



Politecnico di Torino

DEPARTMENT OF ENGINEERING
Master degree in Mechanical Engineering

**STUDY OF A MACHINE LEARNING BASED
METHODOLOGY APPLIED TO FAULT DETECTION AND
IDENTIFICATION IN AN ELECTROMECHANICAL
SYSTEM**

Candidate:
Davide Palmisano

Thesis advisors:
Prof. Miguel Delgado Prieto
Prof. Enrico Galvagno
Research supervisor:
Dott. Francisco Arellano Espitia

Contents

1	Abstract	6
2	Introduction	7
2.1	Topic	7
2.2	Motivation	8
2.3	Objectives	10
2.4	Structure of the document	11
3	State of the art	12
3.1	Fault frequencies approach	12
3.2	Machine Learning	13
4	Case of study	15
5	Theoretical fault frequencies method	20
5.1	Introduction	20
5.2	Vibration measurements	22
5.3	Results obtained	24
5.4	Conclusions	32
6	Machine Learning Method	33
6.1	Introduction	33
6.2	Creation of the data base	36
6.3	PCA technique	39
6.3.1	Introduction	39
6.3.2	Results obtained	41
6.4	LDA technique	47
6.4.1	Introduction	47
6.4.2	Results obtained	49
6.5	Design of the Neural Network	55
6.6	Results Obtained	59
6.6.1	PCA reduction	59
6.6.2	LDA reduction	63
6.7	Optimization of the Neural Network	67
6.7.1	PCA reduction	67
6.7.2	LDA reduction	76
7	Final conclusions	83

List of Figures

2.1	Scheme of an electric motor [20]	7
2.2	Distribution of the failures in an induction motor [17]	10
4.1	Induction motor	15
4.2	Metallic and ceramic bearing mounted on the shaft	16
4.3	Scheme of a radial bearing [2]	16
4.4	Example of vibration signal in time, for metallic bearing at 25 Hz	18
4.5	Example of current signal in time, for metallic bearing at 25 Hz	18
4.6	Accelerotemer <i>KS943B100</i>	19
4.7	Test bench used for the measurements	19
5.1	Steps followed in the development of this section.	20
5.2	Spectrum of vibration signal for metallic bearings at 25 Hz	22
5.3	Spectrum of vibration signal for metallic bearings at 50 Hz	22
5.4	Spectrum of vibration signal for ceramic bearings at 25 Hz	23
5.5	Spectrum of vibration signal for ceramic bearings at 50 Hz	23
5.6	Vibrations around the fault frequencies at different harmonics for metallic bearing at 25 Hz	25
5.7	Vibrations around the fault frequencies at different harmonics for metallic bearing at 50 Hz	26
5.8	Vibrations around the fault frequencies at different harmonics for ceramic bearing at 25 Hz	28
5.9	Vibrations around the fault frequencies at different harmonics for ceramic bearing at 25 Hz	30
6.1	Scheme of a Neural Network [9]	33
6.2	Scheme of a perceptron [9]	34
6.3	Comparison between sigmoid and perceptron output	34
6.4	Steps followed in the development of this section.	35
6.5	Visualization of how PCA works. [4]	39
6.6	Scatter plot PCA, for vibration measurements of metallic bearings	41
6.7	Scatter plot PCA, for stator current measurements of metallic bearings	42
6.8	Scatter plot PCA, for vibration measurements of ceramic bearings	43
6.9	Scatter plot PCA, for stator current measurements of ceramic bearings	44
6.10	Scatter plot PCA, for all the scenarios considered	45
6.11	Schematization of LDA	47
6.12	Scatter plot LDA, for vibration measurements of metallic bearings	49

6.13	Scatter plot LDA, for stator current measurements of metallic bearings	50
6.14	Scatter plot LDA, for vibration measurements of ceramic bearings	51
6.15	Scatter plot LDA, for stator current measurements of ceramic bearings	52
6.16	Evolution of Kurtosis, Shape factor, Crest factor and Impulse factor with respect to the size of the fault [13]	53
6.17	Scatter plot LDA, for all the scenarios considered	54
6.18	Example of a cost function	57
6.19	Scheme of the Neural Network	57
6.20	Performance of the network	58
6.21	Confusion Matrix	58
6.22	Confusion Matrix PCA, for vibration measurements of metallic bearings	59
6.23	Confusion Matrix PCA, for stator current measurements of metallic bearings	60
6.24	Confusion Matrix PCA, for vibration measurements of ceramic bearings	61
6.25	Confusion Matrix PCA, for stator current measurements of ceramic bearings	62
6.26	Confusion Matrix LDA, for vibration measurements of metallic bearings	63
6.27	Confusion Matrix LDA, for stator current measurements of metallic bearings	64
6.28	Confusion Matrix LDA, for vibration measurements of ceramic bearings	65
6.29	Confusion Matrix LDA, for stator current measurements of ceramic bearings	66
6.30	Effect of the number of epochs on the classification of the Neural Network with PCA reduction	68
6.31	Effect of the hidden layer size on the classification of the Neural Network with PCA reduction	69
6.32	Optimized solution for PCA of metallic bearings with vibration measurement	70
6.33	Optimized solution for PCA of metallic bearings with current measurement	70
6.34	Optimized solution for PCA of ceramic bearings with vibration measurement	71
6.35	Optimized solution for PCA of ceramic bearings with current measurement.	71
6.36	Effect of the number of epochs on the classification of the Neural Network with PCA reduction, considering alle the scenarios	73
6.37	Effect of the hidden layer size on the classification of the Neural Network with PCA reduction, considering all the scenarios	74
6.38	Optimized solution for PCA considering all the scenarios	75
6.39	Effect of the number of epochs on the classification of the Neural Network with LDA reduction	76
6.40	Effect of the hidden layer size on the classification of the Neural Network with LDA reduction	77

6.41	Optimized solution for LDA of metallic bearings with vibration measurement	78
6.42	Optimized solution for LDA of metallic bearings with current measurement	78
6.43	Optimized solution for LDA of ceramic bearings with vibration measurement	79
6.44	Optimized solution for LDA of ceramic bearings with current measurement	79
6.45	Effect of the number of epochs on the classification of the Neural Network, considering all the scenarios	80
6.46	Effect of the hidden layer size on the classification of the Neural Network, considering all the scenarios	81
6.47	Optimized solution for LDA of ceramic bearings with vibration measurement	82

List of Tables

4.1	Mechanical features of the bearing	17
5.1	Fault frequencies	21
5.2	Vibration amplitude at fault frequencies for metallic bearings at 25 Hz	24
5.3	Vibration amplitude at fault frequencies for metallic bearings at 50 Hz	27
5.4	Vibration amplitude at fault frequencies for ceramic bearings at 25 Hz	29
5.5	Vibration amplitude at fault frequencies for ceramic bearings at 50 Hz	31
6.1	Most important time domain features for vibration measurements of metallic bearings, with PCA reduction	41
6.2	Most important time domain features for current measurements of metallic bearings, with PCA reduction,	42
6.3	Most important time domain features for vibration measurements of ceramic bearings, with PCA reduction	43
6.4	Most important time domain features for current measurements of ceramic bearings, with PCA reduction	44
6.5	Most important time domain features for all the scenarios, with PCA reduction	45
6.6	Most important time domain features for vibration measurements of metallic bearings, with LDA reduction	49
6.7	Most important time domain features for current measurements of metallic bearings, with LDA reduction	50
6.8	Most important time domain features for vibration measurements of ceramic bearings, with LDA reduction	51
6.9	Most important time domain features for current measurements of ceramic bearings, with LDA reduction	52
6.10	Most important time domain features for all the scenarios, with LDA reduction	53

Chapter 1

Abstract

This work addresses the application of two different methods in order to detect faults in bearings operating within an electromechanical system, based on the measurement of vibrations and stator currents. The electromechanical system considered is a shaft connected to an electric induction motor. Two bearings are mounted on the shaft; these bearings can be metallic or ceramic. The bearings can be found in three different conditions: healthy or with a inner race hole of 1 *mm* or 2 *mm*. First of all the analysis of theoretical fault frequencies was explored. The goal of this method is to identify theoretical fault frequencies, depending on features of the bearing, in order to verify the presence of peaks in the frequency signals obtained from laboratory measurements. The accuracy of the theoretical frequency calculation was demonstrated by the actual presence of these peaks in the frequency signals, however it was expected to be found a proportion between the peak heights, and the severity of the fault, but this didn't happened. That led to the development of the second method, based on the building of a neural network able to classify the bearings with respect to their conditions, starting from 15 different statistical time domain features as input. Two reduction techniques, LDA and PCA, were implemented in order to reduce the number of input to the two most significant features; after that the neural network was built. The results obtained with this second method are very satisfactory as they allow to classify with a good performance both considering the different scenarios of bearing material and measured signal taken individually, but also considering all four different scenarios at the same time.

Chapter 2

Introduction

2.1 Topic

Electrical motors are crucial devices in modern life, with lots of applications in various fields, such as industrial machinery, motor vehicles, aerospace, transportations and other commercial applications. Of all these fields, industrial applications are the most important ones, where they are widely used in pumps, conveyor systems, compressor, fans and other industrial machineries. The global electric market size was \$96.9 billions and is projected to reach \$136.4 billions in 2025 growing at an annual rate of 4.5% [1]. Electric motors can be founded in different configurations such as DC motors, brushed or brushless, or AC motors, synchronous or at induction. As shown in figure 2.1 all the electric motors are

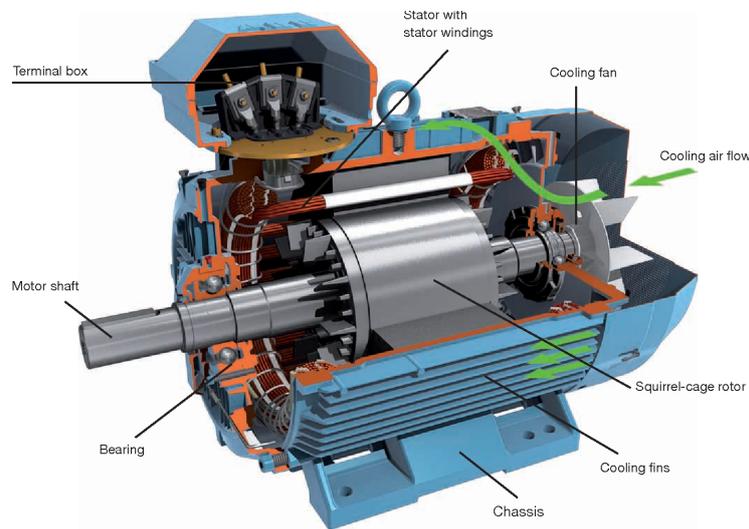


Figure 2.1: Scheme of an electric motor [20]

characterized by the same main components that are:

- stator;

- rotor;
- windings;
- shaft;
- bearings;
- power electronics.

2.2 Motivation

Linked to the different components that characterize an electric motor there are many types of faults that can happen, such as : [5]

1. **Stator faults** including stator open phase, stator unbalance do to short circuits or increased resistance.
2. **Rotor electrical faults** such as rotor open phase, rotor unbalance due to short circuits or increased resistance connections for wound rotor machines and broken bar or cracked end-ring for squirrel-cage machines, and rotor magnetic faults as demagnetization .
3. **Rotor mechanical faults** that include bearing damage, eccentricity, bent shaft and misalignment.
4. **Power electronics failures** .

This means that a lot of failures can happen, thus causing breakdown of machinery, unproductive times, high maintenance costs and in some particular sectors, such as hospitals, loss of life. The ability to detect these damages when they are forming, so that targeted maintenance can be carried out in a timely manner, can really make a difference within an industry. Fault detection has been studied in great depth in the academic field, the most used techniques are listed below [17] :

1. **Signal-based fault diagnosis** which use data collected by sensors mounted on the system under examination, which provide information on particular physical quantities, the evolution of which over time can provide information on the state of health of the machine. Different approaches can be found:
 - mechanical vibration analysis;
 - shock pulse monitoring;
 - stator current analysis;

- acoustic noise analysis;
 - temperature measurement.
2. **Model-based fault diagnosis** this approach involves the creation of a mathematical model, defined by a series of parameters, the number of which determine how sophisticated this model is and therefore able to predict the real behavior of the system. Some typical techniques used are:
- neural network;
 - fuzzy logic analysis;
 - genetic algorithm;
 - artificial intelligence.
3. **Machine-theory-based fault analysis** i.e. a series of models based on the physical principles underlying the operation of the machine and which make it possible to predict its change as certain physical properties change as a result of the present damage. For example there are:
- winding function approach;
 - modified winding function approach;
 - magnetic equivalent circuit.
4. **Simulations-based fault analysis** which use complex mathematical models to generate data sets in case of damage and automatically evaluate them. All this is then validated by comparing the data obtained from the simulations, with measurements made in real cases. Some methods used are:
- finite-element analysis;
 - time-step coupled finite element state space analysis.

All these techniques can be used to detect all the different faults seen before, but not all the faults happen with the same probability. As shown in figure 2.2, **bearings** are the most critical components, with 69% of the total failures. This is why in this work the two most important techniques used to detect bearing faults were explored, that are:

1. **Fault frequencies analysis**
 this method is based on the identification of certain particular frequencies, which can be identified in the spectrum of a measured physical quantity (e.g. bearing vibration). This frequency is a function of the geometric

characteristics of the bearing and in its correspondence it is possible to detect peaks of the measured quantity, due to the presence of the damage.

2. **Machine learning method** this method, as it will be widely described later in this work, exploits a neural network that, properly trained, is able to recognize the presence of a damage by exploiting a series of quantities given as input to the network.

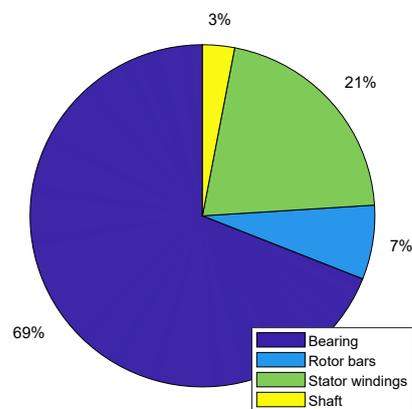


Figure 2.2: Distribution of the failures in an induction motor [17]

2.3 Objectives

The main objective of this work is the detection and classification of ball bearing faults within an electromechanical system, using vibration and stator current measurements, taken in the laboratory. These measurements will be used to carry out a fault frequency analysis and to build a classifier neural network, with the main goal to not only detect the presence of a fault in a rolling bearing, but also understand the severity of these faults. The real challenge of this work is therefore the realization of a system of recognition and classification of damage that is **global**, i.e. applicable in different operating situations (in this case the two power supply frequencies 25 Hz and 50 Hz will be analyzed) and considering different technologies for the realization of bearings, in particular both metal and **ceramic** bearings will be studied here. The study of ceramic bearings, combined with metal bearings, represents one of the main novelties of this study compared to the current state of the art. The objective for the future is then to expand

this model, making it adaptable to a number of operating cases that is as varied and universal as possible, and therefore applicable to bearings of different sizes, materials, operating in various conditions and capable of recognizing damage of different shape, position and size.

2.4 Structure of the document

After a brief introduction about the topic studied, the problems connected and the proposal of the work, in chapter 3 is described a state of the art study about fault frequencies and machine learning methods of bearing fault detection. Then in chapter 4 the case of study is described, with also all the experimental apparatus, the signal measurement process and the whole electromechanical system. In chapter 5 the fault frequencis analysis is described, then are reported the vibration measurements and the results obtained. Finally in chapter 6 the machine learning method is reported; this section contains the linear reduction techniques, the results obtained and te optimization of the neural network.

Chapter 3

State of the art

As said previously bearing fault detection is a crucial topic in modern industry this is why a lot of academic research were carried out in order to develop increasingly sophisticated techniques to address this issue. This detection can be done by taking into account different magnitudes, of witch the two most important are:

1. vibrations ;
2. stator current .

These two quantities are significant regarding the presence of a fault because when the bearing rolls over a defect, then an impact is produced, generating a vibration that is acquired with an accelerometer. The presence of this defect can also produce torque ripples and eccentricity between the rotor and the stator causing asymmetries in the system, variations in the air gap and therefore in the magnetic flux through the motor and consequently in the current passing through the stator. Vibrations and stator current are then analyzed into two different domains :

1. frequency domain;
2. time domain.

The idea behind the first approach is to identify some peaks in the signals acquired at certain characteristic fault frequencies provided from the literature. The presence of these peaks is related to the presence of a fault. On the other hand the second approach uses the calculation of statistical quantities calculated from vibration and current signals to define clusters and classify them according to different damage conditions.

3.1 Fault frequencies approach

In [2] the first approach is used, by analyzing in the frequency-domain both vibration and stator current signals, in order to compare the results obtained with the two different measurements. The electromechanical system studied is composed by a three-phase induction motor, supported by metalling bearings in

three different conditions: healthy, simulated brinneling defect, and outer race hole. The conclusions obtained are that vibration signal is a robust indicator since peaks in the frequency spectre were effectively found around the characteristic fault frequencies predicted. The problem with this solution is that the equipment needed for measuring vibrations is more complicated and expensive and moreover the accelerometers has to be mounted as nearest as possible to the bearing in order to acquire the correct vibration. Measuring stator current is easier and cheaper but the main issue is that the effects of the faults are visible only in certain operative conditions, since they can be totally buried in the present noise, this is why sophisticated filtering techniques may be needed. In [18] is also compared the detection of bearing faults using vibration and stator current analysis. The metallic bearings support a two pole-apirs induction machine and can be founded into two different configurations that are healthy or with a 3 mm hole into the outhter race. In this work too is stated that the signal of vibration is a good indicator of faults since peaks where find in the frequency spectrum around the frequencies predicted using the classic theory, but accelerometers are expansive and more complex to use. On the other hand the stator current signal is easier to get since often is already measured for control and detection purpose. A different approach is used and consists in using a fault detector defined by extracting energies on frequency ranges related to the characteristic frequency of the fault. The difference between the healthy and the fault scenario is performed for each one of these frequencies and then a cumulative sum is calculated. This detector shows good results for the vibration signal, but the effect on the current signal is lower. For amplifying that peaks on the current signal the supply frequency of the motor is tuned in order that the characteristic frequency of the fault is equal to the frequence of resonance of the motor, that can be obtained from the datasheet supplied by the manufacturer. In this way the energy spectre detector can recognize faulty coditions also with the stator current signal.

