Politecnico di Torino

Politecnico di Torino

Master's degree in Mechatronic Engineering

A.a. 2021/2022

Sessione di Laurea Ottobre 2022

Application of Artificial Intelligence algorithms for the classification of the health status of rotating systems

Relatori:
Prof.ssa Cristiana Delprete
Ing. Luigi Di Maggio

Candidato:
Luca Cibrario

ABSTRACT

Intelligent Fault Diagnosis (IFD) refers to the application of machine learning theories to machine fault diagnosis and seems to be a promising way to release the contribution from human labor, since it allows the automatic recognition of the state of health of machines.

The work presented in the thesis has as objective the identification, training, application and evaluation of an artificial intelligence algorithm suitable for assessing the state of health of a spherical roller bearing rotating system through the analysis of its vibration path.

The first step consists in the creation, using the approach of nonlinear multi-body models, of the dynamic model of a real bearing, which can simulate its behavior in presence of localized defects on the inner or outer raceway. The interaction between bearing elements is based on the theory of elastic contact elaborated by Hertz; the decrease in deformation of the rolling elements due to the thickness of the lubricant film and damping effects of the fluid itself according to Elastohydrodynamic Lubrication (EHL) theory are also included in the model.

Once the database of vibration paths, related to different rotational speeds, radial loads localized defects conditions, has been created, it can be analyzed in time and frequency-domain to extract the features necessary to train the artificial intelligence algorithm. At the same time, this operation is also carried out for vibration paths obtained experimentally under the same operating conditions used for numerical simulations to compare the results.

The last step of this research consists in choosing the best algorithm among those belonging to supervised learning, such as the Support Vector Machine (SVM) and the k-Nearest Neighbors (kNN), evaluating their performance on both numerical and experimental data and a finally choosing the features of greatest impact through the Shapley values method.

TABLE OF CONTENTS

LIST OF SYMBOLS

- a contact ellipse major axis
- B bearing width
- $BPFI$ Ball Pass Frequency relative to Inner Race
- $BPFO$ Ball Pass Frequency relative to Outer Race
- BSF Ball Spin Frequency
- b contact ellipse minor axis
- C damping matrix
- $\boldsymbol{C_B}$ bearing damping matrix
- CLF clearance factor
- CF crest factor
- c damping coefficient
- c_d bearing diametral clearance
- D outer diameter, dataset
- D_1 outer raceway contact diameter
- D_r rolling element diameter
- d bore diameter
- d_2 inner raceway contact diameter
- d_{IR} inner ring mean diameter
- d_m pitch diameter
- d_{OR} outer ring mean diameter
- E Young modulus
- E' equivalent modulus of elasticity
- $\bar{E}$ elliptical integral of the first kind
- E_D energy dissipated per load cycle
- E_S strain energy
- E_{train} training error
- e load factor
- F force, effective force in radial direction
- FTF Fundamental Train Frequency
- $FTFI$ Fundamental Train Frequency relative to Inner Race
- $FTFO$ Fundamental Train Frequency relative to Outer Race
- $\bar{F}$ elliptical integral of the second kind
- F_a axial load
- F_c centrifugal force
- F_{inlet} EHL inlet-zone length factor
- F_r radial load
- $F(\rho)$ curvature difference
- f force, frequency, function
- G dimensionless material parameter

- H_c dimensionless central lubricant film height
- H_d defect height
- H_{min} dimensionless minimum lubricant film height
- h hypothesis
- $\hat{h}$ predicted hypotesis
- h_{def} defect height
- h_{ehl} lubricant film thickness
- hd Vickers hardness
- I inertia modulus
- IF impulse factor
- J cost function
- J_a axial integral
- J_r radial integral
- K stiffness matrix, Kernel matrix
- $\boldsymbol{K}_B$ bearing stiffness matrix
- K_{in-ou} resultant contact stiffness
- K_l Hertzian contact stiffness for line-contact
- k stiffness, discrete sample
- $\hat{k}$ stiffness per unit length
- k_e ellipticity parameter
- $Kurtosis$ kurtosis value
- L loss function
- L_{def} defect linear extension
- L_r roller length
- l_{inlet} lubricant inlet-zone length
- M mass matrix, number of features
- M features set
- m mass
- $\hat{m}$ mass per unit width
- N number of samples
- n rotational speed
- n_h Hertzian contact exponent
- p pressure, feature
- p_0 atmospheric pressure
- p_a contact asperities pressure
- Q rolling element load
- Q_ϕ angular dependent rolling element load
- q load per unit length, gene
- R equivalent curvature
- RMS root mean square

