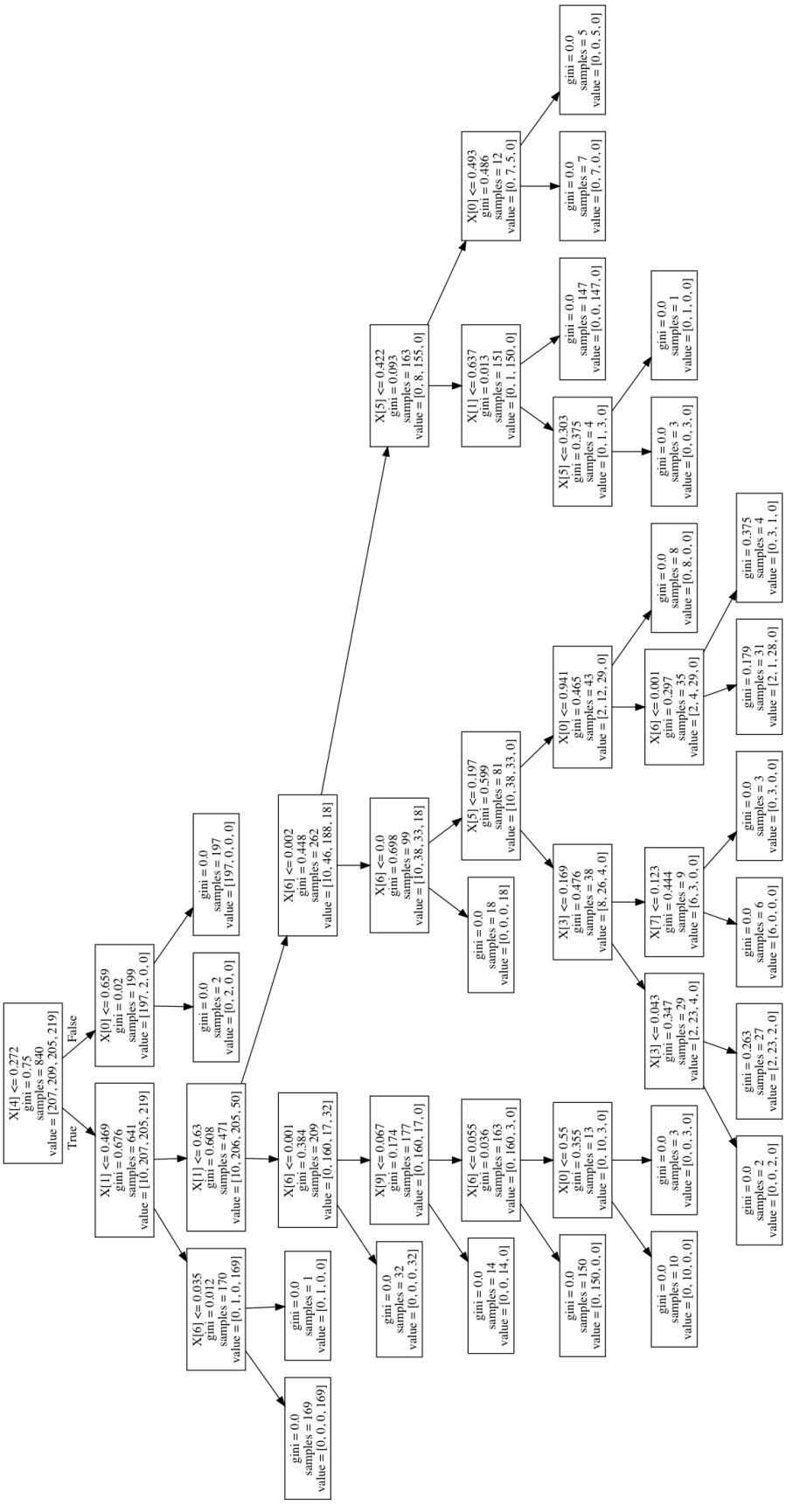


Figure C.10: Decision Tree modelling REB fault classification with feature selection and filtered signal, $\boldsymbol{BR_y}$ sensor.

APPENDIX D

Classification Results: No Feature Selection

This part of the document will list all results obtained, with and without feature selection and comparing Non-filtered and filtered signals. This analysis however will only cover the sensor that best benefited from filtering (shown by PCA analysis), $\boldsymbol{BR_y}$ sensor, along with the Mic and AE sensors for their instrumentalization benefits.