3.2 Machine Learning

Machine learning is the method behind how machines learn from data, by a continuous training. It's a subset of *Artificial Intelligence* (A.I.), that represents the science of training machines to perform human tasks. This technique was invented during 1950's but now is widespreading because of the incredible amount of data that can be collected with modern sensors and also because of more powerful computers to process that data. The fields of application are various, such as:

- **financial services** in order to identify investments opportunities and to prevent frauds;
- **health care** since sensors can use data to asses patients' health in real time and can identify trends or red flags that can diagnose a disease in advance;
- **oil and gas extraction** in this field machine learning can help find new energy sources, analyzing minerals in the soil and predicting failures of

the equipment used;

- **government and political administration** for detecting frauds, public safety and better administration of money ;
- **retail** for recommending items based on customers' previous researches, and also to improve marketing compaignes and price-policy optimization;
- **transporation** in order to make routes more efficient and to predict potential problems linked with more profitable routes.

Several methods can be used in order to train the neural network, of which the two most important ones are:

1. **Supervised learning** in which the training is performed using labeled examples where the desired output are known. It is used in applications where historical data predict likely future events.
2. **Unsupervised learning** that uses with no historical labels with the goal to explore that data and find patterns within. This method works better on transactional data. Popular techniques used are :
 - nearest-neighbour mapping;
 - k-means clustering;
 - singular value decomposition.

There are also **semi-supervised** methods in which both labeled and unlabeled data are used, usually a small amount of labeled ones and a huge amount of unlabeled ones. [19]

In this work is explored machine learning with the goal to detect faults in ball bearings so the state of the art in this field is reported below. In the work [15] vibration and stator current measurements are used to detect bearing faults, using the machine learning approach. The two signals are anlyzed in the time domain by extracting 10 statistical features used for classifying four different conditions, that are: healthy, inner and outer race scraped, partial demagnetization of the machine and static eccentricity. First of all the statistical features are obtained, then using a linear reduction technique the number of input variables is reduced to 2 and finally, an algorithm of machine learning is able to identify and classify a novel condition, with an accuracy near to 99%, every times it shows up (incremental learning).

In the work [14] are also used vibrations and stator current measurements but here the features are extracted from both time-domain and frequency-domain. Two experimental test bench are considered: a pulley-belt system and a gearbox-based system. Different operating conditions (in terms of torque load and speed of rotation) are considered and 5 fault conditions are taken into account: healthy, shaft misaligned, shaft unbalanced, broken rotor bar and defected bearing. Once obtained the features, then the number of input variables is reduced to 2 using the LDA technique and then a Neural Network classifier is used. The results in the classification are really good, with an accuracy up to 99% obtained thanks to the hybrid extraction of features from both time and frequency domain .

Chapter 4

Case of study

This work deals with fault detection of bearings working within an electro-mechanical system. This system is composed by an induction motor **ABB M2AA112M-2** (figure 4.1), with a rotating shaft connected, supported by bearings (figure 4.2). The bearing can be of two types:

Metallic bearing	model 6205-ETN9/C3
Ceramic bearing	model 6205-CE-ZRO2-ZEN

Only one type of bearing, for both metallic and ceramic, has been studied because considering bearings with different internal and external diameters, or with different number of balls, would have excessively increased the variance of the problem and would have considerably complicated the classification.



Figure 4.1: Induction motor

The faults in the bearings were simulated by doing in laboratory a hole in the inner race of the bearing, so that can be found in three different conditions:

- healthy;
- 1 mm inner race fault;
- 2 mm inner race fault.



Figure 4.2: Metallic and ceramic bearing mounted on the shaft

It is important to specify how the dimensions of the holes are much larger than those that usually occur in reality and which particularly occur in case of incipient damage. However, the choice to analyze damages of this size is to identify more easily the presence of differences between different damage conditions, as well as patterns within the same type of damage. However, it is proposed for the future to extend the study to different types of damage in terms of size, location and shape.

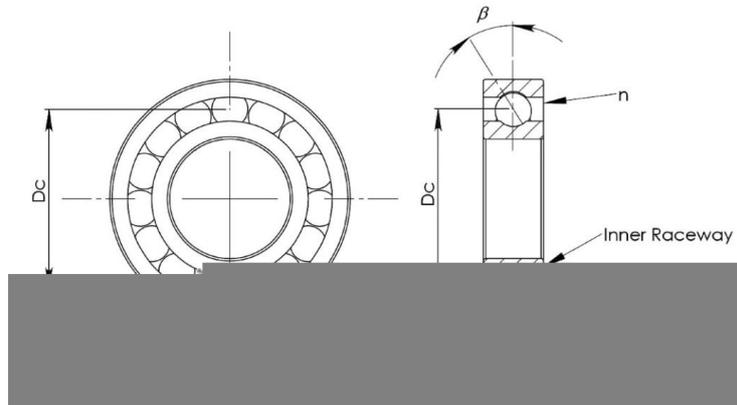


Figure 4.3: Scheme of a radial bearing [2]

The most important mechanical features of the bearings are described in table 4.1.

The induction motor is subjected to two supply frequencies that are:

$$\begin{cases} f_{s,1} = 25 \text{ Hz} \\ f_{s,2} = 50 \text{ Hz} \end{cases}$$

Number of balls	$n = 9$
Ball diameter	$D_b = 7,95 \text{ mm}$
Pitch diameter	$D_c = 39,05 \text{ mm}$
Angle of contact	$\beta = 0^\circ$

Table 4.1: Mechanical features of the bearing

since the tests were carried out without any load on the shaft the slip is negligible, so the frequency of rotation of the shaft can be considered equal to the supply frequency ($f_r \simeq f_s$). However, it is necessary to point out that the measurements were not taken under completely stationary conditions because, due to a series of tolerances due to the speed sensors, the inverter control and the test bench itself, the power requirement fluctuates by a few Hz from the nominal value. The fault detection is realized starting from vibration and stator current signals. Vibration signals are taken by means of an accelerometer **KS943B100** (fig. 4.6) and then acquired by the **PXIe 1062** acquisition system, provided by *National Instrument*. The sensitivity of the accelerometer is

$$S_{acc} = 100 \frac{mV}{\frac{m}{s^2}}$$

The accelerometer provides the components of the vibrations in the three directions of the space $[a_x, a_y, a_z]$ but only the x and z directions are taken into account since the y-direction is parallel to the angular speed one and that means that vibrations in that direction don't load radially the bearing. An example of the vibration time-signal for a metallic bearing is reported in figure 4.4. The current is measured with probe **A622** by *Farnell*, whose sensitivity is

$$S_{probe} = 100 \frac{mV}{A}$$

An example of the current time-signal for a metallic bearing is reported in figure 4.5.

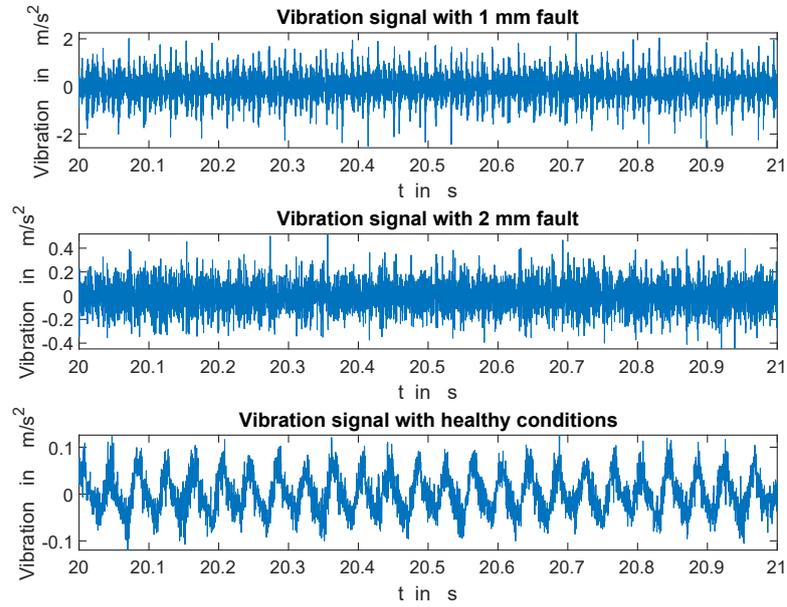


Figure 4.4: Example of vibration signal in time, for metallic bearing at 25 Hz

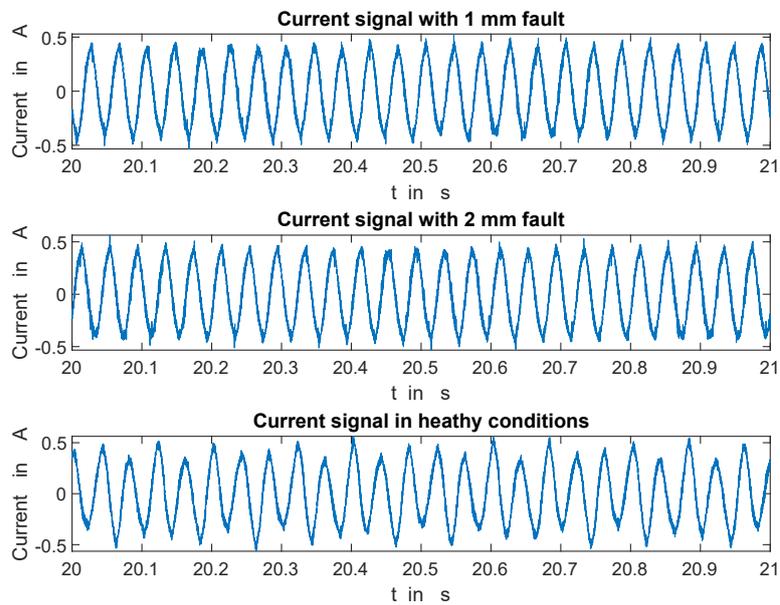


Figure 4.5: Example of current signal in time, for metallic bearing at 25 Hz

The sampling frequency of the signal is

$$f_{\text{sampl}} = 5000 \text{ Hz}$$

This value allow to respect the Nyquist theroem

$$f_{\text{sampl}} \geq 2 \cdot f_{\text{max}}$$

where $f_{\text{max}} = 2500 \text{ Hz}$. In this way it is possible to obtain a good resolution of the data, as the signal variations are intercepted well and there is no loss of information. In the first part of the work, the one concerning fault frequencies, only vibration measurements will be used. Subsequently, stator current measurements will be added to these to implement the machine learning algorithm, so as to have a larger sample of measurements with which to train the neural network.

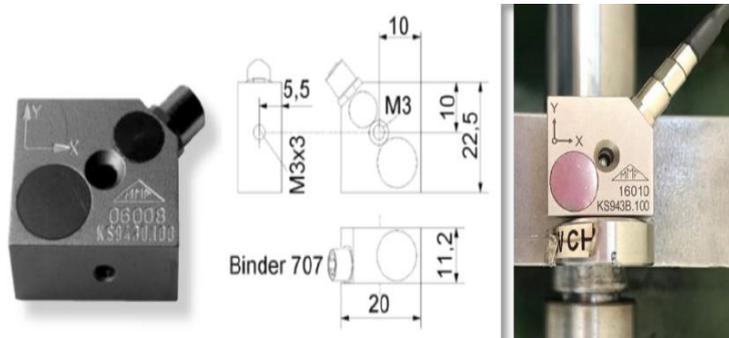


Figure 4.6: Accelerotemer *KS943B100*

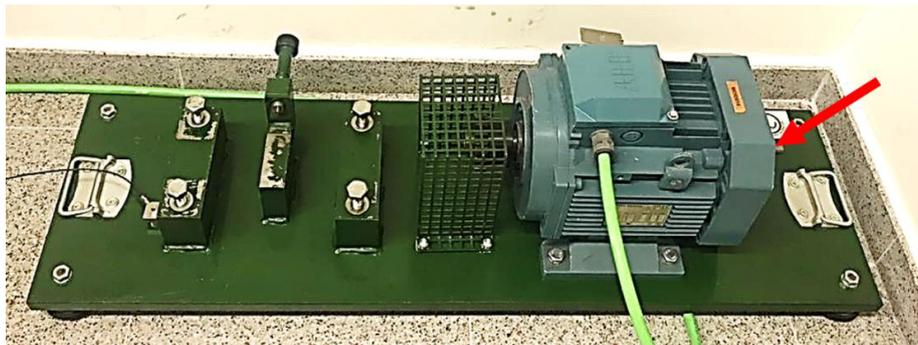


Figure 4.7: Test bench used for the measurements

Chapter 5

Theoretical fault frequencies method

5.1 Introduction

A bearing operating within an electromechanical system, over time, due to various factors such as fatigue, overloading, poor lubrication or a contaminated environment can become damaged. Whenever rolling elements pass through the damaged area, impacts occur that can be detected by analyzing the vibration spectrum of the bearings. These impacts occur at particular characteristic frequencies that depend on the frequency of shaft rotation and the mechanical characteristics of the bearing.

The work in this section will follow the flow chart in figure 5.1 .

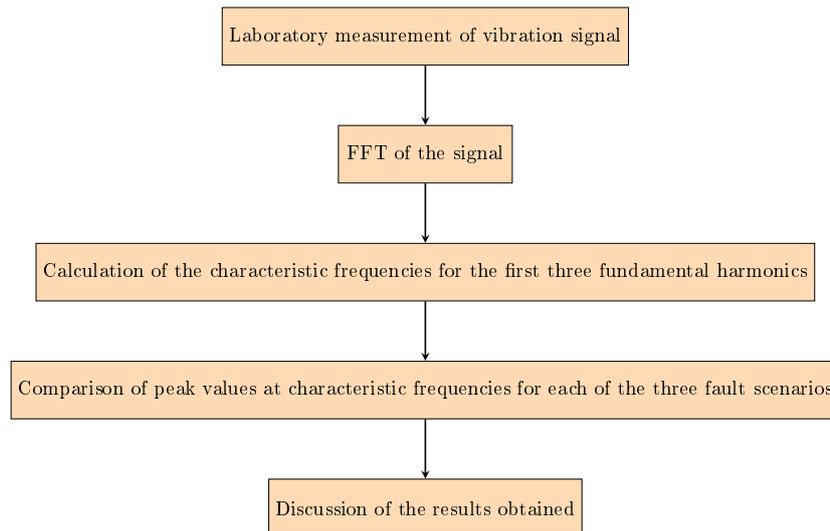


Figure 5.1: Steps followed in the development of this section.

For a single point defect, the literature [2] suggests the following expressions in order to find the fault frequencies:

$$\left\{ \begin{array}{l} f_C = \frac{f_r}{2} \left(1 - \frac{D_b \cdot \cos \beta}{D_c}\right) \\ f_O = \frac{f_r \cdot n}{2} \left(1 - \frac{D_b \cdot \cos \beta}{D_c}\right) \\ f_I = \frac{f_r \cdot n}{2} \left(1 + \frac{D_b \cdot \cos \beta}{D_c}\right) \\ f_B = \frac{f_r \cdot D_C}{D_b} \left(1 - \left(\frac{D_b \cdot \cos \beta}{D_c}\right)^2\right) \end{array} \right.$$

where the meaning of the frequencies is listed in table 5.1 and the other variables are the same listed in table 4.1 . Since the bearings considered are affected by

f_C	Cage fault frequency
f_O	Outer raceway fault frequency
f_I	Inner raceway fault frequency
f_B	Ball raceway fault frequency
f_R	Rotor mechanical frequency frequency

Table 5.1: Fault frequencies

inner race fault, the fault frequencies are:

$$\left\{ \begin{array}{ll} f_{I,1} = 135.4 \text{ Hz} & \text{for } f_R = 25 \text{ Hz} \\ f_{I,2} = 270.8 \text{ Hz} & \text{for } f_R = 50 \text{ Hz} \end{array} \right.$$

5.2 Vibration measurements

Metallic bearings

25 Hz

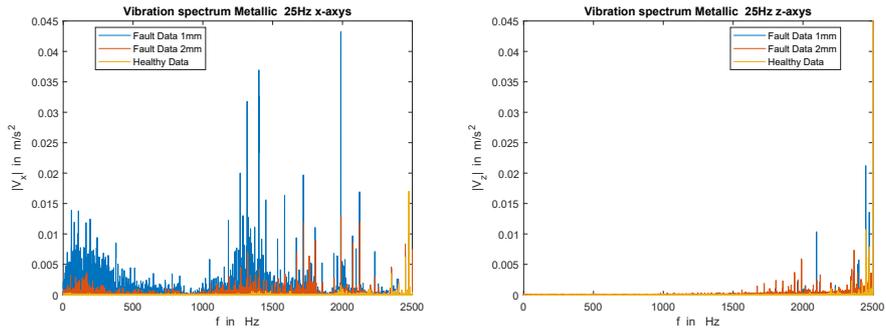


Figure 5.2: Spectrum of vibration signal for metallic bearings at 25 Hz

50 Hz

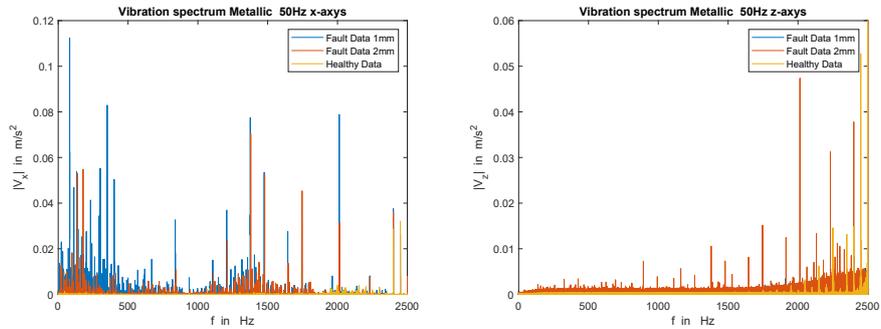


Figure 5.3: Spectrum of vibration signal for metallic bearings at 50 Hz

Ceramic bearings

25 Hz

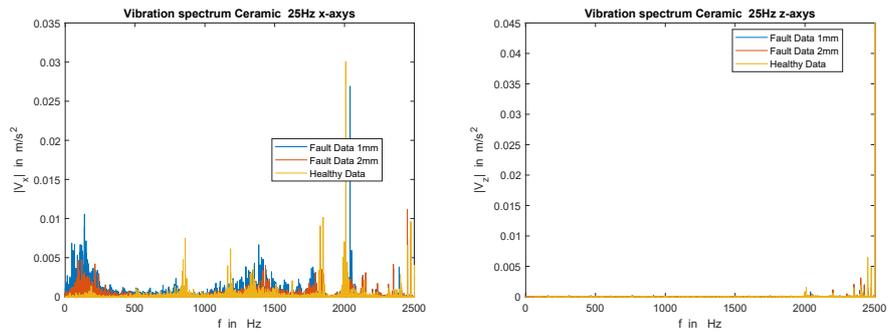


Figure 5.4: Spectrum of vibration signal for ceramic bearings at 25 Hz

50 Hz

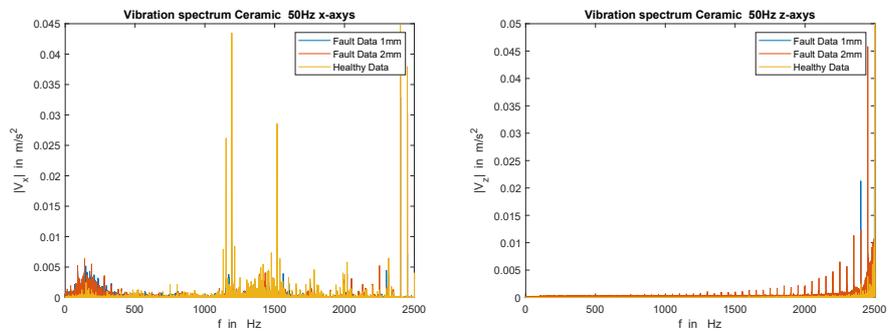


Figure 5.5: Spectrum of vibration signal for ceramic bearings at 50 Hz

5.3 Results obtained

To obtain more information, the entire measurement period was divided into small windows of 1 s each. Then the FFT was applied to each window in order to analyze the frequency domain, and in particular to see the behaviour of the signal measured near the fault frequency. The harmonics considered were the first three, obtained by the formula

$$f_I = k \cdot \frac{f_r \cdot n}{2} \left(1 + \frac{D_b \cdot \cos \beta}{D_c}\right)$$

with $k = 1, 2, 3$. A random window of the 100 available was taken into account; the results are reported in the following plots.