- $RMSF$ root mean square frequency
- R_{emp} empirical risk
- r radius
- r_{in} roller internal radius
- r_{out} roller external radius
- r_r roller contour radius
- S features subset
- SF shape factor
- STD standard deviation
- $STDF$ standard deviation frequency
- s generalized coordinate, spectrum value
- $Skew$ skewness
- T temperature, period
- t time
- U dimensionless speed parameter
- u rolling speed
- u_s sum of surface velocities
- V dimensionless hardness parameter
- v linear velocity, gain function
- v_{rand} uniformly distributed random variable
- W dimensionless load parameter
- w parameter weight, gain function
- X_a maximum deflection amplitude
- x x-coordinate, feature
- $\mathbf{x}$ feature vector
- $\bar{x}$ mean vibration level
- $\overline{x_f}$ mean frequency
- x_{fc} frequency center
- y y-coordinate, label
- Z number of rolling elements
- z_{PV} viscosity-pressure ratio

Greek Symbols

- α contact angle
- α_{PV} lubricant pressure-viscosity coefficient
- α_{sensor} accelerometer sensing parameter
- β angle between the rolling element center and the defect edges, penalty constant
- β_0 regression model constant parameter
- β_j regression model weights
- $\gamma = \frac{D_r \cos\alpha}{d_m}$
- $\gamma_h = \frac{8e^{\frac{1}{2(1-\nu)}}}{\pi E' L_r}$
- Δf frequency difference
- $\Delta \mathrm{t}$ time difference
- Δx accelerometer position in x-direction
- Δy accelerometer position in y-direction
- $\Delta\theta_{def}$ defect angular extension
- δ deformation
- δ_a axial contact deformation
- δ_r radial contact deformation
- δ_ϕ angular-dependent contact deformation
- $\varepsilon = \frac{1}{2}\left(1 + \frac{\delta_a}{\delta_r}\tan\alpha\right)$
- ζ switch variable, slack variable
- η_0 kinematic viscosity
- θ_i defect centre angular position
- Λ lambda-ratio
- μ_0 dynamic viscosity
- μ_m mass per unit length
- ν Poisson ratio, contribution function
- ρ curvature
- ρ_{steel} steel density
- $\Sigma\rho$ curvature sum
- σ RMS roughness
- $\bar{\sigma}$ dimensionless roughness parameter
- $\mathbf{\Phi}$ nonlinear vector functions matrix
- ϕ angular position, nonlinear vector function, Shapley value
- ϕ_l half of the loading zone angular extension
- ψ_L loss factor
- ω angular speed
- ω_n natural frequency

Subscripts

- 0 initial position
- 1 first body
- 2 second body
- c central
- d damping
- def defect
- ehl related to elastohydrodynamic theory
- ext external
- h related to Hertzian contact theory
- i i-th element
- IR inner raceway
- j j-th element
- m related to bearing cage
- max maximum value
- min minimum value
- OR outer raceway
- r rolling element
- $roller$ rolling element
- $raceway$ related to raceways
- $slip$ slipping
- x x-axis direction
- y y-axis direction

Superscripts

- C complementar
- in inner raceway
- out outer raceway

1 INTRODUCTION

Rolling elements bearings are used almost in any kind of rotating machinery across various types of industries that include aerospace, construction, mining, steel, paper, textile, railways and renewable energy. Their damage and consequent failure cause significant economic losses and can lead to dangerous and even mortal situations for human lives, such as in the case of an aircraft engine failure or a train derailment due to a bearing seizure.