D.1 Non-filtered Signal

Table D.1: Accuracy scores for Decision Tree: Non-filtered Signal and No Feature Selection.

	Accuracy Scores /% in rounds $R_{1...10}$									
BR_Y	93,02	93,02	94,11	91,76	95,18	93,97	92,77	95,18	92,77	92,77
BR_{AE}	94,18	98,83	91,76	85,88	93,97	95,18	85,54	91,56	96,39	90,36
BR_{Mic}	87,20	86,04	89,41	90,58	91,56	77,10	83,13	81,92	86,74	83,13

Table D.2: Accuracy scores for Random Forest: Non-filtered Signal and No Feature Selection.

	Accuracy Scores /% in rounds $R_{1...10}$									
BR_Y	97,02	97,02	98,21	95,83	98,81	95,23	97,61	97,02	96,42	95,83
BR_{AE}	98,21	98,21	99,40	98,21	98,81	98,21	97,61	98,21	96,42	97,02
BR_{Mic}	90,47	94,04	92,26	89,28	96,42	91,66	93,45	93,45	91,66	91,66

Table D.3: Accuracy scores for SVM.SVC: Non-filtered Signal and No Feature Selection.

	Accuracy Scores /% in rounds $R_{1...10}$									
BR_Y	88,69	88,09	89,28	86,31	94,04	88,09	94,64	89,28	85,71	88,09
BR_{AE}	53,57	54,76	60,71	62,5	58,33	57,14	55,35	54,76	53,57	57,14
BR_{Mic}	70,23	75,00	70,83	72,02	73,81	75,00	75,00	69,64	68,45	69,04

Table D.4: Accuracy scores for SVM.LinearSVC: Non-filtered Signal and No Feature Selection.

	Accuracy Scores /% in rounds $R_{1...10}$									
BR_Y	91,66	89,88	94,04	94,05	95,24	92,86	96,43	93,45	89,88	94,05
BR_{AE}	79,16	79,76	85,11	81,55	85,12	83,93	79,17	79,76	81,55	83,33
BR_{Mic}	79,76	79,16	77,38	81,55	86,31	81,55	80,36	80,95	76,79	76,79

Table D.5: Accuracy scores for SVM.NuSVC: Non-filtered Signal and No Feature Selection.

	Accuracy Scores /% in rounds $R_{1...10}$									
BR_Y	89,88	89,88	92,26	86,31	95,83	89,88	95,23	88,69	89,28	89,28
BR_{AE}	73,21	75,59	82,14	79,16	80,35	82,14	75,59	74,40	79,16	79,76
BR_{Mic}	78,57	83,33	84,52	79,76	86,31	80,35	83,33	79,16	77,38	78,57

Table D.6: Confusion Matrix /% for Decision Tree and No Feature Selection: $\boldsymbol{BR_y}$ - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	96,04	0,89	1,96	0,95
Inner	1,49	91,96	7,35	0,00
Ball	1,49	6,70	89,71	1,43
Healthy	0,99	0,00	2,94	96,19

Table D.7: Confusion Matrix /% for Random Forest and No Feature Selection: $\boldsymbol{BR_y}$ - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	98,81	0,00	0,48	0,71
Inner	4,76	97,14	2,86	0,00
Ball	1,67	4,76	92,86	0,71
Healthy	0,95	0,00	0,24	98,81

Table D.8: Confusion Matrix /% for SVM.SVC and No Feature Selection: $\boldsymbol{BR_y}$ - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	88,57	5,95	0,48	5,00
Inner	0,00	87,62	12,38	0,00
Ball	0,00	11,19	87,62	1,19
Healthy	0,95	5,95	0,00	93,10

Table D.9: Confusion Matrix /% for SVM.LinearSVC and No Feature Selection: $\boldsymbol{BR_y}$ - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	96,19	1,67	0,00	2,14
Inner	1,43	85,48	13,1	0,00
Ball	0,71	4,29	94,52	0,48
Healthy	0,95	2,62	0,00	96,43

Table D.10: Confusion Matrix /% for SVM.NuSVC and No Feature Selection: $\boldsymbol{BR_y}$ - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	84,76	10,48	0,00	4,76
Inner	0,00	96,43	3,57	0,00
Ball	0,00	11,19	88,33	0,48
Healthy	0,48	6,43	0,00	93,10

Table D.11: Confusion Matrix /% for Decision Tree and No Feature Selection: AE - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	98,02	0,45	0,98	0,48
Inner	0,50	92,41	3,92	3,81
Ball	0,00	4,46	91,18	3,81
Healthy	0,00	6,7	4,9	88,10