Metallic Bearings

25 Hz

Metallic 25 Hz x-axys	Amplitude ($\times 10^{-3}$) $\frac{m}{s^2}$		
	I Harmonic	II Harmonic	III Harmonic
Healthy	0,6615	0,1903	0,2306
Fault 1mm	0,6652	2,2438	0,8535
Fault 2mm	1,8474	7,0743	8,3316

Metallic 25 Hz z-axys	Amplitude ($\times 10^{-3}$) $\frac{m}{s^2}$		
	I Harmonic	II Harmonic	III Harmonic
Healthy	0,3296	0,3882	0
Fault 1mm	15,8632	4,5961	2,1104
Fault 2mm	4,3299	1,5746	10,9270

Table 5.2: Vibration amplitude at fault frequencies for metallic bearings at 25 Hz

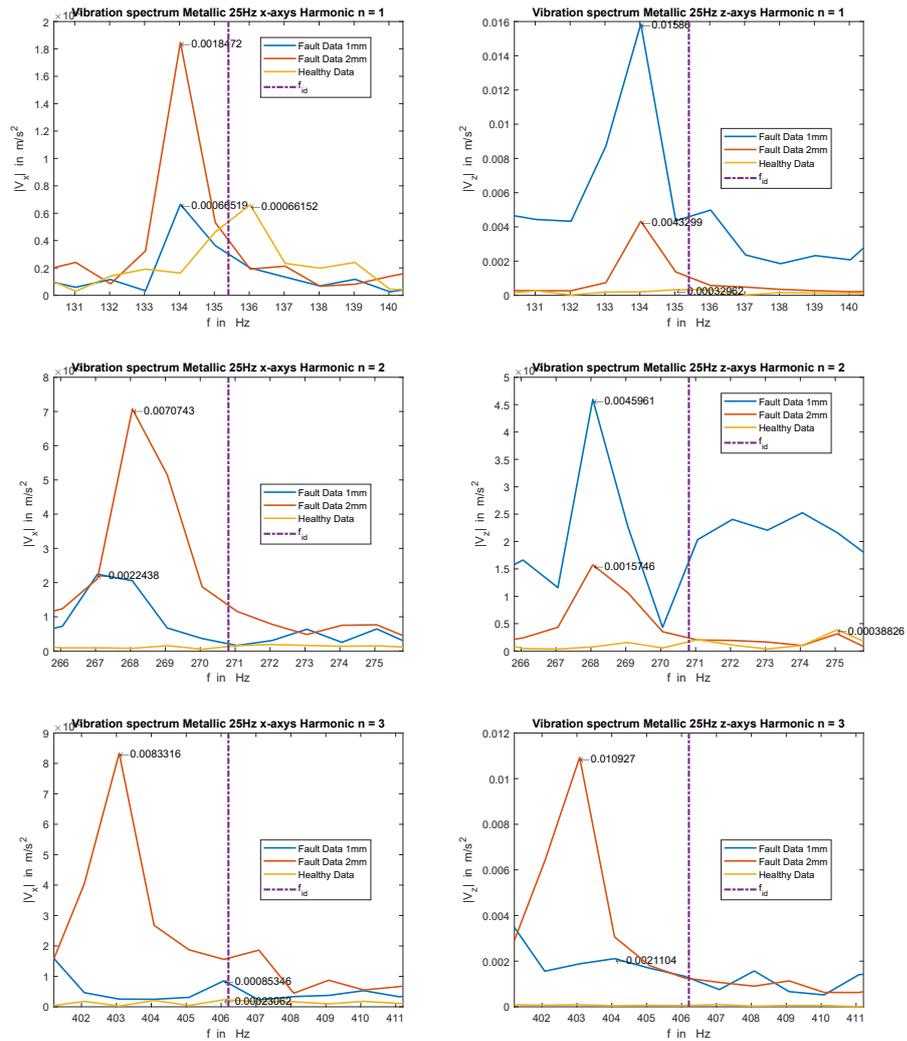


Figure 5.6: Vibrations around the fault frequencies at different harmonics for metallic bearing at 25 Hz

50 Hz

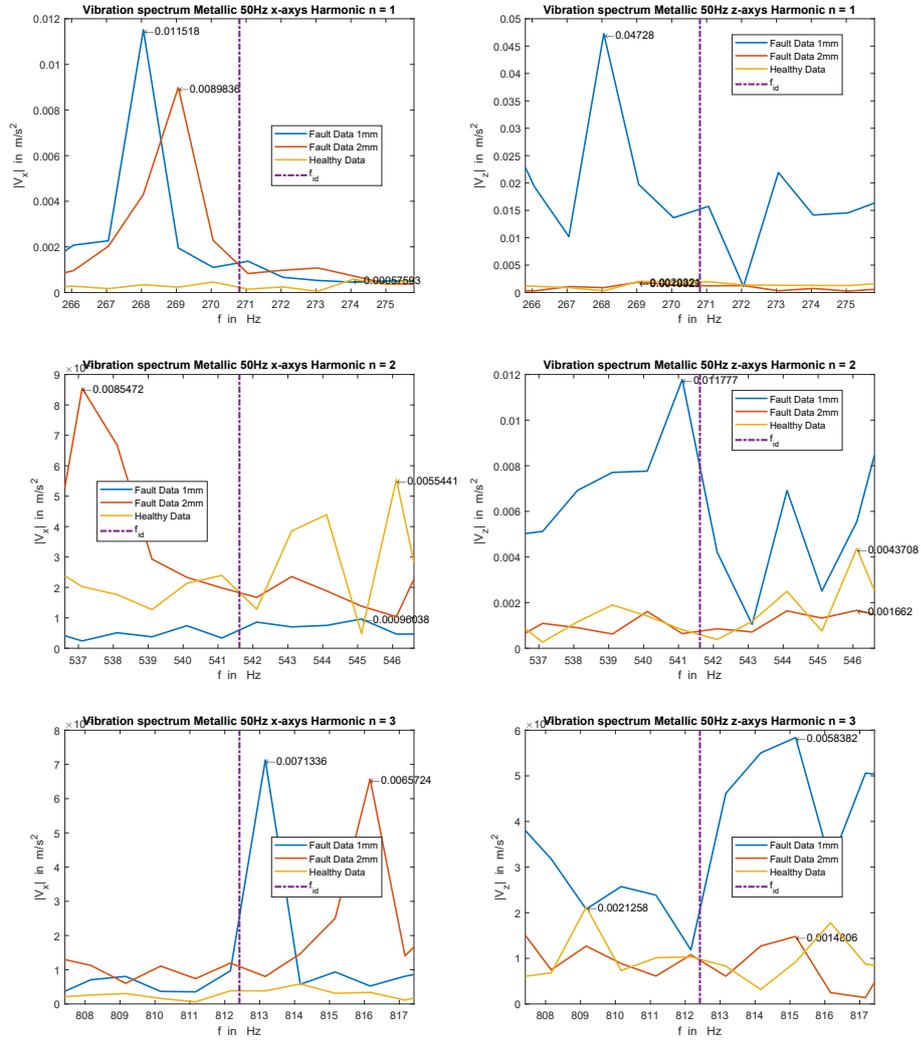


Figure 5.7: Vibrations around the fault frequencies at different harmonics for metallic bearing at 50 Hz

Metallic 50 Hz x-axys	Amplitude ($\times 10^{-3}$) $\frac{m}{s^2}$		
	I Harmonic	II Harmonic	III Harmonic
Healthy	0,5759	5,5441	0,5856
Fault 1mm	11,5180	0,9604	7,1336
Fault 2mm	8,9836	8,5472	6,5724

Metallic 50 Hz z-axys	Amplitude ($\times 10^{-3}$) $\frac{m}{s^2}$		
	I Harmonic	II Harmonic	III Harmonic
Healthy	2,0321	4,3708	2,1258
Fault 1mm	47,2832	11,7773	5,8382
Fault 2mm	1,8929	1,6624	1,4806

Table 5.3: Vibration amplitude at fault frequencies for metallic bearings at 50 Hz

Ceramics

25 Hz

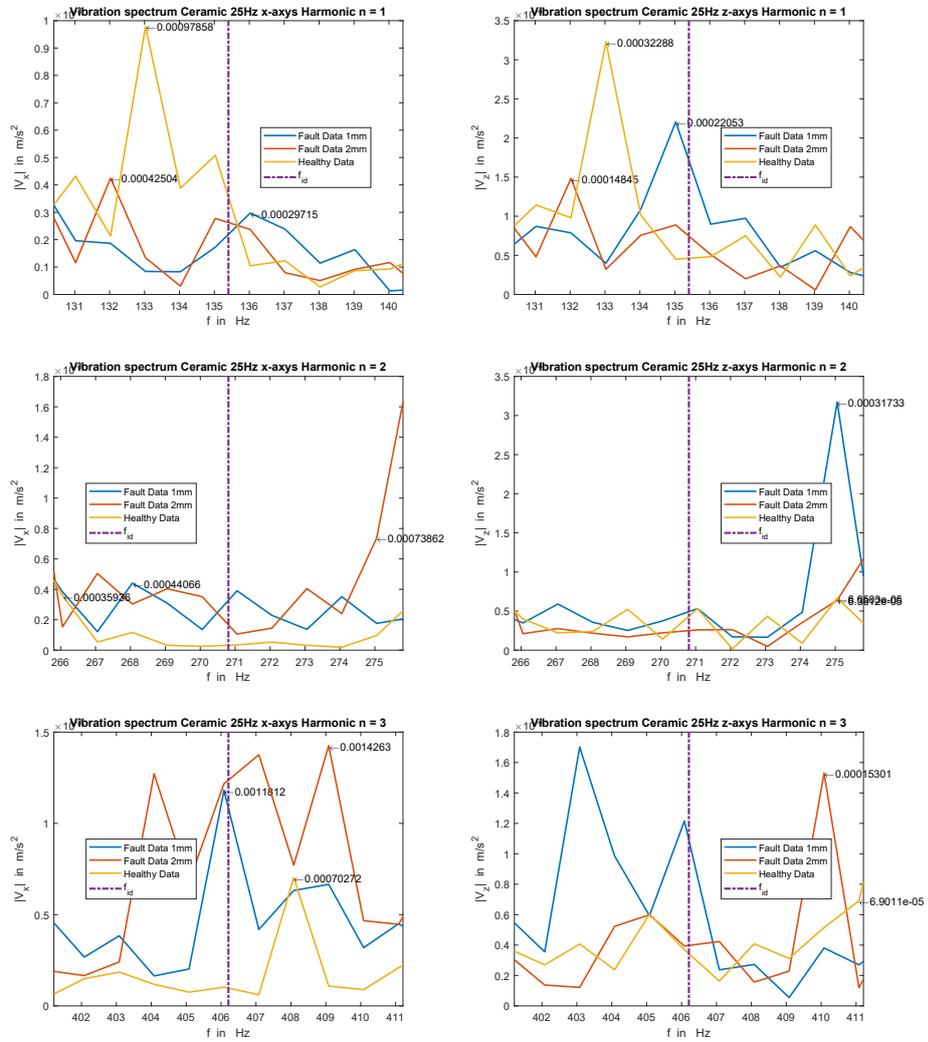


Figure 5.8: Vibrations around the fault frequencies at different harmonics for ceramic bearing at 25 Hz

Ceramic 25 Hz x-axys	Amplitude ($\times 10^{-3}$) $\frac{m}{s^2}$		
	I Harmonic	II Harmonic	III Harmonic
Healthy	0,9786	0,3594	0,7027
Fault 1mm	0,2972	0,4407	1,1812
Fault 2mm	0,4250	0,7386	1,4263

Ceramic 25 Hz z-axys	Amplitude $\frac{m}{s^2}$		
	I Harmonic	II Harmonic	III Harmonic
Healthy	0,3229	0,0065	0,0069
Fault 1mm	0,2205	0,3173	0,1746
Fault 2mm	0,1484	0,0064	0,1530

Table 5.4: Vibration amplitude at fault frequencies for ceramic bearings at 25 Hz

50 Hz

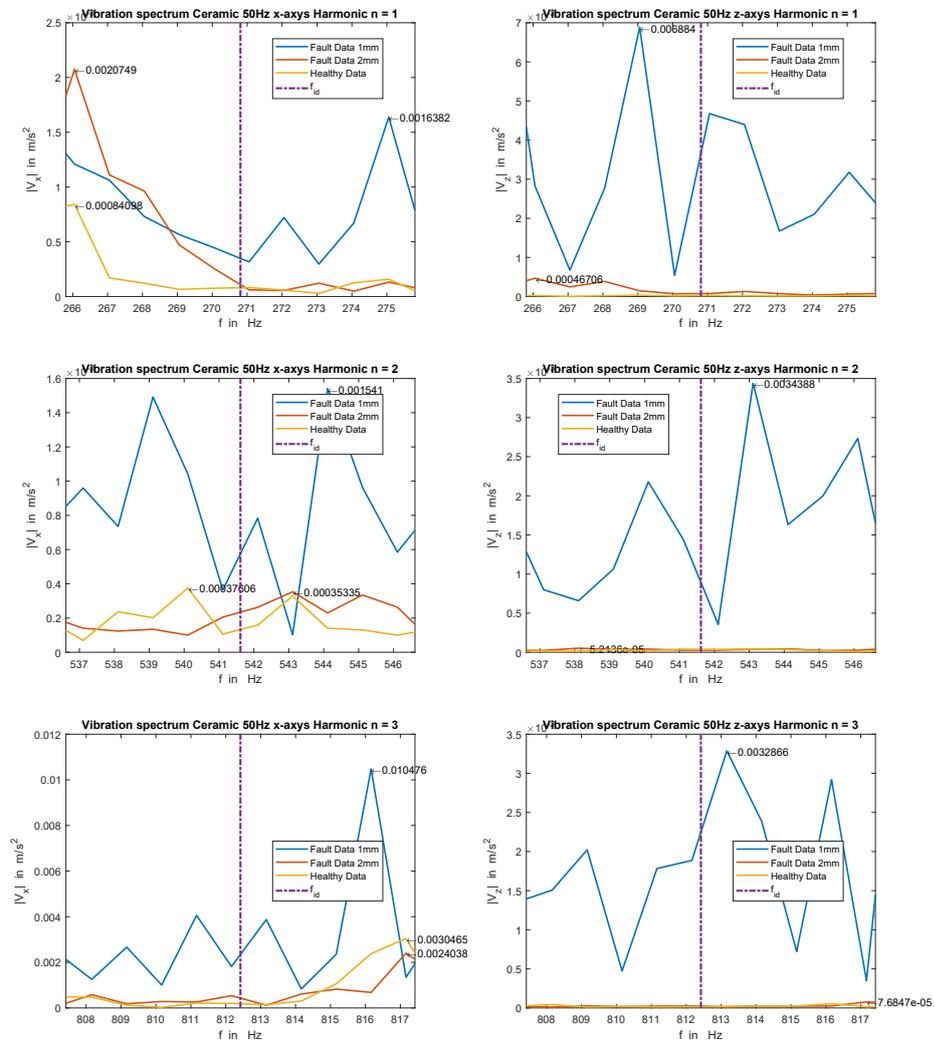


Figure 5.9: Vibrations around the fault frequencies at different harmonics for ceramic bearing at 25 Hz

Ceramic 50 Hz x-axys	Amplitude ($\times 10^{-3}$) $\frac{m}{s^2}$		
	I Harmonic	II Harmonic	III Harmonic
Healthy	0,8410	0,3761	3,0465
Fault 1mm	1,6382	1,5410	10,4760
Fault 2mm	2,0749	0,3534	2,4038

Ceramic 50 Hz z-axys	Amplitude $\frac{m}{s^2}$		
	I Harmonic	II Harmonic	III Harmonic
Healthy	0	0	0
Fault 1mm	6,8841	3,4388	3,2867
Fault 2mm	0,4706	0	0

Table 5.5: Vibration amplitude at fault frequencies for ceramic bearings at 50 Hz

5.4 Conclusions

The aim of this section was firstly to verify the validity of the formulas of the classical literature by detecting peaks in the vibration spectrum near the characteristic fault frequencies and secondly to look for a correlation between these peaks and the severity of the damage. The values reported in the tables prove the accuracy of theoretical formula, since are effectively shown peaks around the fault frequencies predicted, however there is no proportionality between the size of the failure and the magnitude of the peaks, as it could have been expected. In fact, overall in the three different harmonics analyzed, the 1mm damage most often has the highest vibration amplitude, both for metal and ceramic bearings and with both supply frequencies. In conclusion this method can be useful for detecting the presence of a fault, but has no effectiveness in classifying the fault according to its severity.