Engineers usually use fatigue as the normal failure mode on the assumption that the bearings are properly installed, operated and maintained. However, thanks to the improvements in manufacturing technology and materials, bearing fatigue life is not the limiting factor and accounts for less than 3% of failures in service [1]. In fact, many bearings may fail prematurely in service due to faulty installation (misalignments, unbalances and poor fittings), poor maintenance (contamination, insufficient lubrication) and handling practice, that lead to vibration. [2]

Therefore, the most important and, unfortunately, the most expensive task in term of labour time and cost in machinery maintenance is fault detection and diagnosis. Fault detection is the process of observing the measured system data and status information and comparing them with a range of observed attributes, coming from healthy conditions of the machine, to determine if they fall outside the acceptable limits. Diagnosis is the process of determining the state of failing components and identifying the causes of failure. [3]

The recognition of bearing health problems can be caught by experienced engineers that can diagnose faults depending on the abnormal sound or locating them using advanced signal processing method to analyse vibration signals. [4]

Vibration monitoring has now become a well-accepted part of many planned maintenance regimes and relies on known characteristic vibration signatures which rolling bearings exhibit as the rolling surfaces degrades. However, in most situations bearing vibration cannot be measured directly, so the vibration signature is affected by noise coming from the structure and other equipment of the system, such as electric motors, gears, belts, hydraulics and structural resonances. This makes the analysis of vibration data difficult even for an experienced engineer and this could lead to misdiagnosis resulting in unnecessary maintenance, parts replacements, machine downtime and consequent costs. [1].

For that reason, machine users would prefer an automatic method to shorten the maintenance cycle and improve the diagnosis accuracy. The solution is represented by Intelligent Fault Diagnosis (IFD), that uses machine learning theories, such as support vector machine (SVM), k-Nearest Neighbour (kNN) or Artificial Neural Networks (ANN), to compute fault diagnosis, relying on previously collected data from the machine under test, instead of using the experience and knowledge of engineers. [4]

Since the collection of a wide a set of vibration data coming from the machine under test may result in a huge waste of time and money to perform an accurate testing activity on a sufficient wide range of load, rotational speed and defect-type conditions, the development of a suitable dynamic model, which is able to simulate the mechanical system behavior, could lead to a great improvement in IFD. Machines could be tested even if they have recently been installed, because an experimental database of vibration paths would not be necessary.

1.1 RESEARCH OBJECTIVE

This thesis is the report of the modelling and analysis work done to setup, train and test an artificial intelligence algorithm able to classify the health status of rotating mechanical systems with spherical roller bearings.

The first activity was aimed at the identification and implementation of a suitable model to simulate the behavior of the bearing under different loads, shaft rotational velocities localized defect conditions. The developed model was then used to create a database of numerical vibration paths related to several operating and damage conditions to train the AI algorithm.

The second part of the work was dedicated at the identification of the most suitable machine learning algorithm among those belonging to supervised learning, such as Support Vector Machine (SVM) and k-Nearest Neighbor (kNN), to classify the type of defect. Firstly, vibration paths were segmented to get a wider set of data and then the most relevant features among those of time and frequency domains were extracted from each of them; therefore, a database of examples labelled with type of defect each time segment is referred to was generated. Secondly, a number of examples, taken randomly from the database, was selected to train the AI algorithms. Finally, they were tested on a new set of randomly selected data to evaluate they accuracy. Machine learning algorithms were trained and tested using numerical data coming from the bearing model and then with experimental data taken from a bearing test rig. Testing accuracy was also evaluated for algorithms trained with numerical data and tested on experimental ones, to verify whether simulations could be used to test the health status of machine, eliminating the need to carry out numerous tests on the machine to build a database for diagnostic operations.

The final task of the thesis was aimed at the selection of the most relevant features for fault detection using the Shapley values method.

1.2 THESIS OUTLINE

The thesis is divided in two main parts that present the theoretical background, computations and the discussion of obtained results regarding the two main tasks of the work: the development of a dynamic model of the bearing system able to simulate its vibration behavior and the individuation of a machine learning algorithm for fault detection.