Table D.12: Confusion Matrix /% for Random Forest and No Feature Selection: AE - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	99,52	0,00	0,24	0,24
Inner	0,48	99,05	0,24	0,24
Ball	0,24	1,43	96,9	1,43
Healthy	0,00	2,86	0,48	96,67

Table D.13: Confusion Matrix /% for SVM.SVC and No Feature Selection: AE - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	82,14	0,48	0,00	17,38
Inner	26,67	2,86	0,00	70,48
Ball	6,67	0,00	49,29	44,05
Healthy	3,81	0,00	0,95	92,86

Table D.14: Confusion Matrix /% for SVM.LinearSVC and No Feature Selection: AE - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	98,81	1,19	0,00	0,00
Inner	1,43	88,1	0,24	10,24
Ball	5,00	0,95	55,71	38,33
Healthy	2,86	10,95	1,43	84,76

Table D.15: Confusion Matrix /% for SVM.NuSVC and No Feature Selection: AE - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	96,90	2,62	0,24	0,24
Inner	2,86	76,43	0,00	20,71
Ball	5,48	1,19	50,24	43,10
Healthy	3,81	6,90	0,24	89,05

Table D.16: Confusion Matrix /% for Decision Tree and No Feature Selection: Mic - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	81,68	13,84	1,47	1,43
Inner	8,42	87,95	4,90	0,00
Ball	3,96	6,70	83,33	5,24
Healthy	2,97	0,00	7,84	89,52

Table D.17: Confusion Matrix /% for Random Forest and No Feature Selection: Mic - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	93,33	4,76	0,71	1,19
Inner	6,19	90,24	3,57	0,00
Ball	1,67	4,52	91,67	2,14
Healthy	2,14	0,24	3,10	94,52

Table D.18: Confusion Matrix /% for SVM.SVC and No Feature Selection: Mic - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	68,81	26,90	2,38	1,9
Inner	15,48	75,48	9,05	0,00
Ball	3,57	12,38	48,10	35,95
Healthy	3,81	0,00	0,95	95,24

Table D.19: Confusion Matrix /% for SVM.LinearSVC and No Feature Selection: Mic - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	72,14	19,76	3,81	4,29
Inner	15,71	75,71	8,57	0,00
Ball	4,52	7,38	76,43	11,67
Healthy	2,62	0,48	0,95	95,95

Table D.20: Confusion Matrix /% for SVM.NuSVC and No Feature Selection: Mic - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	82,86	10,95	4,29	1,90
Inner	6,19	83,10	10,71	0,00
Ball	1,43	9,76	63,57	25,24
Healthy	3,81	0,00	1,19	95,00

D.2 Filtered Signal

Table D.21: Accuracy scores for Decision Tree: Filtered Signal and No Feature Selection.

	Accuracy Scores /% in rounds $R_{1...10}$									
BR_Y	97,66	96,47	96,47	96,47	96,47	97,62	96,43	96,49	96,39	97,53
BR_{AE}	95,29	96,47	95,29	90,59	90,59	92,86	89,29	87,95	91,57	92,59
BR_{Mic}	87,06	92,94	90,59	90,59	95,29	92,86	95,24	92,77	87,95	100,00

Table D.22: Accuracy scores for Random Forest: Filtered Signal and No Feature Selection.

	Accuracy Scores /% in rounds $R_{1...10}$									
BR_Y	98,81	100,00	98,21	98,21	99,41	98,21	99,41	98,81	98,81	98,81
BR_{AE}	97,62	97,62	98,21	98,81	97,62	97,62	98,21	95,83	98,21	93,45
BR_{Mic}	98,81	98,21	98,81	98,21	97,02	98,81	97,02	98,21	98,81	98,21

Table D.23: Accuracy scores for SVM.SVC: Filtered Signal and No Feature Selection.

	Accuracy Scores /% in rounds $R_{1...10}$									
BR_Y	95,83	97,02	95,83	95,24	95,24	94,64	95,83	96,43	95,83	92,86
BR_{AE}	80,95	74,41	81,55	79,76	74,41	81,55	76,79	73,81	70,83	79,85
BR_{Mic}	95,24	96,43	93,45	90,48	94,05	94,64	94,05	92,26	89,89	93,45

Table D.24: Accuracy scores for SVM.LinearSVC: Filtered Signal and No Feature Selection.