Chapter 6

Machine Learning Method

6.1 Introduction

The use of neural networks is in continuous diffusion in modern industry, especially in the field of real-time diagnostics, where it will be used in this work. The idea behind the neural network is to provide as input a series of data characterizing a phenomenon, obtained experimentally. Such data will then be used to train the neural network in such a way as to recognize within this phenomenon a series of categories, thus allowing to classify these categories. Once trained, this neural network will be able to make this classification starting from input data collected real-time. In the case of this work, the neural network will therefore be able not only to recognize if the bearing in operation is fully functional or has defects, but also to classify the defect into two different categories (1 mm or 2 mm inner race fault). A neural network is an ensemble of neurons, just as the human brain (figure 6.1). The most simple type of neuron is called *perceptron*

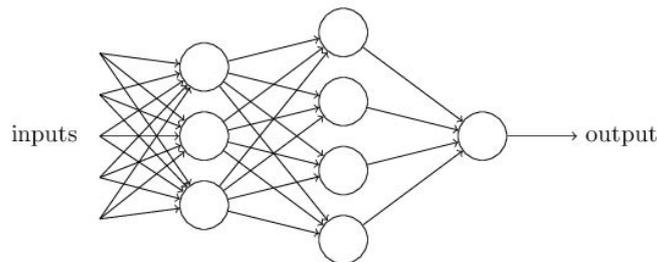


Figure 6.1: Scheme of a Neural Network [9]

(figure 6.2). A perceptron is characterized by different inputs x_i , a weight w_i for each input, a single output, that can be either 0 or 1 and a bias b_i that represents the threshold that determines if the perceptron is activated or not.

$$output = \begin{cases} 0 & \text{if } \sum(x_i \cdot w_i) + b \leq 0 \\ 1 & \text{if } \sum(x_i \cdot w_i) + b > 0 \end{cases}$$

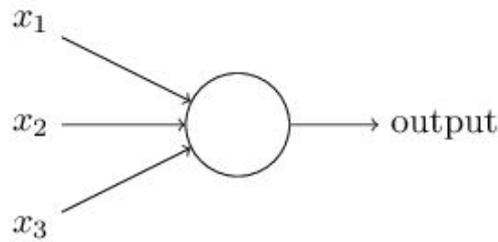


Figure 6.2: Scheme of a perceptron [9]

The problem with perceptrons is that a small change in the weights or bias of any single perceptron in the network can sometimes cause the output of that perceptron to completely flip, say from 0 to 1, thus completely affecting the accuracy of the classification [9]. This is why another type of neurons, called *sigmoids* were introduced. The scheme of the neuron is basically the same, with inputs, biases and weights, the only difference is that the output $\sigma(z)$ of the neuron can be any value between 0 and 1 (fig. 6.3).

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

where

$$z = \sum (x_i \cdot w_i) + b$$

As already said, a set of neurons form a network. Within the network, the

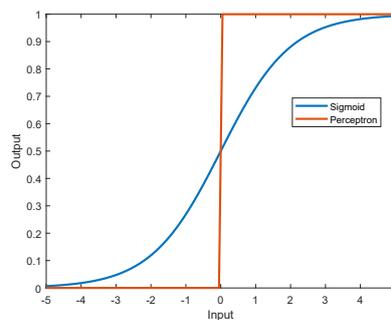


Figure 6.3: Comparison between sigmoid and perceptron output

neurons are grouped in layers. The first one (leftside of the network), is called input layer; the last one (rightside of the network), is called output layer. In the middle, between the input and the output layers there are the so called *hidden layers*.

This part of the work will be organized according to the steps of fig. 6.4 .

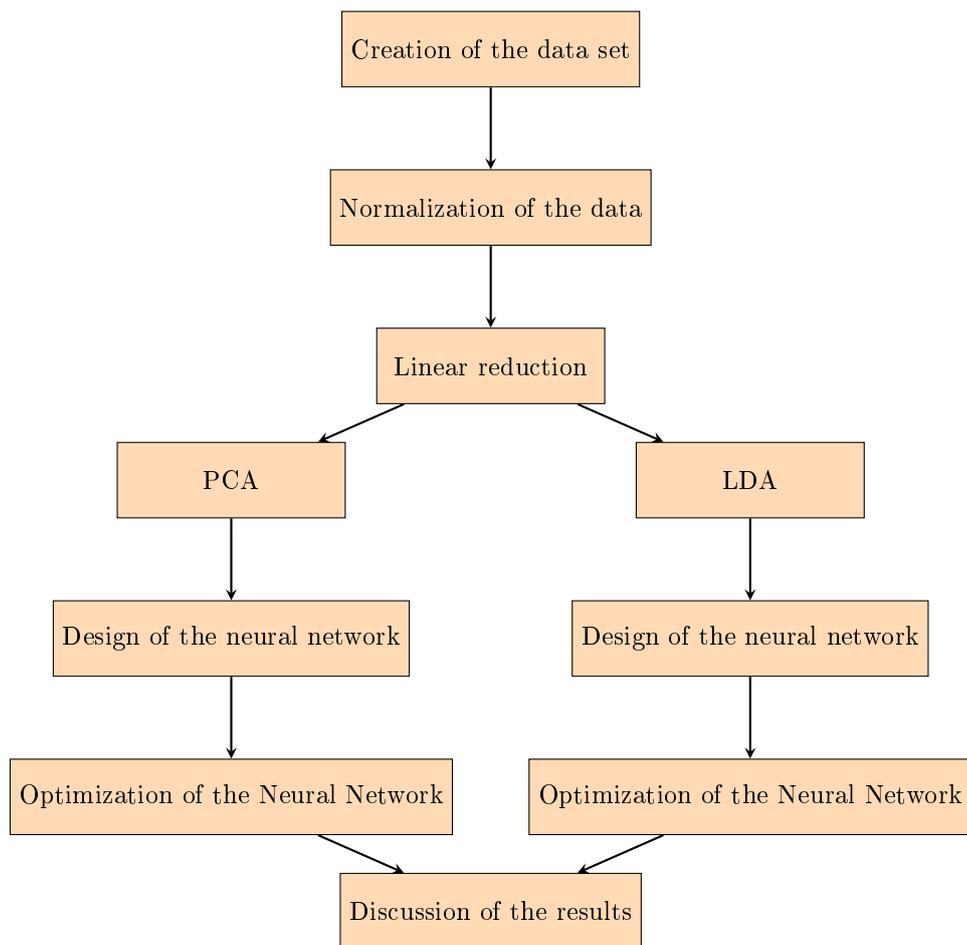


Figure 6.4: Steps followed in the development of this section.

6.2 Creation of the data base

The first step for building a neural network is to define a data set that will be used for training and the testing. In that sense is necessary to characterize the data obtained in the laboratory through a series of features that would allow to classify the bearings into the different categories. For this application were chosen 15 *time domain features* that are listed below.

1. Mean

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N}$$

2. Maximum Value

$$x_{max} = \max(x_i)$$

3. Root Mean Square

$$x_{RMS} = \sqrt{\frac{\sum_{i=1}^N x_i^2}{N}}$$

That feature is the root value of the mean squared. In vibration measurement applications, RMS gives an indication of system energy.

4. Square Root Mean

$$x_{SRM} = \left(\frac{\sum_{i=1}^N \sqrt{x_i}}{N} \right)^2$$

The SRM represent the square of the root value of the mean.

5. Standard Deviation

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N - 1}}$$

In statistics, the standard deviation is a measure of the amount of variation or dispersion of a set of values. A low standard deviation indicates that the values tend to be close to the mean of the set, while a high standard deviation indicates that the values are spread out over a wider range.

6. Variance

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N - 1}$$

Variance in statistics is a measurement of the spread between numbers in a data set. That is, it measures how far each number in the set is from the mean and therefore from every other number in the set.

7. Shape Factor RMS

$$SF_{RMS} = \frac{x_{RMS}}{\frac{\sum |x_i|}{N}}$$

RMS divided by the mean of the absolute value. Shape factor is dependent on the signal shape while being independent of the signal dimensions.

8. Shape Factor MRS

$$SF_{MRS} = \frac{x_{MRS}}{\frac{\sum |x_i|}{N}}$$

This value represents MRS divided by the mean of the absolute value.

9. Crest Factor

$$CF = \frac{x_{max}}{x_{RMS}}$$

Peak value divided by the RMS. Faults often first manifest themselves in changes in the peakiness of a signal before they manifest in the energy represented by the signal root mean squared. The crest factor can provide an early warning for faults when they first develop.

10. Latitude Factor

$$LF = \frac{x_{max}}{x_{MRS}}$$

Compare the height of a peak to MRS value of the signal.

11. Impulse Factor

$$IF = \frac{x_{max}}{\frac{\sum |x_i|}{N}}$$

Compare the height of a peak to the mean level of the signal.

12. Skewness

$$S_k = \frac{\sum_{i=1}^N (x_i - \bar{x})^3}{(N - 1) \cdot \sigma^3}$$

Asymmetry of a signal distribution. Faults can impact distribution symmetry and therefore increase the level of skewness.

13. Kurtosis

$$Kur = N \cdot \frac{\sum_{i=1}^N (x_i - \bar{x})^4}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

Length of the tails of a signal distribution, or equivalently, how outlier prone the signal is. Developing faults can increase the number of outliers, and therefore increase the value of the kurtosis metric. The kurtosis has a value of 3 for a normal distribution.

14. Normalized 5th Center of Mass

$$CM^{5th} = \frac{E[(x_i - \bar{x})^5]}{\sigma^5}$$

Just like skewness, the fifth center of mass measures symmetry of the data, but it is more impacted by outliers in the tails than the third moment (skewness).

15. Normalized 6th Center of Mass

$$CM^{6th} = \frac{E[(x_i - \bar{x})^6]}{\sigma^6}$$

The sixth center of mass measures similarly to the second moment (variance) but with an even heavier focus on outliers than the fourth moment (kurtosis).

This features were calculated for both of the significant axys, that are **x-axys** and **z-axys** and for both of the supply frequencies of $25 Hz$ and $50 Hz$. In this way the features in input at the neural network are 30. In order to obtain a big enough amount of significant data from total measurement time (equal to $100 s$), both the vibration and stator current spectres were divided into 200 windows, each one of $0.5 s$; the matrix of input data is so obtained. That matrix is composed by 200 samples for each one of the three conditions of the 30 time domain features.

Once obtained the data, the next step is to normalize them with respect to the **healthy** ones, since the healthy condition is supposed to be the one of reference for the bearings during their operative condition. The normalization was made using the **z-score** method, according to which each feature y_j of the $i - th$ sample is obtained by subtracting it for the mean value of that feature in the healthy condition and divided for the standard deviation of the same condition:

$$y_{j,i} = \frac{y_{i,j} - \bar{y}_H}{\sigma_H}$$

In addition to the data set in input, it needs to be provided to the neural network, the **targets vector**. This vector contains the classification of each sample of the data set to his own condition and is used during the training part to calculate the weights and the biases of the neural network and, during the testing part to obtain the performance of the network by comparing the targets with the results predicted by the network. The target vector has the following shape

$$\mathbf{t} = \begin{bmatrix} 1 & 0 & 0 \\ \vdots & & \\ 0 & 1 & 0 \\ \vdots & & \\ 0 & 0 & 1 \end{bmatrix}$$

These steps were made for each one of the cases analyzed that are

$$\left\{ \begin{array}{l} \text{Metallic Bearings} \\ \text{Ceramic Bearings} \end{array} \right\} \left\{ \begin{array}{l} \text{Vibration measures} \\ \text{Stator Current measures} \end{array} \right.$$

6.3 PCA technique

6.3.1 Introduction

PCA is a linear reduction technique defined as an *unsupervised* algorithm, meaning that uses input that are neither classified nor labeled for the training of the network [12]. The idea behind PCA is to find a low-dimension set of axes that summarize data, looking for some properties that strongly differ across the classes [8]. This task is achieved by solving an eigenvalue problem that allows to find the *principal components* of the initial data, that are the directions along which the data set has the maximum variance and, consequently, the highest class separability. Once found these principal components, the original data will be projected into that subspace, through a linear combination of them.

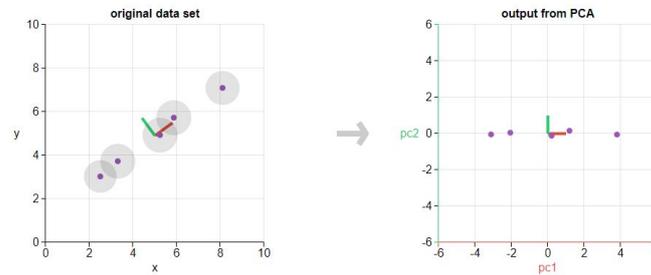


Figure 6.5: Visualization of how PCA works. [4]

The linear reduction is obtained starting from the matrix of normalized data set $[Z]$. The covariance matrix is obtained by performing $[Z]^T[Z]$. Then the eigenvalues and eigenvectors of the covariance matrix are obtained by solving the eigen problem

$$[Z]^T[Z] \cdot [V] = [V] \cdot [D]$$

where $[V]$ is the matrix with the eigenvectors as columns, and $[D]$ is a matrix with the eigenvalues λ_i on the diagonal. The principal component are obtained by sorting the eigenvalues and the correspondent eigenvectors in descending order. It was chosen to take into account only the two most significant features, that means that from the eigenvector matrix sorted, only the first two column were extracted, thus obtaining the matrix $[V^*]$. The projection of the original dataset along the principal axis is then obtained by computing the dot product

$$[Z^*] = [Z] \cdot [V^*]$$

All this procedure is obtained on *Matlab* by using the function `pca()`.

```
1 [coeff,scoreTrain,~,~,explained,mu] = pca(XTrain,'Centered',0);
2 idx = 2 ;
3 scoreTrain95 = scoreTrain(:,1:idx);
4 scoreTest95 = (XTest-mu)*coeff(:,1:idx);
```

The last step of this analysis consist in a focus on the weight coefficients of each features with which the final two features projected are obtained. Once known the three most important ones for each condition, the main goal of this step is trying to figure it out if the feature is effectively related to the fault

in exam. This means that maybe a feature can be characterized by an higher class separability, but not dued to his ability to describe the particular fault condition, but because it is affected from elements of error characterising the measurement, such as noise or offset. For this reason in the following section will be reported for each condition the three most important features with their weight coefficients.

6.3.2 Results obtained

Metallic bearings

Vibration measurements

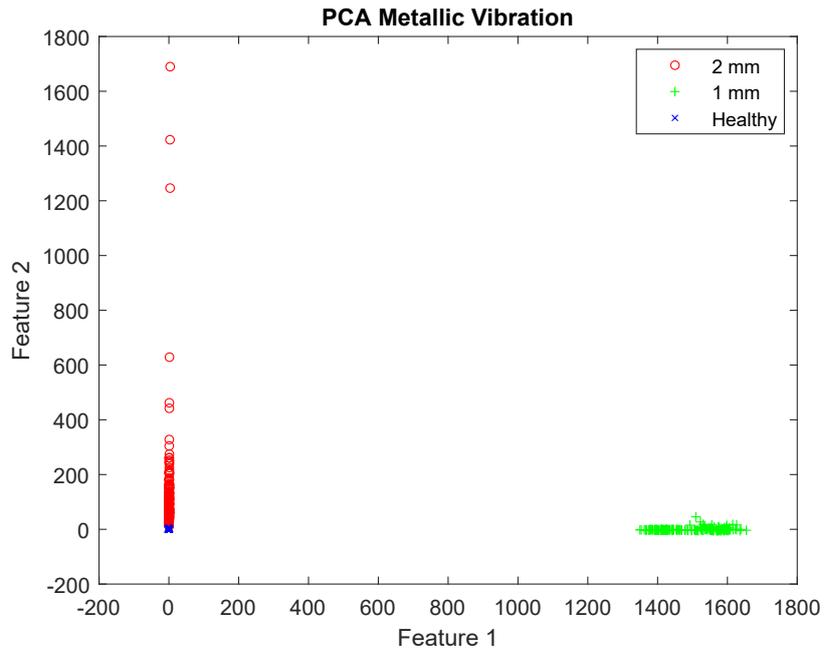


Figure 6.6: Scatter plot PCA, for vibration measurements of metallic bearings

Time Feature	Coefficient
Variance x	$9,71 \cdot 10^{-1}$
Root Mean Square x	$1,93 \cdot 10^{-1}$
Standard Deviation x	$1,39 \cdot 10^{-1}$

Table 6.1: Most important time domain features for vibration measurements of metallic bearings, with PCA reduction

Current measurements

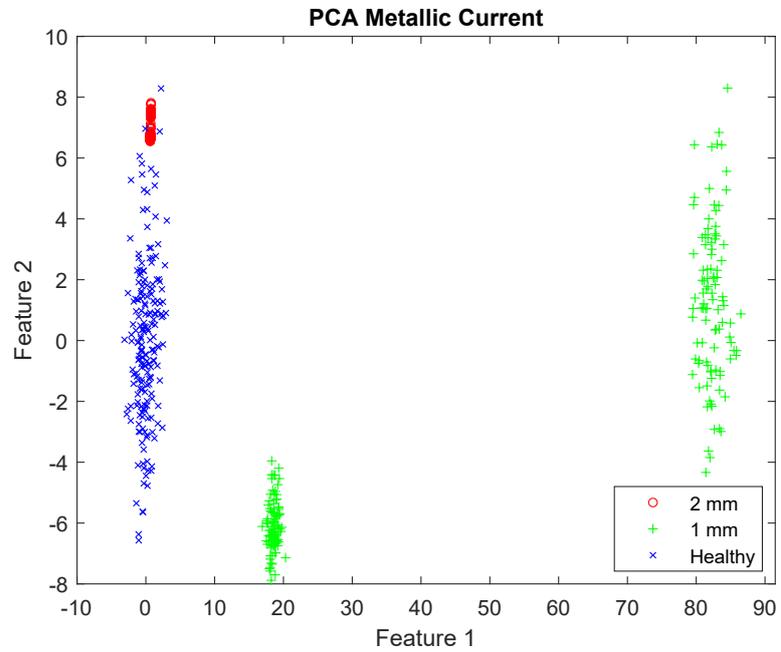


Figure 6.7: Scatter plot PCA, for stator current measurements of metallic bearings

Time Feature	Coefficient
Mean z	$8,85 \cdot 10^{-1}$
Mean x	$-4,42 \cdot 10^{-1}$
Skewness z	$-7,86 \cdot 10^{-2}$

Table 6.2: Most important time domain features for current measurements of metallic bearings, with PCA reduction,