Chapters 1 contains the thesis introduction, that explains the importance of machine fault detection and the contribute of AI algorithms to fault diagnostics. The research objective and the thesis outline are also presented in this chapter.

In Chapter 2 a general introduction on rolling element bearings from the historical and technical points of view are presented; the main characteristics of spherical roller bearings, which is the type of bearing analysed in this work, is described and in the final part a general discussion of vibration and on its causes is presented.

Chapter 3 reports the main mathematical relationships regarding rolling element bearings geometry, kinematics and loads, adopting the Hertzian theory of contact; an introduction on Elastohydrodynamic Lubrication (EHL) and the empirical formulas to compute the lubricant film thickness and damping coefficient are described; finally, a lumped parameter model to simulate the resultant stiffness and damping in each contact between bearing elements is presented.

Chapter 4 describes the main types of models used to simulate the bearing behaviour and the development of the 4-DoF nonlinear multi-body dynamic model of the spherical roller bearing used in this work.

In Chapter 5 the main features of the developed model are reported and discussed with the help of plots and graph, which were obtained from numerical simulation based on the real spherical roller bearing characteristics.

In Chapter 6 the test rig, located at Politecnico di Torino and used to record the experimental vibration paths, which were obtained in the same operating conditions of the numerical simulations, is described.

Chapter 7 presents and introductory discussion on the main failure causes in rolling element bearings and the most used traditional procedures in fault detection, such as thermal, oil debris and vibration analysis. The last one then analysed in more details, since it is the one employed in this work, and most used features in time and frequency domain, which are used for fault diagnosis, are reported. In the last part of the chapter Intelligent Fault Diagnosis (IFD) is presented and the machine learning algorithms used for this work, SVM and kNN, as well as the feature selection technique using Shapley values are described.

Chapter 8 reports the main steps followed to setup, train and test AI algorithms: database generation, feature extraction, training and test phases using both numerical and experimental data. The accuracy of each algorithm is presented though plots and discussed. In the last part of the chapter the feature selection activity through Shapley values and its results are described.

Chapter 9 presents the conclusions of the work carried out.

2 ROLLING ELEMENT BEARINGS AND VIBRATION

The machine elements that support a rotating shaft against a stationary housing in rotating machinery systems are known as bearings, which, simply by the nature of their function, have proven to be the most critical elements governing the overall system performance.

Of the many bearing types, rolling bearings have been popular because of their low power consumption, high stiffness, and high load support capabilities. For example, while the high-load capabilities have been effective for high-speed turbine applications, the high-stiffness and low-friction characteristics of rolling bearings have promoted their use in super precision, high-speed spindles and guidance systems. In general, rolling bearing applications cover a wide range of operating loads and speeds. [5]

2.1 HISTORICAL BACKGROUND

After the invention of the wheel, it was learned that less effort was required to move an object on rollers than to slide it over the same surface. Even after lubrication was discovered to reduce the work required in sliding, rolling motion still required less work when it could be used. Although the concept of rolling motion was known and used for thousands of years, and simple forms of rolling bearings were in use ca. 50 AD during the Roman civilization, the general use of rolling bearings did not occur until the Industrial Revolution.

The universal adoption of rolling bearings by design engineers was impeded by the inability of manufactures to supply rolling bearings that could compete in endurance with hydrodynamic sliding bearings. Until approximately 1940 the design and application of bearings could be considered more an art than a science, because only little was understood about the physical phenomena that occur during their operation. In the following years a more scientific approach led to a better understanding of how they perform under varied operation conditions. However, most information and data contained in manufacturers' catalogs were almost entirely empirical, because they were obtained from testing activities, based on information contained in the American National Standard Institute (ANSI) or International Organization of Standards (ISO) publications. For that reason, those data were obtained only for slow speed, simple loading and nominal operating temperatures.