	Accuracy Scores /% in rounds $R_{1...10}$									
BR_Y	96,43	97,02	96,43	95,24	95,83	97,02	95,24	96,43	94,05	95,24
BR_{AE}	82,74	80,36	85,12	83,33	85,12	80,36	82,14	84,52	76,79	75,60
BR_{Mic}	94,05	94,05	91,07	91,67	94,05	91,67	92,26	92,86	89,29	89,88

Table D.25: Accuracy scores for SVM.NuSVC: Filtered Signal and No Feature Selection.

	Accuracy Scores /% in rounds $R_{1...10}$									
BR_Y	90,48	91,07	91,67	89,29	92,86	86,31	91,07	89,29	86,91	86,91
BR_{AE}	83,33	74,41	82,14	77,38	79,76	73,81	81,55	76,19	74,41	70,83
BR_{Mic}	90,48	86,31	88,09	85,12	90,48	86,91	89,29	86,91	80,36	88,09

Table D.26: Confusion Matrix /% for Decision Tree: Filtered Signal and No Feature Selection, $\boldsymbol{BR_y}$.

	Outer	Inner	Ball	Healthy
Outer	96,14	0,96	2,93	0,00
Inner	1,93	95,69	1,95	0,46
Ball	2,42	0,96	95,61	0,91
Healthy	0,00	0,00	0,49	99,54

Table D.27: Confusion Matrix /& for Random Forest: Filtered Signal and No Feature Selection, $\boldsymbol{BR_y}$.

	Outer	Inner	Ball	Healthy
Outer	98,81	0,48	0,48	0,24
Inner	0,24	99,52	0,24	0,00
Ball	0,95	0,95	97,62	0,48
Healthy	0,24	0,24	0,00	99,52

Table D.28: Confusion Matrix /% for Random Forest: Filtered Signal and No Feature Selection, $\boldsymbol{BR_y}$.

	Outer	Inner	Ball	Healthy
Outer	99,29	0,48	0,24	0,00
Inner	0,95	95,0	4,05	0,00
Ball	0,24	6,67	91,19	1,90
Healthy	0,48	0,95	2,14	96,43

Table D.29: Confusion Matrix /% for SVM.LinearSVC: Filtered Signal and No Feature Selection, $\boldsymbol{BR_y}$.

	Outer	Inner	Ball	Healthy
Outer	99,29	0,71	0,00	0,00
Inner	0,95	95,71	2,86	0,48
Ball	0,24	5,24	91,67	2,86
Healthy	0,71	1,19	1,19	96,90

Table D.30: Confusion Matrix /% for SVM.NuSVC: Filtered Signal and No Feature Selection, $\boldsymbol{BR_y}$.

	Outer	Inner	Ball	Healthy
Outer	86,19	9,76	0,71	3,33
Inner	0,00	87,38	11,67	0,95
Ball	0,00	8,33	88,33	3,33
Healthy	0,00	3,57	0,00	96,43

Table D.31: Confusion Matrix /% for Decision Tree: Filtered Signal and No Feature Selection, AE.

	Outer	Inner	Ball	Healthy
Outer	98,07	0,00	0,98	0,91
Inner	0,48	88,52	3,41	7,31
Ball	0,00	2,39	92,20	5,02
Healthy	0,00	7,18	2,93	90,41

Table D.32: Confusion Matrix /% for Random Forest: Filtered Signal and No Feature Selection, AE.

	Outer	Inner	Ball	Healthy
Outer	100,00	0,00	0,00	0,00
Inner	0,24	97,38	0,71	1,67
Ball	0,24	1,90	94,76	3,10
Healthy	0,00	1,90	0,95	97,14

Table D.33: Confusion Matrix /% for SVM.SVC: Filtered Signal and No Feature Selection, AE.

	Outer	Inner	Ball	Healthy
Outer	96,90	2,62	0,48	0,00
Inner	1,90	81,19	0,00	16,90
Ball	5,24	2,86	50,00	41,90
Healthy	3,57	13,57	0,71	82,14

Table D.34: Confusion Matrix /% for SVM.LinearSVC: Filtered Signal and No Feature Selection, AE.

	Outer	Inner	Ball	Healthy
Outer	98,81	1,19	0,00	0,00
Inner	1,43	88,57	0,24	9,76
Ball	5,00	0,71	55,48	38,81
Healthy	3,33	11,67	1,43	83,57

Table D.35: Confusion Matrix /% for SVM.NuSVC: Filtered Signal and No Feature Selection,
acAE.