Ceramic Bearings

Vibration measurements

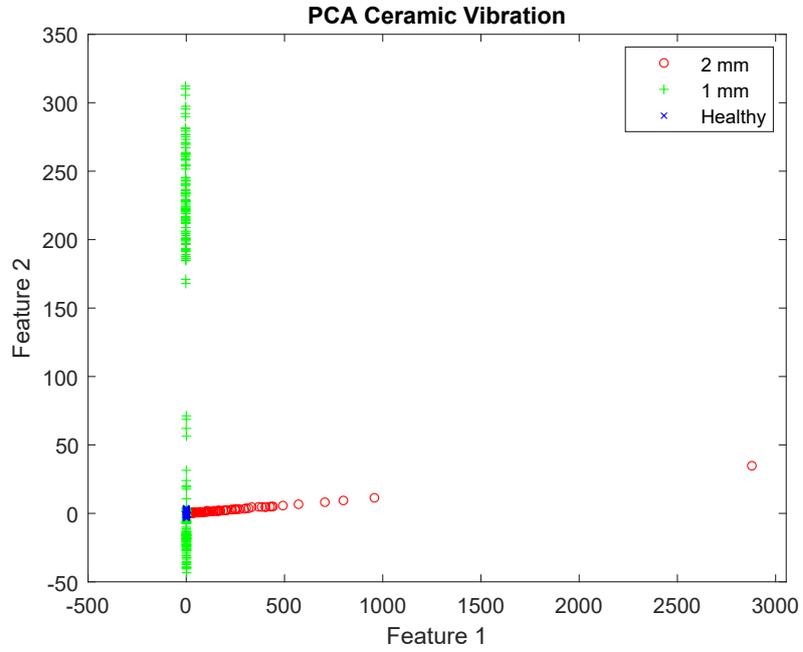


Figure 6.8: Scatter plot PCA, for vibration measurements of ceramic bearings

Time Feature	Coefficient
Normalized 6th CM z	$9,83 \cdot 10^{-1}$
Normalized 5th CM z	$1,70 \cdot 10^{-1}$
Normalized 6th CM x	$4,31 \cdot 10^{-2}$

Table 6.3: Most important time domain features for vibration measurements of ceramic bearings, with PCA reduction

Current measurements

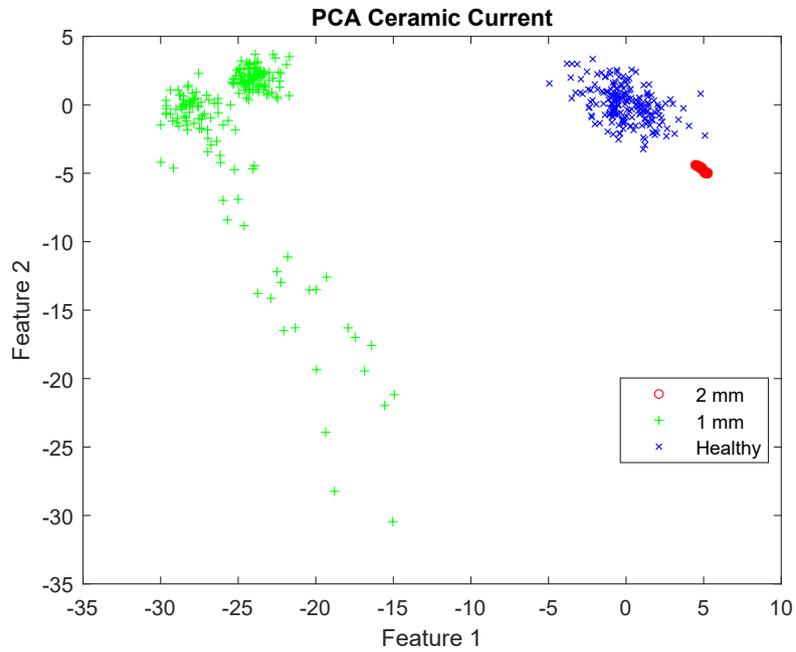


Figure 6.9: Scatter plot PCA, for stator current measurements of ceramic bearings

Time Feature	Coefficient
Mean z	$4,92 \cdot 10^{-1}$
Mean x	$3,48 \cdot 10^{-1}$
Latitude Factor x	$2,84 \cdot 10^{-1}$

Table 6.4: Most important time domain features for current measurements of ceramic bearings, with PCA reduction

General classification

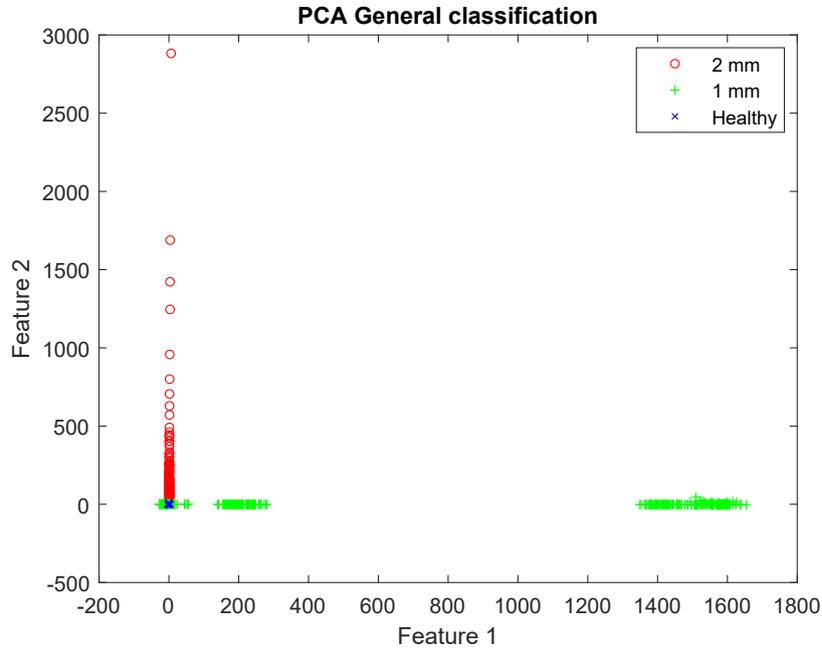


Figure 6.10: Scatter plot PCA, for all the scenarios considered

Time Feature	Coefficient
Variance x	$9,69 \cdot 10^{-1}$
Root Mean Square x	$1,43 \cdot 10^{-1}$
Standard Deviation x	$1,43 \cdot 10^{-1}$

Table 6.5: Most important time domain features for all the scenarios, with PCA reduction

The goal of analyzing the most important time features is to understand if a correlation can be founded between the statistical definition of the feature and the physics of the phenomenon. This can be crucial to understand if the class-separation provided from that feature is effectively related to the effect of different conditions on the bearing, or maybe is due to other reasons that are not related to the physics of the problem, such as noise or offset. As it can be seen, for each case taken individually, there are different time domain features that separates the most the classes. In general one can notice the following features show a good class separability:

- mean;

— root mean square;

— variance.

When during its rotation a bearing passes through a point where the single defect is present, an impact is generated which, as already seen, manifests itself within the vibration spectrum with a peak at a given characteristic presence. However, this peak does not seem to be representative, at least as far as the vibration signal is concerned. However there is another feature that changes according to the fault severity and it is the *RMS*. That feature is related to the mean energy of the vibratory signal. This mean energy changes very slowly in the early stages of damage formation, since the vibrations related are not wide, but increases dramatically with the size of the fault [13], this is why *RMS* can be considered a good class separator.

The mean value is a good class-separator for stator current measurements. It's known that the presence of the fault produces an air-gap that changes the magnetic flux through the motor, and this, in addition to the disturbance torques generated by the presence of the damage, affect the stator current signal. Thus the severity of the fault affect in different ways the air-gap and the disturbance torque so that stator current with different mean values are obtained.

The variance represents by definition the dispersion of samples with respect to the mean value. For a healthy bearing the vibration signal is expected to be more regular thus with very grouped values. But when a fault shows up then impacts happen, so that periodically some peaks are shown in the spectre. Moreover when the size of the fault become higher and higher, then the vibration signals became more random [13] and values are more dispersed, thus increasing the variance.

6.4 LDA technique

6.4.1 Introduction

The other reduction technique explored in this work is **Linear Discriminant Analysis** (LDA). As in PCA the idea behind this technique is the projection of the original dataset onto a lower dimensional space of features [16]. Instead of PCA, LDA is a **supervised** algorithm, because in LDA the goal is to find starting from the original axes, the ones that show the higher class separability, thus keeping the original features, instead of what happened for PCA.

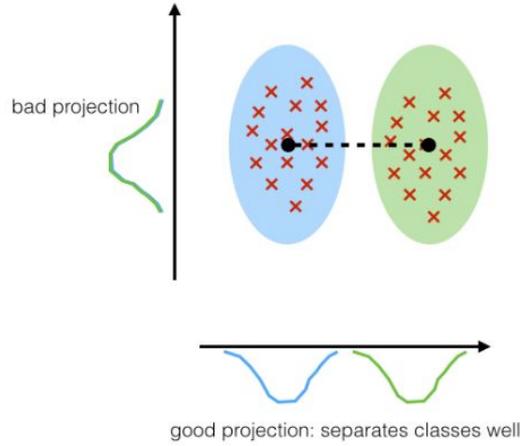


Figure 6.11: Schematization of LDA

As first step, starting from the original dataset, the **within-class** matrix (S_W) and the **between-class** matrix (S_B) were obtained.

$$S_W = \sum_{j=1}^c \sum_{i=1}^{n_i} (x_{i,j} - m_j) \cdot (x_{i,j} - m_j)^T$$

where c is the number of different conditions, n_i is the number of samples for each condition, $x_{i,j}$ is the i -th sample of the j -th condition, and m_j is the mean of the features of the j -th condition.

$$S_B = \sum_{j=1}^c (m_j - \bar{m}) \cdot (m_j - \bar{m})^T$$

where \bar{m} is the overall mean of the features. Then was solved the eigenvalues problem

$$[S_W^{-1} \cdot S_B] \cdot [V] = [V] \cdot [D]$$

where $[V]$ is the matrix with the eigenvectors as columns, and $[D]$ is a matrix with the eigenvalues λ_i on the diagonal. At this point the eigenvectors are sorted in descending order according to the corresponding eigenvalues. Then only the first two eigenvectors are taken, obtaining the matrix $[W]$. The new dataset is then obtained with performing the dot product

$$[Y] = [X] \cdot [W]$$

```

1 c = max(Targets3C);
2 m = mean(x,2);
3 for i = 1:1:c
4     tar = (Targets3C==i) ;
5     xi = x (:,tar);
6     mi = mean(xi,2)';
7     for j = 1:size(xi,2)
8         Sw = Sw +(xi(:,j)-mi)*(xi(:,j)-mi)';
9     end
10 end
11
12 for i = 1:1:c
13     tar = (Targets3C==i) ;
14     xi = x(:,tar);
15     mi = mean(xi,2)';
16     Sb = Sb+size(xi,2)*((m-mi)*(m-mi)');
17
18 end
19 [v,l] = eig(inv(Sw)*Sb,'vector') ;
20 [l,ind1] = sort(l,'descend');% sorting the eigenvectors
21 v = v(:,ind1);
22 W = v(:,1:2) ;
23 Y = x'*W ;

```

6.4.2 Results obtained

Metallic Bearings

Vibration measurements

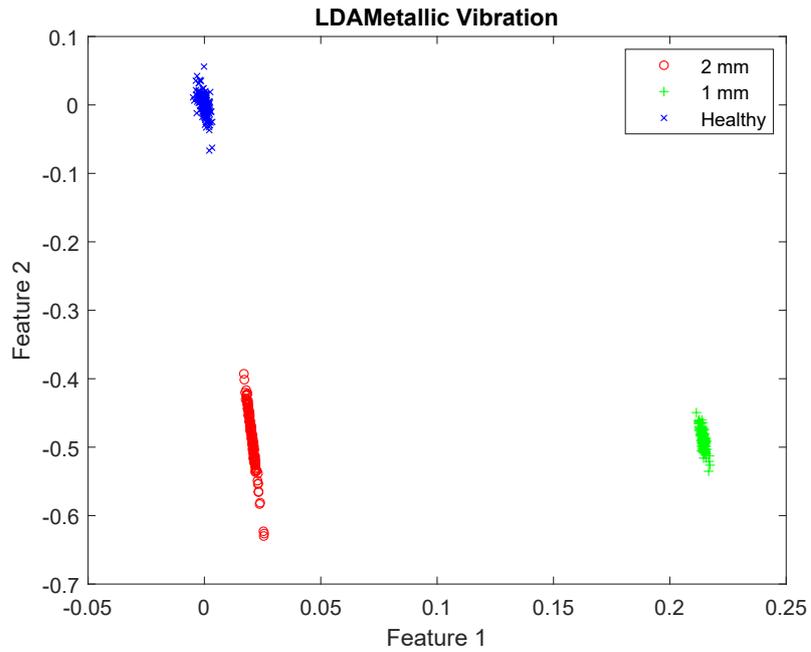


Figure 6.12: Scatter plot LDA, for vibration measurements of metallic bearings

Time Feature	Coefficient
Root Mean Square x	$7,07 \cdot 10^{-1}$
Standard Deviation x	$-7,05 \cdot 10^{-1}$
Impulse Factor x	$-3,96 \cdot 10^{-1}$

Table 6.6: Most important time domain features for vibration measurements of metallic bearings, with LDA reduction

Current measurements

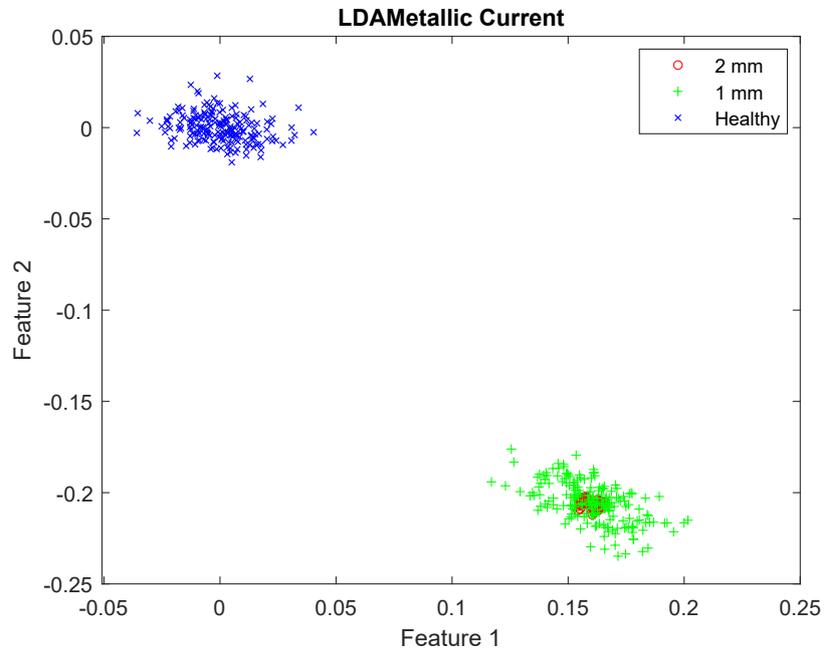


Figure 6.13: Scatter plot LDA, for stator current measurements of metallic bearings

Time Feature	Coefficient
Standard Deviation z	$-5,38 \cdot 10^{-1}$
Impulse Factor x	$-3,62 \cdot 10^{-1}$
Impulse Factor z	$-3,54 \cdot 10^{-1}$

Table 6.7: Most important time domain features for current measurements of metallic bearings, with LDA reduction

Ceramic Bearings

Vibration measurements

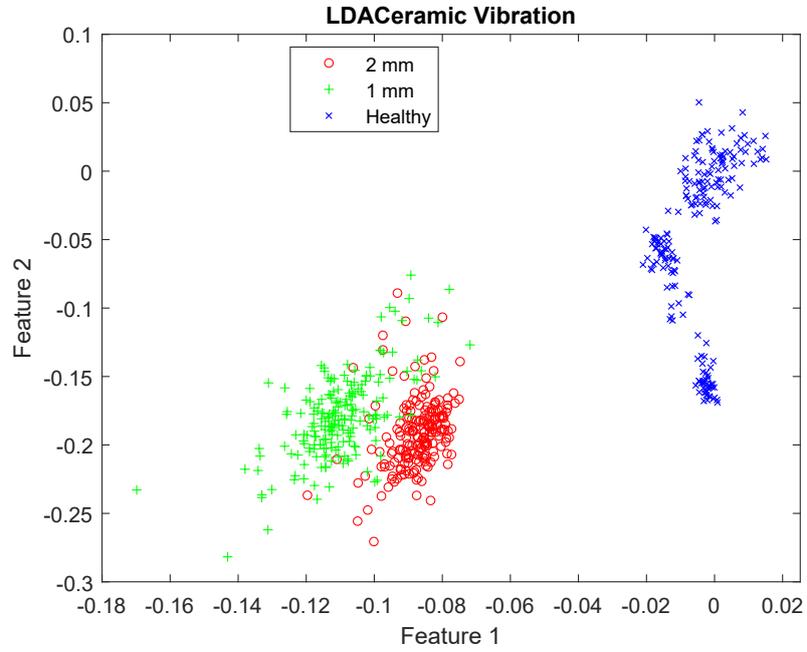


Figure 6.14: Scatter plot LDA, for vibration measurements of ceramic bearings

Time Feature	Coefficient
Impulse Factor x	$8,13 \cdot 10^{-1}$
Latitude Factor x	$-4,51 \cdot 10^{-1}$
Crest Factor x	$-3,64 \cdot 10^{-2}$

Table 6.8: Most important time domain features for vibration measurements of ceramic bearings, with LDA reduction

Current measurements

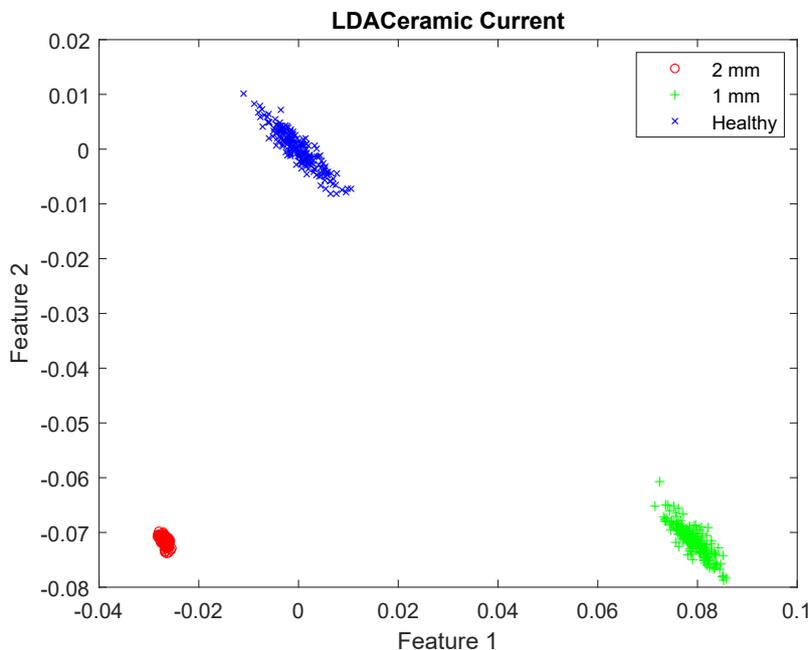


Figure 6.15: Scatter plot LDA, for stator current measurements of ceramic bearings

Time Feature	Coefficient
Standard Deviation x	$6,76 \cdot 10^{-1}$
Root Mean Square x	$-6,01 \cdot 10^{-1}$
Root Mean Square z	$-2,61 \cdot 10^{-1}$

Table 6.9: Most important time domain features for current measurements of ceramic bearings, with LDA reduction

The same analysis about the time domain features is then carried out also for the *LDA* reduction. As it can be seen from the tables reported in this case, *RMS* and *Standard deviation* shows again a very good class separability, moreover other feature related with the "shape" of the signal can be founded, such as:

- crest factor;
- impulse factor;
- latitude factor.