The concept of rolling bearing fatigue life and its prediction was seen for the first time in the book *Ball and Roller Bearing Engineering* by Arvid Palmgren, Technical Director of ABSKF form many years. Palmgren and Gustav Lundberg, Professor of Mechanical Engineering at Chalmers Institute of Technology in Goteborg, Sweden, wrote the theory and equations on which ANSI and ISO based their standards for the calculation of rolling bearing fatigue life.

Starting from 1960 the situation changed, thanks to the development of superior rolling bearing steels and the constant improvement in manufacturing, providing extremely accurate geometry and assemblies with long lives.

Along with the technical improvements, much research was conducted on rolling bearing and rolling contact, using modern laboratory equipment, such as scanning and transmission electron microscopes, x-ray diffraction devices and digital computers, that permitted a better comprehension of mechanical, hydrodynamic, metallurgical and chemical phenomena involved in bearing operations. Therefore, many technical papers, which analyzed the performance of rolling bearings in exceptional applications, such as high-speed, heavy-load and particular design, were published. Particular attention was also given to the theory of lubrication and to rheology of lubricants.

In the 1970s the competition between ball and roller bearing manufactures for worldwide markets provided consumers with low-cost, standard design and reliable bearings. [5] [6]

2.2 ROLLING ELEMENT BEARINGS

The term rolling bearings includes all the form of bearings that utilize the rolling action of balls or rollers to permit minimum friction from the constrained motion of one body relative to another. Most rolling bearings are employed to permit the relative rotation of a shaft with respect to a fixed housing structure. [6]

A rolling bearing consists of three fundamental parts: a number of rolling elements, the inner and outer raceways and the cage. Depending on the type of bearing, the rolling elements can be balls, cylindrical rollers, spherical rollers, tapered rollers or spherical tapered rollers. Therefore, the geometry of races and cage varies with the type of bearing and this also determines the capacity to sustain radial, axial or both the types of loads.

The inner race is usually fastened to the shaft, while the outer ring is mounted in a housing; rolling elements space the two raceways and provide for smooth relative motion between them, while the cage keeps them uniformly spaced in the bearing, preventing from rubbing on each other or bunching on one side of the bearing. [7]

2.2.1 Spherical roller bearings

As Figure 2.1 shows, spherical roller bearings contain two rows of rollers, an outer ring with spherical raceways, an inner ring with two spherical raceways perpendicular to bearing axes and a cage. The locking feature given by geometry makes the inner ring captive within the outer ring in the axial direction only.

Thanks to the two rows of rolling elements, this kind of bearing have a high load-carrying capacity both in axial and in radial directions, making it irreplaceable in many heavy industrial applications.

Thanks to symmetrical rollers and spherical raceways, spherical roller bearings compensate shaft tilting and misalignments without increasing friction or reducing the service life. For that reason, they are used in those applications, where the shaft may need to change the alignment of its axis or to carry heavy loads. [8]

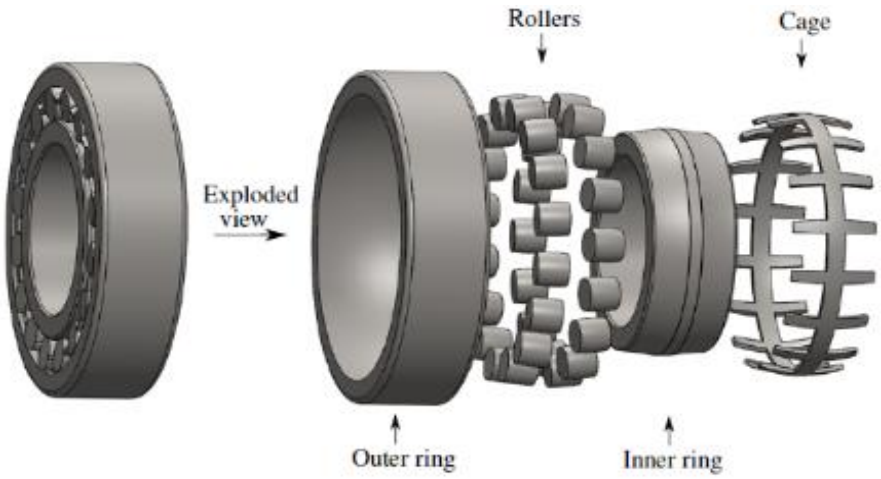

Figure 2.1: Spherical roller bearing components

The spherical roller bearing used for this research is the SKF 22240 CCK/W33 and its characteristics are reported in Figure 2.2 and in Table 2.1. [9]