	Outer	Inner	Ball	Healthy
Outer	97,14	2,38	0,48	0,00
Inner	3,57	77,62	0,00	18,81
Ball	5,48	1,43	50,00	43,10
Healthy	1,19	4,29	9,76	84,76

Table D.36: Confusion Matrix /% for Decision Tree: Filtered Signal and No Feature Selection, Mic.

	Outer	Inner	Ball	Healthy
Outer	91,79	5,74	0,98	1,37
Inner	2,90	93,78	3,41	0,00
Ball	2,42	3,83	89,27	4,11
Healthy	1,93	0,48	2,93	94,98

Table D.37: Confusion Matrix /%for Random Forest: Filtered Signal and No Feature Selection, Mic.

	Outer	Inner	Ball	Healthy
Outer	97,62	2,14	0,00	0,24
Inner	0,95	98,57	0,48	0,00
Ball	0,00	0,95	98,57	0,48
Healthy	0,95	0,00	0,95	98,10

Table D.38: Confusion Matrix /% for SVM.SVC: Filtered Signal and No Feature Selection, Mic.

	Outer	Inner	Ball	Healthy
Outer	94,76	3,10	1,43	0,71
Inner	3,57	94,29	2,14	0,00
Ball	1,90	11,90	86,19	0,00
Healthy	1,67	0,00	0,00	98,33

Table D.39: Confusion Matrix /% for SVM.LinearSVC: Filtered Signal and No Feature Selection, Mic.

	Outer	Inner	Ball	Healthy
Outer	93,33	3,57	1,90	1,19
Inner	3,81	93,33	2,86	0,00
Ball	3,33	11,90	83,33	1,43
Healthy	1,19	0,00	0,48	98,33

Table D.40: Confusion Matrix /% for SVM.NuSVC: Filtered Signal and No Feature Selection, Mic.

	Outer	Inner	Ball	Healthy
Outer	76,90	7,38	3,57	12,14
Inner	2,86	94,76	2,14	0,24
Ball	2,62	11,19	77,14	9,05
Healthy	0,00	0,00	0,00	100,00

APPENDIX E

Classification Results: Automatic Feature Selection (mRMR)

E.1 Non-filtered Signal

Table E.1: Accuracy scores for Decision Tree: Non-filtered Signal and mRMR Feature Selection.

	Accuracy Scores /% in rounds $R_{1...10}$									
BR_Y	95,35	94,19	95,29	92,94	92,77	93,98	93,98	97,59	92,77	92,77
BR_{AE}	95,35	98,84	94,12	91,76	95,18	93,98	89,16	91,57	93,98	90,36
BR_{Mic}	87,21	90,70	97,65	88,240	87,95	83,13	91,57	89,16	84,34	90,36

Table E.2: Accuracy scores for Random Forest: Non-filtered Signal and mRMR Feature Selection.

	Accuracy Scores /% in rounds $R_{1...10}$									
BR_Y	94,64	95,83	96,43	95,83	97,62	96,43	95,24	95,83	94,64	97,02
BR_{AE}	97,02	95,83	98,21	95,83	96,43	95,24	96,43	94,64	95,83	96,43
BR_{Mic}	83,93	95,83	91,07	89,29	94,05	89,29	91,07	91,67	91,07	90,48

Table E.3: Accuracy scores for SVM.SVC: Non-filtered Signal and mRMR Feature Selection.

	Accuracy Scores /% in rounds $R_{1...10}$									
BR_Y	84,52	79,76	85,71	80,95	92,26	83,93	92,86	86,31	80,95	88,09
BR_{AE}	64,29	64,29	71,43	67,86	64,88	67,26	64,29	63,09	60,71	69,05
BR_{Mic}	67,26	67,86	64,88	68,45	65,48	67,26	69,64	60,12	64,29	65,48

Table E.4: Accuracy scores for SVM.LinearSVC: Non-filtered Signal and mRMR Feature Selection.

	Accuracy Scores /% in rounds $R_{1...10}$									
BR_Y	86,91	82,14	88,09	85,71	92,86	85,12	92,26	92,86	83,33	91,07
BR_{AE}	72,02	73,81	78,57	74,41	79,17	77,98	77,38	72,02	75,00	76,19
BR_{Mic}	67,86	67,86	65,48	70,83	70,83	68,45	72,02	66,67	63,09	66,07

Table E.5: Accuracy scores for SVM.NuSVC: Non-filtered Signal and mRMR Feature Selection.