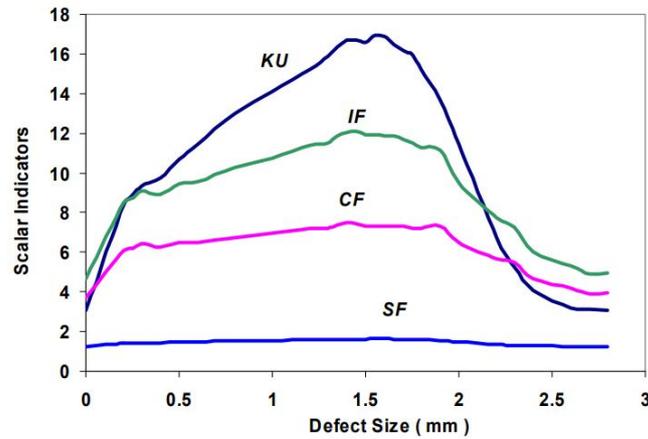


Figure 6.16: Evolution of Kurtosis, Shape factor, Crest factor and Impulse factor with respect to the size of the fault [13]

As it can be seen in figure 6.16 all these features show more or less the same behaviour. In the first part of growth of the failure, they increase (with different slopes), because basically the peak values increase but on the other hand the *RMS* and the mean value of the signals change less. But when the size of the fault grows over a certain threshold, then the *RMS* rises dramatically so the value of the features decreases. This means that these factors are no more good classifiers over a certain threshold, but as it can be seen from figure 6.16 they reach their peaks around 2 mm of fault, this is why in this case are good classifiers.

Finally the general classification confirms again, as previously for the *PCA*, that *variance* and *standard deviation* are the best classifiers.

General classification

Time Feature	Coefficient
Standard Deviation x	$5,19 \cdot 10^{-1}$
Variance x	$-5,17 \cdot 10^{-1}$
Standard Deviation z	$-4,80 \cdot 10^{-1}$

Table 6.10: Most important time domain features for all the scenarios, with LDA reduction

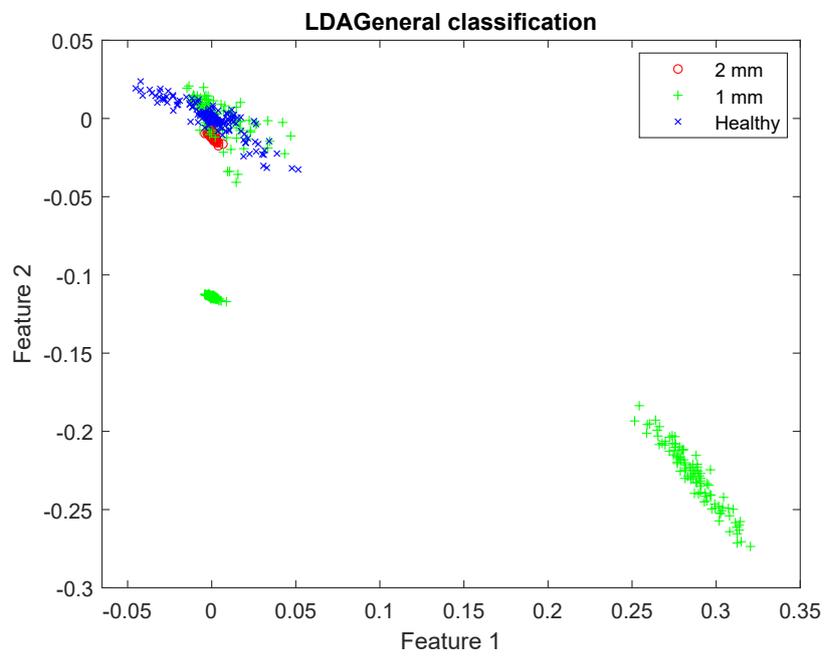


Figure 6.17: Scatter plot LDA, for all the scenarios considered

6.5 Design of the Neural Network

Once obtained the normalized data set of input and the target vector, the design of the neural network can be started. For this step the *Deep Learning Toolbox* of *Matlab* was used.

```
1 net = patternnet(hiddenLayerSize);
2
3 net.divideParam.trainRatio = 0.7; %70/100;
4 net.divideParam.valRatio = 0.15; %15/100;
5 net.divideParam.testRatio = 0.15; % 15/100;
6
7 % Train the Network
8 net.trainFcn = 'trainscg';
9 net.trainParam.epochs=epochs;
10 net.trainParam.max_fail=1500;
11 net.trainParam.min_grad=1e-10;
12
13 [net,tr] = train(net,x,t);
14
15 % Test the Network
16 y = net(x);
```

With this lines of code the neural network was generated starting from two input parameters that are:

1. **hidden layer size;**
2. **number of epochs.**

The hidden layer size, as said previously, is the number of layer between the input and the output ones, while the number of epochs are the iterations made by the network to calculate and refine the weights and biases in order to obtain the best performance. Subsequently, an analysis will be conducted on the effect of the variation of these two parameters in the network's ability to classify. The set of data is randomly divided as follows. Another important parameter to set

Training data	75%
Validation data	15%
Testing data	15%

is the training function that the network will use to update his coefficients. The one chosen is the *gradiend descendent* function. The training of the network infact is made by minimizing the so called *cost function*

$$C(w, b) = \frac{1}{2n} \cdot \sum \|y(x) - \sigma(z)\|^2$$

where $y(x)$ is the output of the target vector, $\sigma(z)$ is the output of the sigmoid neuron and n is the number of samples. So basically the cost function calculates the difference between the results expected and the ones predicted by the network. Since C is function of the weights and the biases the training of the network consists in changing these coefficients in order to minimize the cost function. The C can be seen as a surface in the 3D-space (figure 6.18), so

minimizing this function means literally "descending" that surface, this is what is made by the gradient descent algorithm. Let's call

$$\Delta v = [\Delta w ; \Delta b]$$

then from calculus it's known that

$$\Delta C(w, b) \simeq \nabla C \cdot \Delta v$$

if the Δv is chosen as follows [9]

$$\Delta v = \eta \cdot \nabla C$$

where η is a positive parameter, called **learning rate**, then is obtained

$$\Delta C(w, b) \simeq -\eta \|\nabla C\|^2$$

in this way the $\Delta C(w, b)$ is always negative so that the "descending of the valley" is always guaranteed. The learning rate weights the steps of the gradient for each iteration, if too small then a lot of iteration will be needed to find the minimum, if too big, then there will be the risk of never finding with precision the minimum. However in this work it won't be used the cost function defined before, but a more performing one, called **cross-entropy** cost function, that is defined as follows

$$C' = -\frac{1}{n} \sum [(y \cdot \ln \sigma(z) + (1 - y) \cdot \ln 1 - \sigma(z))]$$

it can be proved that, as the one defined before, the cross-entropy function is always non negative and tends to zero when $\sigma(z) \rightarrow y(x)$, so has the properties of a cost function. The most important property of the cross-entropy is that with some algebra it can be proved that

$$\frac{\partial C'}{\partial v_j} = \frac{1}{x} \sum x_i \cdot (\sigma(z) - y)$$

thus meaning that the gradient of this function is directly proportional to the error, so that the learning rate is higher at the beginning of the iterations, when the difference between the targets and the outputs are higher, and then it slows down during the training [9]. Once finished the setup, then the network is created, trained and tested. As example the data about vibration measurements of metallic bearings without linear reduction (30 parameters in input).

As it can be seen from figure 6.19 the neural network of this example has 30 neurons in the input layer, 3 neurons in the output and 10 neurons in the hidden layer. Figure 6.20 shows that the network reaches its best performance after 27 epochs. Finally in the confusion matrix (figure 6.21) the results predicted are compared with the ones expected, and since there are no values outside the principle diagonal, then the classification has an accuracy of 100% .

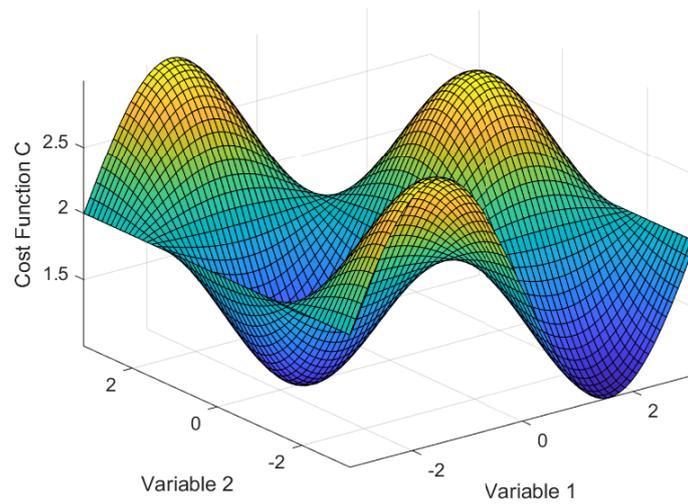


Figure 6.18: Example of a cost function

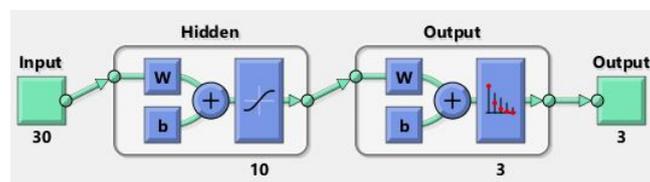


Figure 6.19: Scheme of the Neural Network

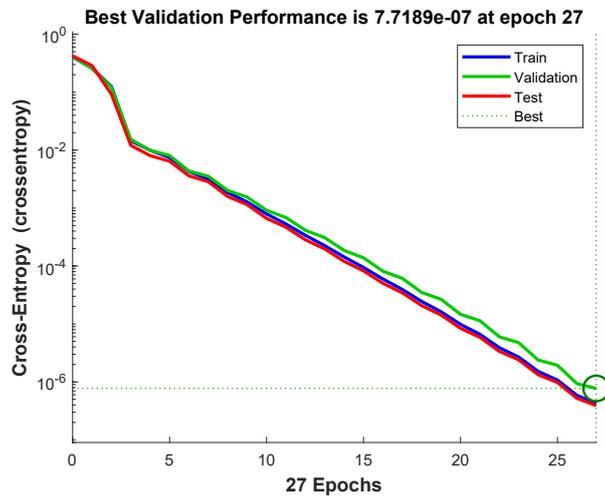


Figure 6.20: Performance of the network

Output Class	1	200 33.3%	0 0.0%	0 0.0%	100% 0.0%
	2	0 0.0%	200 33.3%	0 0.0%	100% 0.0%
	3	0 0.0%	0 0.0%	200 33.3%	100% 0.0%
		100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%
	1	2	3		Target Class

Figure 6.21: Confusion Matrix

6.6 Results Obtained

The neural network described above has therefore been applied to the data set obtained as a result of both reduction techniques, LDA and PCA. As a first attempt the following parameters were chosen:

$$\begin{cases} \text{Hidden layer size} = 1 \\ \text{Neurons in each hidden layer} = 5 \\ \text{Number of epochs} = 500 \end{cases}$$

The results of the classification obtained for both metallic and ceramic bearings are then reported.

6.6.1 PCA reduction

Metallic bearings

— Vibration measurement

PCA Confusion Matrix Metallic Vibration

Output Class	1	200 33.3%	0 0.0%	0 0.0%	100% 0.0%
	2	0 0.0%	200 33.3%	0 0.0%	100% 0.0%
	3	0 0.0%	0 0.0%	200 33.3%	100% 0.0%
		100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%
		1	2	3	
		Target Class			

Figure 6.22: Confusion Matrix PCA, for vibration measurements of metallic bearings

— Current measurement

PCA Confusion Matrix Metallic Current

Output Class	1	193 32.2%	0 0.0%	0 0.0%	100% 0.0%
	2	5 0.8%	200 33.3%	0 0.0%	97.6% 2.4%
	3	2 0.3%	0 0.0%	200 33.3%	99.0% 1.0%
		96.5% 3.5%	100% 0.0%	100% 0.0%	98.8% 1.2%
		1	2	3	
		Target Class			

Figure 6.23: Confusion Matrix PCA, for stator current measurements of metallic bearings

Ceramic Bearings

— Vibration measurement

PCA Confusion Matrix Ceramic Vibration

Output Class	1	200 33.3%	15 2.5%	2 0.3%	92.2% 7.8%
	2	0 0.0%	185 30.8%	0 0.0%	100% 0.0%
	3	0 0.0%	0 0.0%	198 33.0%	100% 0.0%
		100% 0.0%	92.5% 7.5%	99.0% 1.0%	97.2% 2.8%
	1	2	3		
	Target Class				

Figure 6.24: Confusion Matrix PCA, for vibration measurements of ceramic bearings

— Current measurement



Figure 6.25: Confusion Matrix PCA, for stator current measurements of ceramic bearings

6.6.2 LDA reduction

Metallic bearings

— Vibration measurement

Confusion Matrix LDA Metallic Vibration

Output Class	1	200 33.3%	0 0.0%	0 0.0%	100% 0.0%
	2	0 0.0%	200 33.3%	0 0.0%	100% 0.0%
	3	0 0.0%	0 0.0%	200 33.3%	100% 0.0%
		100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%
	1	2	3		
	Target Class				

Figure 6.26: Confusion Matrix LDA, for vibration measurements of metallic bearings

— Current measurement



Figure 6.27: Confusion Matrix LDA, for stator current measurements of metallic bearings

Ceramic Bearings

— Vibration measurement

Confusion Matrix LDA Ceramic Vibration

Output Class	1	200 33.3%	0 0.0%	0 0.0%	100% 0.0%
	2	0 0.0%	184 30.7%	10 1.7%	94.8% 5.2%
	3	0 0.0%	16 2.7%	190 31.7%	92.2% 7.8%
		100% 0.0%	92.0% 8.0%	95.0% 5.0%	95.7% 4.3%
		1	2	3	Total
		Target Class			

Figure 6.28: Confusion Matrix LDA, for vibration measurements of ceramic bearings

— Current measurement



Figure 6.29: Confusion Matrix LDA, for stator current measurements of ceramic bearings

6.7 Optimization of the Neural Network

The final step of this work is the optimization of the neural network. That, was carried out empirically by changing two hyperparameters that are the **hidden layer size** and the number of **epochs**. First of all the effect of the number of epochs was explored by considering a single hidden layer with 5 neurons and 5, 10, 100 and 500 epochs. Then the effect of the hidden layer size was studied, considering 5000 epochs and different networks with 1, 2, 5 and 15 hidden layers each one with 5 neurons. In the second case the number of epochs increases a lot because, since are used to train the totality of the neurons of the all hidden layers, so as the number of hidden layers increases a lot then a lot of iteration are needed to train efficiently the network. The evolution of the performances of the classifier with respect to the hyperparameters was shown only for the first case, that is vibration measurement of metallic bearing, for the others only the optimized configuration was reported.

6.7.1 PCA reduction

As it can be seen from figure 6.30 the separation improves a lot increasing the number of iterations until 100 beyond which the increase in iterations until 500 epochs further improves performance, but less so. The neural network has therefore reached its maximum efficiency and a further increase in iterations over 500 epochs would only affect computational costs, improving efficiency negligibly. On the other hand figure 6.31 shows that since a single condition is considered, then a single hidden layer is more than sufficient, infact with more layers some overfitting problems appear and the boundaries show strange shapes. So the final optimized configuration for a single scenario is:

$$\begin{cases} \text{Hidden layer size} = 1 \\ \text{Number of epochs} = 500 \end{cases}$$

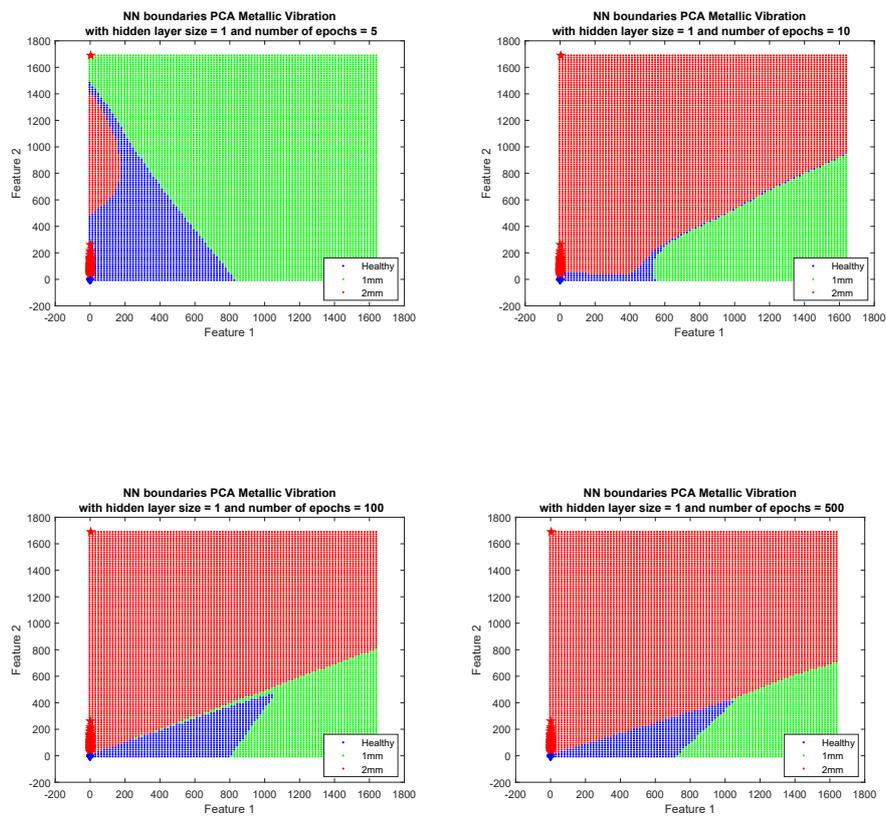


Figure 6.30: Effect of the number of epochs on the classification of the Neural Network with PCA reduction