	Accuracy Scores /% in rounds $R_{1...10}$									
BR_Y	85,12	81,55	85,12	79,17	89,29	82,74	91,07	88,09	81,55	87,5
BR_{AE}	72,02	76,19	81,55	77,98	77,38	79,17	78,57	75,59	77,98	79,17
BR_{Mic}	66,67	76,79	73,21	73,81	73,21	76,79	72,62	66,07	67,86	70,83

Table E.6: Confusion Matrix /% for Decision Tree and mRMR Feature Selection: $\boldsymbol{BR_y}$ - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	96,53	1,34	1,47	0,48
Inner	1,49	90,18	9,31	0,00
Ball	1,98	5,36	91,67	0,48
Healthy	0,50	0,00	0,98	98,57

Table E.7: Confusion Matrix /% for Random Forest and mRMR Feature Selection: $\boldsymbol{BR_y}$ - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	97,38	1,19	1,19	0,24
Inner	0,71	95,71	3,33	0,24
Ball	1,19	4,76	93,57	0,48
Healthy	1,67	0,48	0,71	97,14

Table E.8: Confusion Matrix /% for SVM.SVC and mRMR Feature Selection: $\boldsymbol{BR_y}$ - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	88,10	6,43	5,48	0,00
Inner	0,00	78,10	21,9	0,00
Ball	0,00	10,48	89,52	0,00
Healthy	0,48	5,71	7,38	86,43

Table E.9: Confusion Matrix /% for SVM.LinearSVC and mRMR Feature Selection: $\boldsymbol{BR_y}$ - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	95,95	2,62	1,19	0,24
Inner	1,67	77,38	17,62	3,33
Ball	0,95	9,29	89,76	0,00
Healthy	2,14	4,29	4,52	89,05

Table E.10: Confusion Matrix /% for SVM.NuSVC and mRMR Feature Selection: $\boldsymbol{BR_y}$ - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	83,10	12,86	4,05	0,00
Inner	0,00	81,19	18,81	0,00
Ball	0,24	9,52	90,24	0,00
Healthy	0,48	7,38	6,19	85,95

Table E.11: Confusion Matrix /% for Decision Tree and mRMR Feature Selection: AE - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	97,52	0,45	0,98	0,95
Inner	0,99	92,86	1,96	4,76
Ball	0,50	3,12	92,16	3,81
Healthy	0,0	4,91	3,43	91,43

Table E.12: Confusion Matrix /% for Random Forest and mRMR Feature Selection: AE - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	99,76	0,00	0,24	0,00
Inner	0,24	97,62	0,71	1,43
Ball	0,24	2,86	93,57	3,33
Healthy	0,24	4,76	1,19	93,81

Table E.13: Confusion Matrix /% for SVM.SVC and mRMR Feature Selection: AE - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	89,05	10,95	0,00	0,00
Inner	11,43	35,24	0,00	53,33
Ball	5,95	1,19	49,29	43,57
Healthy	4,52	6,19	0,00	89,29

Table E.14: Confusion Matrix /% for SVM.LinearSVC and mRMR Feature Selection: AE - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	93,33	3,57	3,10	0,00
Inner	6,19	75,95	0,71	17,14
Ball	5,24	3,33	49,29	42,14
Healthy	1,90	14,05	0,00	84,05

Table E.15: Confusion Matrix /% for SVM.NuSVC and mRMR Feature Selection: AE - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	98,10	1,67	0,24	0,00
Inner	4,52	87,38	1,19	6,90
Ball	5,24	7,86	49,29	37,62
Healthy	3,33	19,29	1,90	75,48

Table E.16: Confusion Matrix /% for Decision Tree and mRMR Feature Selection: Mic - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	90,59	7,59	0,00	0,95
Inner	8,91	86,16	6,37	0,00
Ball	1,98	5,80	88,24	3,33
Healthy	3,47	0,00	5,39	91,43

Table E.17: Confusion Matrix /% for Random Forest and mRMR Feature Selection: Mic - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	92,38	5,48	0,95	1,19
Inner	7,14	87,86	5,00	0,00
Ball	1,67	6,67	89,52	2,14
Healthy	2,62	0,24	3,81	93,33

Table E.18: Confusion Matrix /% for SVM.SVC and mRMR Feature Selection: Mic - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	71,67	26,19	2,14	0,00
Inner	25,24	70,95	3,81	0,00
Ball	0,95	21,90	36,43	40,71
Healthy	7,86	5,00	1,9	85,24

Table E.19: Confusion Matrix /% for SVM.LinearSVC and mRMR Feature Selection: Mic - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	68,81	26,43	4,05	0,71
Inner	25,95	60,71	13,33	0,00
Ball	2,38	7,86	55,24	34,52
Healthy	3,57	4,52	5,00	86,9

Table E.20: Confusion Matrix /% for SVM.NuSVC and mRMR Feature Selection: Mic - Non-filtered Signal.

	Outer	Inner	Ball	Healthy
Outer	77,62	18,81	3,57	0,00
Inner	11,43	79,52	9,05	0,00
Ball	0,24	16,9	46,67	36,19
Healthy	8,33	4,29	4,05	83,33

E.2 Filtered Signal

Table E.21: Accuracy scores for Decision Tree: Filtered Signal and mRMR Feature Selection.

	Accuracy Scores /% in rounds $R_{1...10}$									
BR_Y	92,94	97,65	96,47	95,29	92,94	97,62	96,43	95,18	97,59	93,83
BR_{AE}	96,47	94,11	92,94	90,59	92,94	92,86	88,09	90,36	92,77	92,59
BR_{Mic}	89,41	94,12	90,59	94,12	91,76	91,67	94,05	95,18	87,95	93,83

Table E.22: Accuracy scores for Random Forest: Filtered Signal and mRMR Feature Selection.

	Accuracy Scores /% in rounds $R_{1...10}$									
BR_Y	97,62	99,41	98,21	98,21	97,62	98,81	98,81	97,02	95,83	97,62
BR_{AE}	97,62	97,02	97,02	98,21	96,43	96,43	94,64	97,02	97,62	92,86
BR_{Mic}	97,62	97,62	98,21	94,64	96,43	99,41	95,24	98,24	97,62	96,43

Table E.23: Accuracy scores for SVM.SVC: Filtered Signal and mRMR Feature Selection.

	Accuracy Scores /% in rounds $R_{1...10}$									
BR_Y	88,62	91,071	81,55	89,29	85,71	85,71	86,91	85,71	85,71	88,69
BR_{AE}	79,76	75,00	82,14	82,74	77,98	76,19	78,57	77,38	73,21	74,41
BR_{Mic}	77,98	75,00	76,79	73,81	83,33	76,79	80,36	75,00	72,02	77,38

Table E.24: Accuracy scores for SVM.LinearSVC: Filtered Signal and mRMR Feature Selection.

	Accuracy Scores /% in rounds $R_{1...10}$									
BR_Y	90,48	92,26	80,95	88,09	84,52	82,74	86,91	85,71	86,31	86,91
BR_{AE}	75,60	73,21	79,76	77,98	77,38	72,02	77,98	77,38	72,62	70,24
BR_{Mic}	80,95	77,98	75,60	71,43	81,55	79,76	82,74	79,76	76,79	76,79

Table E.25: Accuracy scores for SVM.NuSVC: Filtered Signal and mRMR Feature Selection.

	Accuracy Scores /% in rounds $R_{1...10}$									
BR_Y	87,5	86,91	79,16	86,91	83,93	81,55	84,52	82,74	83,33	87,5
BR_{AE}	79,17	75,00	79,76	80,36	79,17	72,02	74,41	75,59	71,43	72,02
BR_{Mic}	79,17	77,38	74,41	75,60	81,55	77,98	83,33	76,79	77,98	79,17

Table E.26: Confusion Matrix /% for Decision Tree: Filtered Signal and mRMR Feature Selection, $\boldsymbol{BR_y}$.

	Outer	Inner	Ball	Healthy
Outer	96,14	1,91	1,95	0,00
Inner	1,45	93,78	4,39	0,46
Ball	1,45	5,74	92,20	0,46
Healthy	0,00	0,00	0,00	100,00

Table E.27: Confusion Matrix /% for Random Forest: Filtered Signal and mRMR Feature Selection, $\boldsymbol{BR_y}$.

	Outer	Inner	Ball	Healthy
Outer	98,81	0,71	0,48	0,00
Inner	0,48	98,81	1,19	0,24
Ball	0,24	2,62	96,67	0,48
Healthy	0,48	0,71	0,71	98,10

Table E.28: Confusion Matrix /% for SVM.SVC: Filtered Signal and mRMR Feature Selection, $\boldsymbol{BR_y}$.

	Outer	Inner	Ball	Healthy
Outer	98,33	0,48	1,19	0,00
Inner	0,95	83,10	15,48	0,48
Ball	0,48	5,95	89,08	4,52
Healthy	1,43	10,24	11,19	77,14

Table E.29: Confusion Matrix /% for SVM.LinearSVC: Filtered Signal and mRMR Feature Selection, $\boldsymbol{BR_y}$.