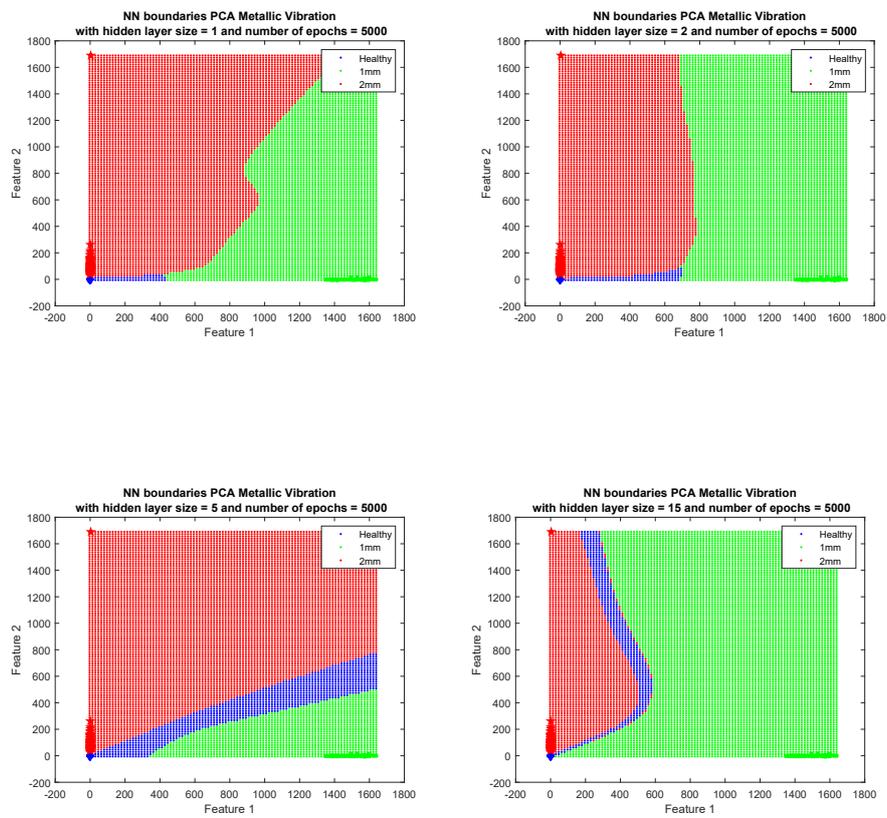


Figure 6.31: Effect of the hidden layer size on the classification of the Neural Network with PCA reduction

Metallic bearings

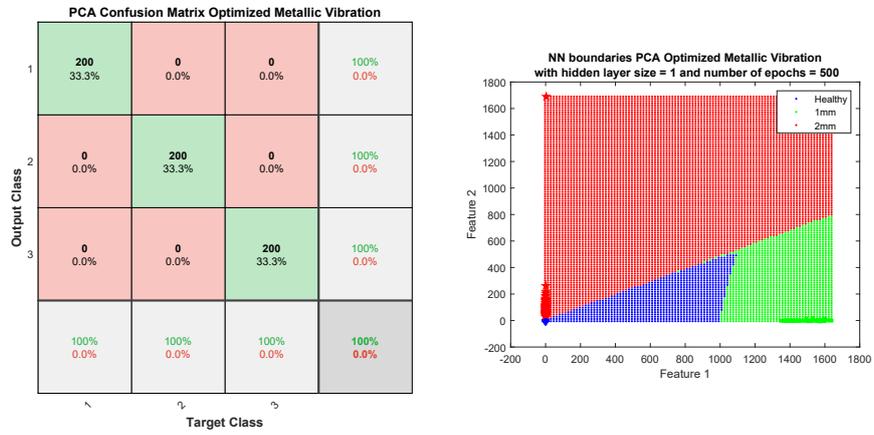


Figure 6.32: Optimized solution for PCA of metallic bearings with vibration measurement

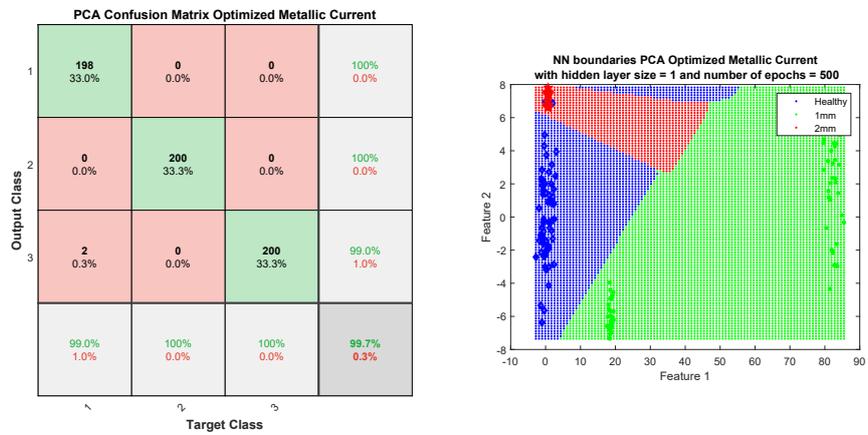


Figure 6.33: Optimized solution for PCA of metallic bearings with current measurement

Ceramic bearings

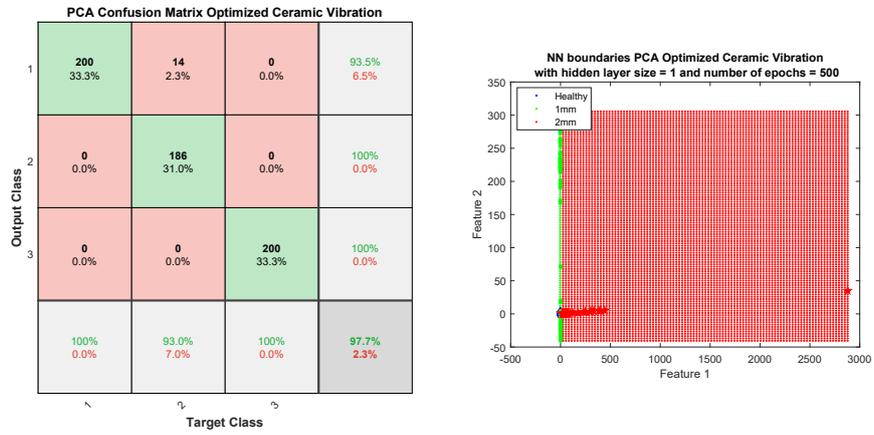


Figure 6.34: Optimized solution for PCA of ceramic bearings with vibration measurement

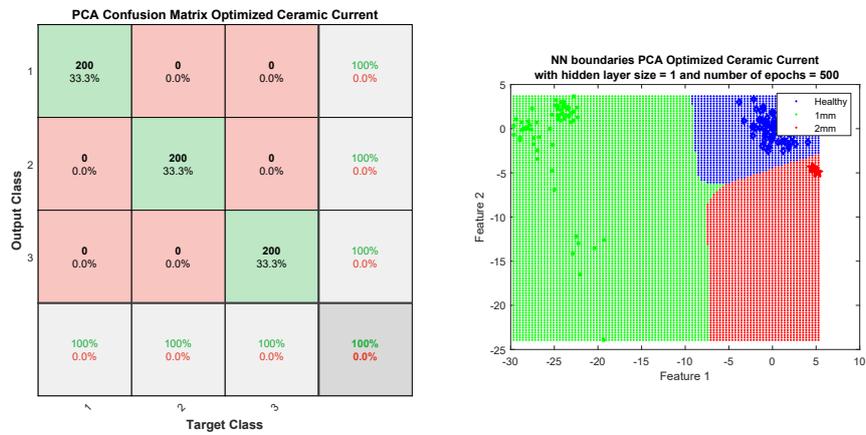


Figure 6.35: Optimized solution for PCA of ceramic bearings with current measurement.

General classification

Finally, in order to create a more *global* approach, all the scenarios concerning technology of the bearings and operating conditions, were taken into account at the same time, by using the features related to the vibration measurement. When all the four scenarios are considered at the same the complexity of the study rises and a more sophisticated neural network is needed. Infact as it can be seen from figure 6.36 with only one hidden layer the performance is still low. By analyzing the confusion matrix for each case, the optimized solution is obtained for

$$\begin{cases} \text{Hidden layer size} = 2 \\ \text{Number of epochs} = 5000 \end{cases}$$

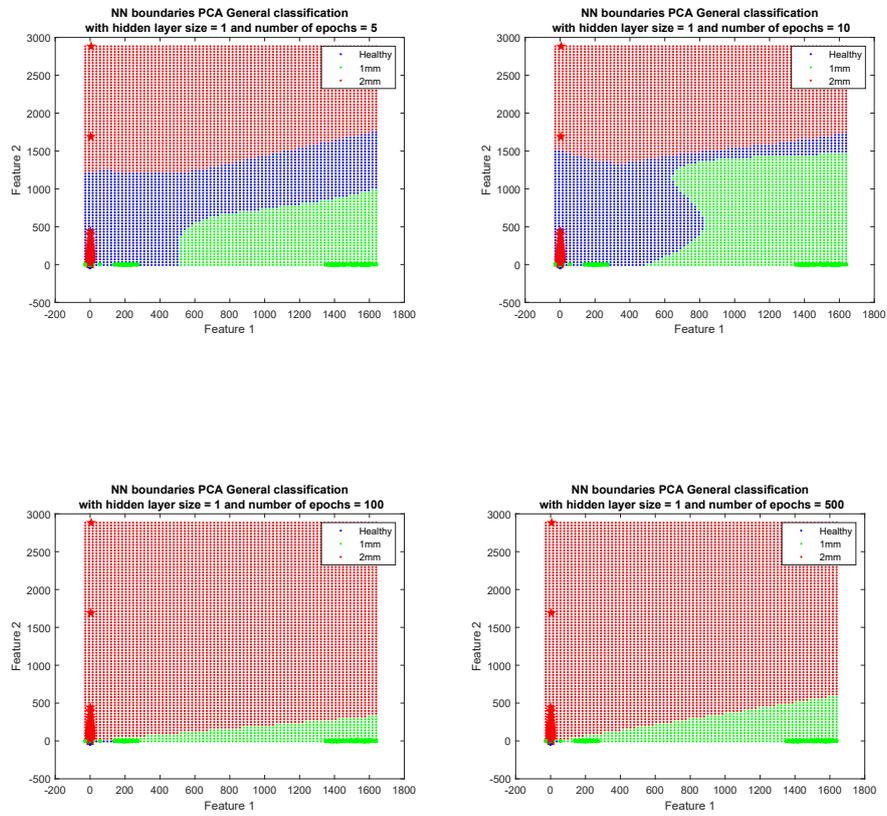


Figure 6.36: Effect of the number of epochs on the classification of the Neural Network with PCA reduction, considering all the scenarios

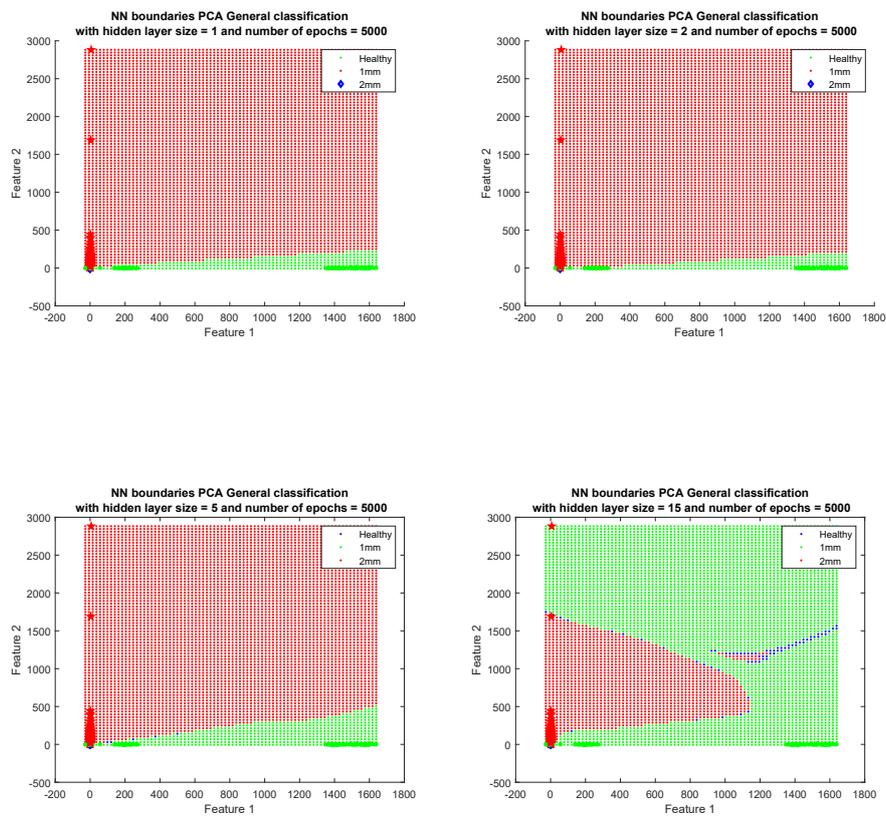


Figure 6.37: Effect of the hidden layer size on the classification of the Neural Network with PCA reduction, considering all the scenarios

PCA Confusion Matrix Optimized General classification

Output Class	Target Class 1	Target Class 2	Target Class 3	Accuracy
1	400 33.3%	16 1.3%	2 0.2%	95.7% 4.3%
2	0 0.0%	384 32.0%	0 0.0%	100% 0.0%
3	0 0.0%	0 0.0%	398 33.2%	100% 0.0%
	100% 0.0%	96.0% 4.0%	99.5% 0.5%	98.5% 1.5%

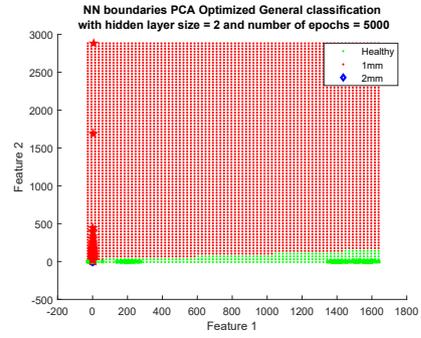


Figure 6.38: Optimized solution for PCA considering all the scenarios

6.7.2 LDA reduction

The same analysis made for the PCA reduction is here reported. The hyper-parameter values considered are 1 hidden layer with 5, 10, 100 and 500 epochs and then 5000 epochs with 1, 2, 5 and 15 hidden layers, each one with 5 neurons. The whole analysis is reported only for the vibrations measurement of the metallic bearing with vibration measurement, then only the optimized solutions are reported.

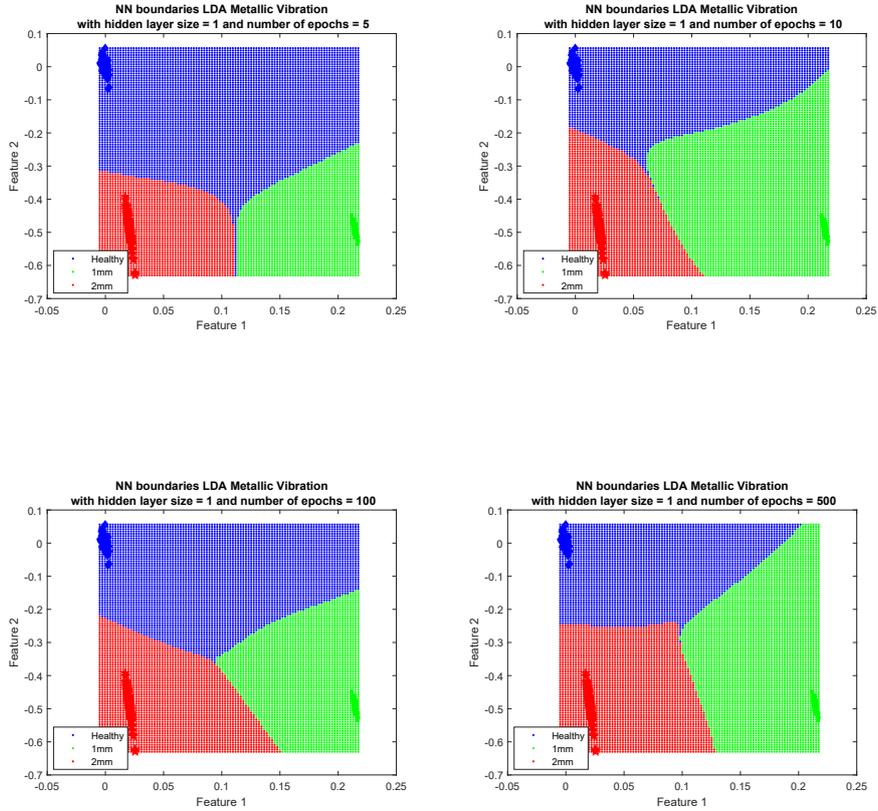


Figure 6.39: Effect of the number of epochs on the classification of the Neural Network with LDA reduction

For this case the same conclusions of the PCA can be drawn, which are that with only 1 hidden layer the best performance is reached between 100 and 500 epochs (fig. 6.39) and since the scenario is simple 1 hidden layer is more than sufficient (fig. 6.40). Thus the optimized solution for a single scenario is

$$\begin{cases} \text{Hidden layer size} = 1 \\ \text{Number of epochs} = 500 \end{cases}$$

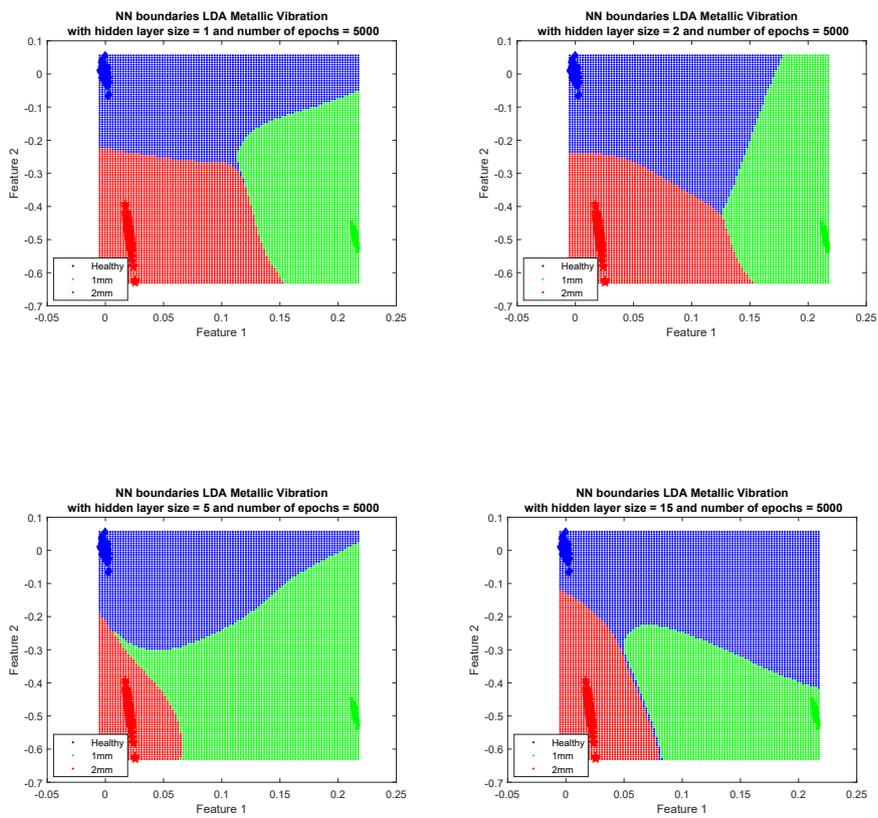


Figure 6.40: Effect of the hidden layer size on the classification of the Neural Network with LDA reduction

Metallic bearings

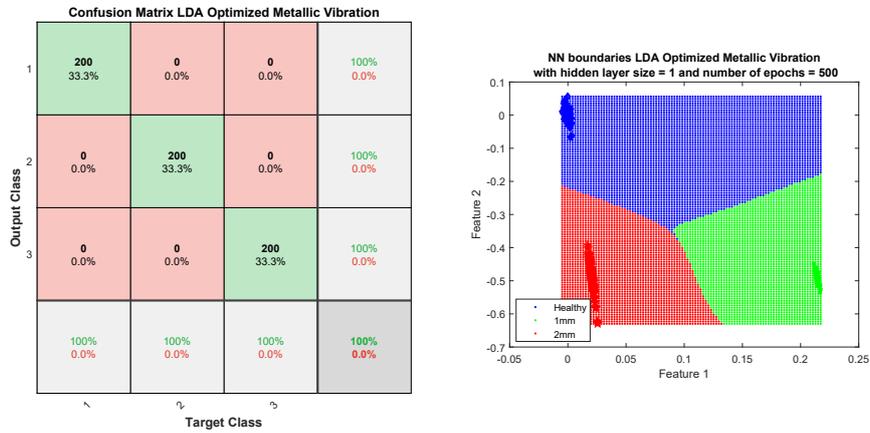


Figure 6.41: Optimized solution for LDA of metallic bearings with vibration measurement

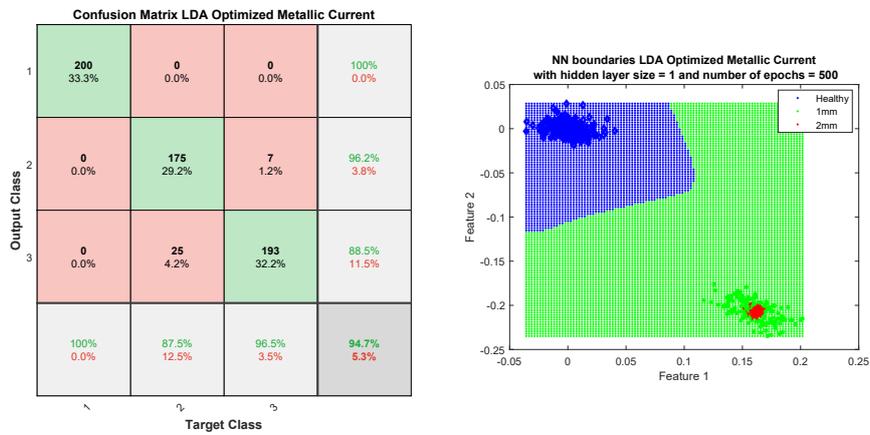


Figure 6.42: Optimized solution for LDA of metallic bearings with current measurement

Ceramic bearings

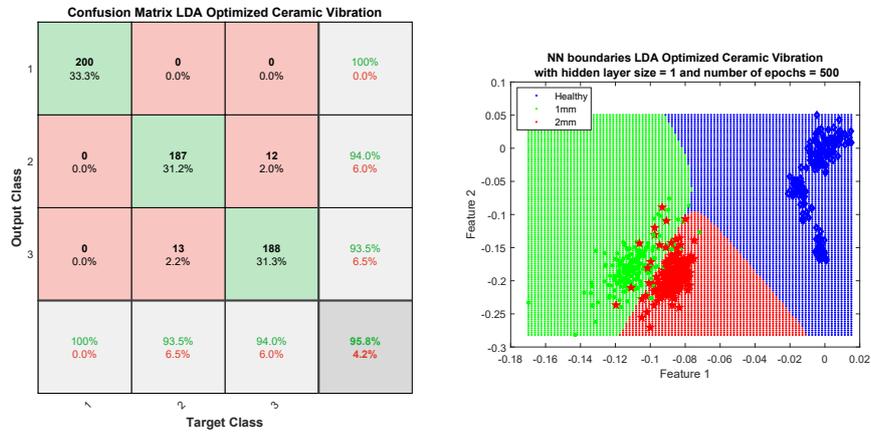


Figure 6.43: Optimized solution for LDA of ceramic bearings with vibration measurement

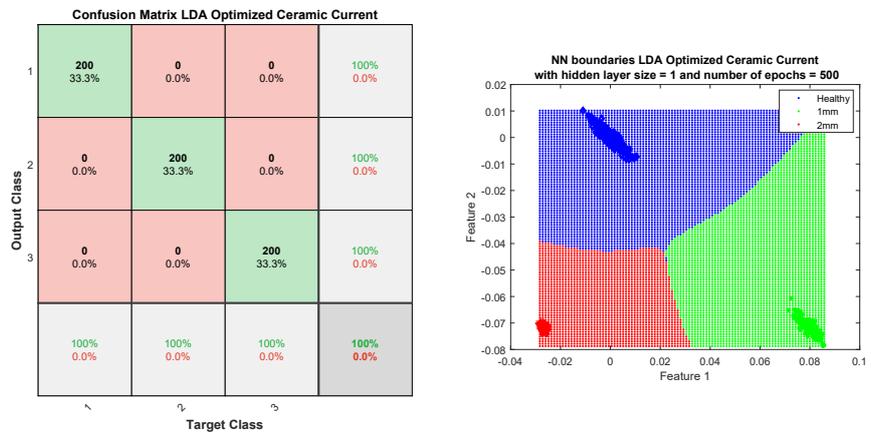


Figure 6.44: Optimized solution for LDA of ceramic bearings with current measurement

General classification

When all the four scenarios are considered at the same time, then more than one hidden layer is needed (figure 6.45 and 6.46). The confusion matrix of each combination of the hyperparameters show that the best performance is obtained for

$$\left\{ \begin{array}{l} \text{Hidden layer size} = 2 \\ \text{Number of epochs} = 5000 \end{array} \right.$$

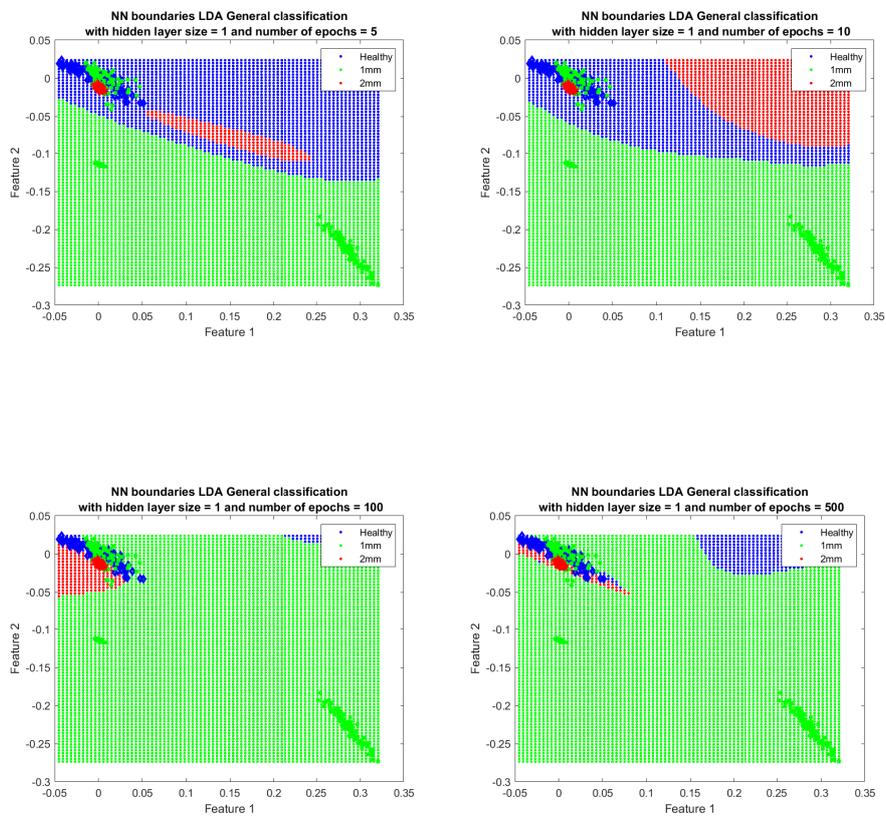


Figure 6.45: Effect of the number of epochs on the classification of the Neural Network, considering all the scenarios

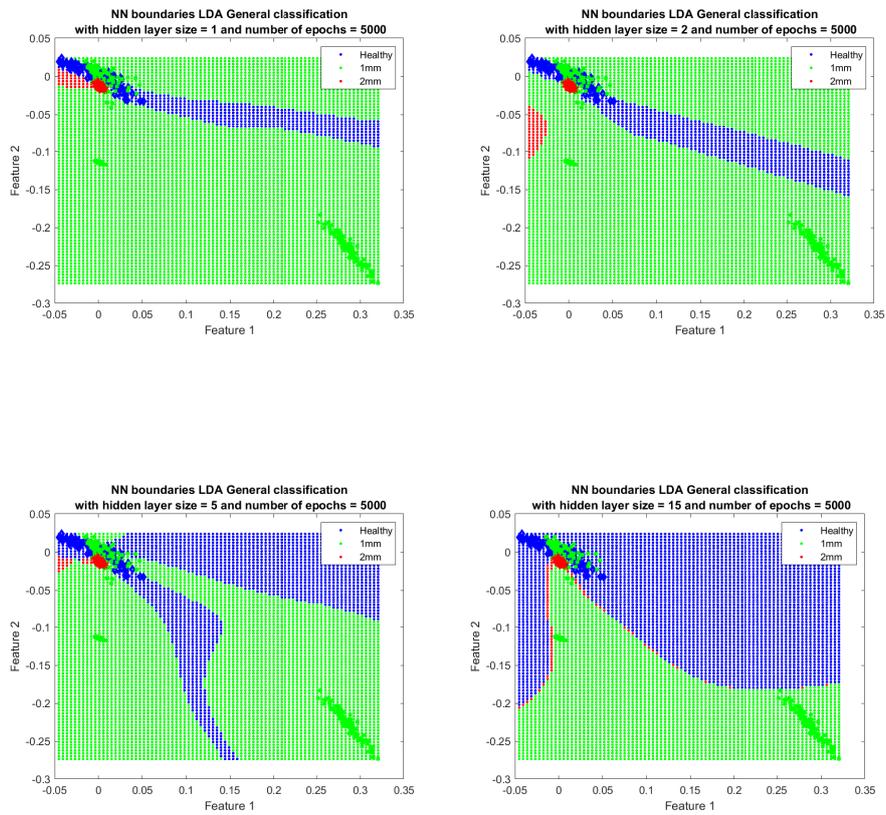


Figure 6.46: Effect of the hidden layer size on the classification of the Neural Network, considering all the scenarios

Confusion Matrix LDA Optimized General classification

Output Class	1	363 30.3%	29 2.4%	0 0.0%	92.6% 7.4%
	2	34 2.8%	366 30.5%	0 0.0%	91.5% 8.5%
	3	3 0.3%	5 0.4%	400 33.3%	98.0% 2.0%
		90.8% 9.3%	91.5% 8.5%	100% 0.0%	94.1% 5.9%
		Target Class			

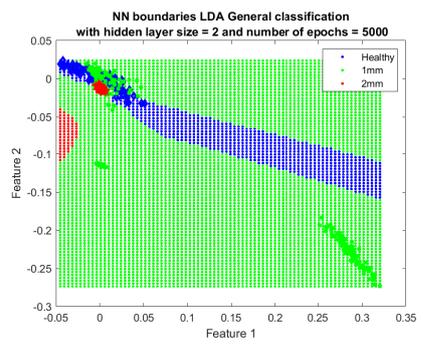


Figure 6.47: Optimized solution for LDA of ceramic bearings with vibration measurement

Chapter 7

Final conclusions

In this work, two techniques have been developed for the detection of damage on bearings operating within an electromechanical system. The two techniques presented concern the identification of characteristic frequencies of damage within the frequency spectrum of bearing vibration measured in the laboratory and the design of a neural network from previous vibration measurements, with added measurements of the induction motor stator current. The bearings are made of metal or ceramic material and may be in a healthy condition, or have a 1 *mm* or 2 *mm* bore hole at the inner raceway, specially made in the laboratory. The development of the first method made it possible to verify the validity of the classical theory concerning fault frequencies. The vibration spectra actually show peaks near these frequencies, which is why bearing vibrations can be considered a solid indicator of whether or not there are such defects in the bearing. However, this work aims not only to detect the presence of damage, but also to identify and classify it. The latter objective is not achieved with this first method because the amplitude of the peaks is not related to the severity of the damage. For this reason, a machine learning algorithm has been developed that is actually capable of implementing this classification. The designed neural network receives in input 15 different time domain features calculated from the measurements made and returns in output the damaged condition of the bearing considered. Given the number of inputs to the neural network, two reduction techniques have been developed: PCA and LDA. In this way the number of input elements has been reduced to only two, allowing a greater efficiency of the neural network and an easier visualization of the hope between classes carried out, representing the input features on a 2-D plane, for each of the studied scenarios. The results obtained with this second technique are excellent because the efficiency in the classification, considering all the different scenarios taken individually, is between 94% and 100%. As a last step of the work, the optimization of the neural network has been empirically carried out by acting on two characteristic parameters: the number of hidden layers and the number of epochs. The optimized neural network thus obtained made it possible to carry out a classification of all scenarios at the same time, with efficiencies of 94,1% for LDA and 98,5% for PCA reduction.

Future developments

As has already been explained above, this work was limited to developing a classifier referring to a single bearing model, albeit of two different materials, under quasi-stationary conditions with two different supply frequencies and with only one type of damage, i.e. a bore of different sizes in the inner bearing race. The goal for the future is to extend the designed neural network to an increasing number of cases, i.e. bearings of different sizes and number of balls and extending it to more and more different damage conditions, which differ in shape, size and position. In this way it will be possible to create a single tool to detect faults, instead of many different models, each intended for a single operating condition. Another goal for the future is the use of more sophisticated techniques for the optimization of the neural network, such as genetic algorithm, in order to find the values of the hyper parameters that guarantee the best efficiency for the neural network.

Bibliography

- [1] V. Banik, A. Chinchane and O. Sumant , "Electric Motor Market", Allied Market Reserch, 2019
- [2] A. Bellini, R. Rubini, C. Tassoni and F. Immovilli, "Diagnosis of Bearing Faults in Induction Machines by Vibration or Current Signals: A Critical Comparison", *IEEE Transactions on industry applications*, vol. 46, no.4, July/August 2010
- [3] G. Bode, S. Thul, M. Baranski, D. Muller, "Real-world application of machine-learning-based fault trained with experimental data ", *Energy*, vol.198, 2020
- [4] M. Brems, "A One-Stop Shop for Principal Component Analysis" , Medium , 2017
- [5] F. Filippetti, A.Bellini and G.A. Capolino, "Monitoring and diagnosis of rotor induction machines: state of art and future perspectives",2013 IEEE Workshop on. :196-209 Mar, 2013
- [6] C. Harlişca, I. Bouchareb, L. Frosini and L. Szabó, "Induction machine bearing faults detection based on artificial neural network," 2013 IEEE 14th International Symposium on Computational Intelligence and Informatics (CINTI), Budapest, 2013, pp. 297-302, doi: 10.1109/CINTI.2013.6705210
- [7] M. He and D. He, "Deep Learning Based Approach for Bearing Fault Diagnosis," in *IEEE Transactions on Industry Applications*, vol. 53, no. 3, pp. 3057-3065, May-June 2017, doi: 10.1109/TIA.2017.2661250
- [8] M. Lopes, "Dimensionality Reduction — Does PCA really improve classification outcome?" , Medium , 2017
- [9] Michael A. Nielsen, "Neural Networks and Deep Learning", Determination Press, 2015
- [10] M. Ojaghi, M. Sabouri, J. Faiz, "Analytic model for induction motors under localized bearing faults", *IEEE TRANSACTIONS ON ENERGY CONVERSION*, VOL.33, NO.2, JUNE 2018
- [11] S. E. Pandarakone, M. Masuko, Y. Mizuno and H. Nakamura, "Deep Neural Network Based Bearing Fault Diagnosis of Induction Motor Using Fast Fourier Transform Analysis," 2018 IEEE Energy Conversion Congress and Exposition (ECCE), Portland, OR, 2018, pp. 3214-3221, doi: 10.1109/ECCE.2018.8557651

- [12] M. Rouse, "unsupervised learning", 2017
- [13] S. Sassi, M. Thomas and B. Bardi, "Numerical simulation for vibration response of a ball bearing affected by localized defect. In Proceedings of the 5th internat. conference on acoustical and vibratory surveillance methods and diagnostic techniques", Senlis, France, paper Vol. 48, 2004
- [14] J. J. Saucedo-Dorantes, M. Delgado-Prieto, R. A. Osornio-Rios and R. d. J. Romero-Troncoso, "Industrial Data-Driven Monitoring Based on Incremental Learning Applied to the Detection of Novel Faults," in IEEE Transactions on Industrial Informatics, vol. 16, no. 9, pp. 5985-5995, Sept. 2020.
- [15] J. J. Saucedo-Dorantes, R. A. Osornio-Rios, R. J. Romero-Troncoso, M. Delgado-Prieto and F. Arellano-Espitia, "Novel condition monitoring approach based on hybrid feature extraction and neural network for assessing multiple faults in electromechanical systems," 2019 IEEE 12th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED), Toulouse, France, 2019, pp. 466-473.
- [16] R. Sebastian, "Linear Discriminant Analysis", August 2014
- [17] H. A. Toliyat, S. Nandi, S. Choi, and H. Meshgin-Kelk, "Electric Machines, Modeling, condition monitoring and Fault diagnosis." CRC Press, 2012
- [18] B. Trajin, J. Regnier, J. Faucher, "Comparison between vibration and stator current analysis for the detection of bearing faults in asynchronous drives", Electric Power applications, May 2009
- [19] "Machine Learning, what it is and why it matters", SAS Institute, 2020
- [20] "Technical Application Papers No.7 Three-phase asynchronous motors Generalities and ABB proposals for the coordination of protective devices.", 2009