	Outer	Inner	Ball	Healthy
Outer	97,38	1,43	1,19	0,00
Inner	1,43	80,24	14,76	3,57
Ball	0,24	10,24	85,00	4,52
Healthy	1,19	8,57	6,90	83,33

Table E.30: Confusion Matrix /% for SVM.NuSVC: Filtered Signal and mRMR Feature Selection, $\boldsymbol{BR_y}$.

	Outer	Inner	Ball	Healthy
Outer	91,67	1,43	6,90	0,00
Inner	0,00	82,14	17,38	0,48
Ball	0,24	6,19	89,05	4,52
Healthy	0,24	10,95	14,05	74,74

Table E.31: Confusion Matrix /% for Decision Tree: Filtered Signal and mRMR Feature Selection, AE.

	Outer	Inner	Ball	Healthy
Outer	97,58	0,00	0,98	1,37
Inner	0,97	90,91	1,95	5,94
Ball	0,97	1,91	92,20	4,57
Healthy	1,45	5,74	4,39	89,04

Table E.32: Confusion Matrix /% for Random Forest: Filtered Signal and mRMR Feature Selection, AE.

	Outer	Inner	Ball	Healthy
Outer	99,76	0,00	0,24	0,00
Inner	0,48	97,38	0,71	1,43
Ball	0,48	1,67	94,05	3,81
Healthy	0,24	3,57	1,43	94,76

Table E.33: Confusion Matrix /% for SVM.SVC: Filtered Signal and mRMR Feature Selection, AE.

	Outer	Inner	Ball	Healthy
Outer	97,86	2,14	0,00	0,00
Inner	5,00	89,05	0,00	5,95
Ball	5,24	9,05	49,29	36,43
Healthy	3,81	21,43	0,00	74,76

Table E.34: Confusion Matrix /% for SVM.LinearSVC: Filtered Signal and mRMR Feature Selection, AE.

	Outer	Inner	Ball	Healthy
Outer	93,33	3,57	3,10	0,00
Inner	6,43	75,0	0,71	17,86
Ball	5,24	3,10	49,53	42,14
Healthy	1,67	14,52	0,00	83,81

Table E.35: Confusion Matrix /% for SVM.NuSVC: Filtered Signal and mRMR Feature Selection, AE.

	Outer	Inner	Ball	Healthy
Outer	92,86	1,67	5,48	0,00
Inner	4,52	87,14	1,43	6,90
Ball	5,71	8,33	48,33	37,62
Healthy	3,33	19,29	2,14	75,24

Table E.36: Confusion Matrix /% for Decision Tree: Filtered Signal and mRMR Feature Selection, Mic.

	Outer	Inner	Ball	Healthy
Outer	87,44	10,05	0,49	1,83
Inner	3,38	91,87	4,88	0,00
Ball	1,93	4,31	92,20	1,37
Healthy	0,48	0,48	1,95	97,26

Table E.37: Confusion Matrix /% for Random Forest: Filtered Signal and mRMR Feature Selection, Mic.

	Outer	Inner	Ball	Healthy
Outer	96,19	3,10	0,48	0,24
Inner	2,62	96,9	0,24	0,24
Ball	0,48	1,43	97,86	0,24
Healthy	1,43	0,00	0,95	97,62

Table E.38: Confusion Matrix /% for SVM.SVC: Filtered Signal and mRMR Feature Selection, Mic.

	Outer	Inner	Ball	Healthy
Outer	68,57	14,52	9,52	7,38
Inner	4,76	70,95	24,29	0,00
Ball	4,76	22,38	70,00	2,86
Healthy	2,14	0,00	0,00	97,86

Table E.39: Confusion Matrix /% for SVM.LinearSVC: Filtered Signal and mRMR Feature Selection, Mic.

	Outer	Inner	Ball	Healthy
Outer	74,05	10,71	7,38	7,86
Inner	7,62	72,14	20,24	0,00
Ball	5,48	21,19	70,00	3,33
Healthy	2,86	0,00	0,00	97,14

Table E.40: Confusion Matrix /% for SVM.NuSVC: Filtered Signal and mRMR Feature Selection, Mic.

	Outer	Inner	Ball	Healthy
Outer	63,81	15,48	8,10	12,62
Inner	4,76	85,00	9,76	0,48
Ball	3,57	24,76	65,00	6,67
Healthy	0,24	0,00	0,24	99,52