d	200 mm
D	360 mm
B	98 mm
d_2	238 mm
D_1	313 mm
b	16.7 mm
K	9 mm
$r_{1,2}$	min. 4 mm

Table 2.1: Dimensions of SKF 22240 CCK/W33 bearing

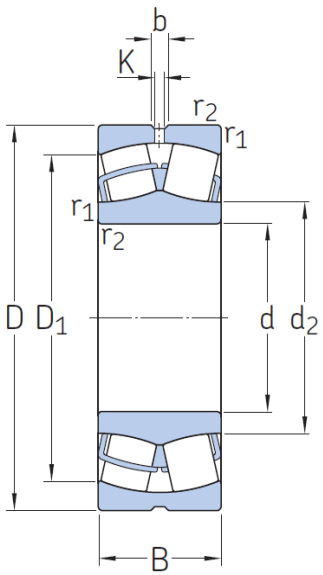

Figure 2.2: SKF 2240 CCK/W33 bearing scheme

2.3 VIBRATION

Rolling element bearings are often used in noise sensitive applications, so vibration is becoming increasingly important from both an environmental perspective and because it is synonymous with quality. In fact, vibration monitoring has now become a well-accepted part of many planned maintenance regimes and relies on the vibration signatures that bearings show as the rolling surfaces degrade.

Rolling contact bearings are a complex vibration system, whose components (rolling elements, raceways and cage) interact to generate complex vibration signatures. Even if bearings are manufactured using high precision machine tools under strict cleanliness and quality control, however they present a degree of imperfection and generate vibration as the surfaces interact through a combination of rolling and sliding. The level of vibration depends on many factors, such as the energy of impact, the mensuration point and the construction of the bearing. [1]

2.3.1 Variable compliance

Vibration is generated in a rotating rolling element bearing under radial or misaligned loads, even if it is geometrically perfect, because the external load is supported by a discrete number of elements (balls or rollers), whose position with respect to the line of action of the load continually changes with time.

In fact, as bearing rotates, individual ball loads, generated though elastic deflection at rolling element-raceway contact areas, change, producing relative movements between the inner and outer races. That movement takes place in two dimensions in case of radial loads, while it is three-dimensional under combined radial and axial loads and under misalignments; it is also periodic, because depends on the rolling elements pass-frequency across the load zone.

Variable compliance vibration is dependent on the number of rolling elements (as their number grows, vibration becomes weaker), but it can be reduced applying a correct axial preload. [1]

2.3.2 Geometrical imperfections

Because of the nature of manufacturing processes, a certain degree of imperfection is always present and it also depends on the accuracy class of the bearing. Therefore, the control waviness and surface finishing of components during the manufacturing process is critical since it can have a significant impact on vibration level and may affect bearing life.

Geometrical imperfections are considered in terms of wavelength compared with the width of the rolling element-raceway contact areas: surface features of the same order or less with respect to the contact width are called roughness, others are called waviness.

- *Roughness*: surface roughness is a significant source of vibration when its level is high with respect to the lubricant film thickness in the contact area between rolling elements and raceways. In this case surface asperities can break through the lubricant and interact directly with the opposing surface, resulting in a metal-to-metal contact.

A common parameter used to estimate the degree of asperity interaction is the lambda ratio, defined as:

$$\Lambda = \frac{h_{ehl}}{\sqrt{\sigma_{roller}^2 + \sigma_{raceway}^2}} \tag{2.1}$$

where h_{ehl} is the lubricant film thickness, σ_{roller}^2 is the RMS roughness of the rolling elements and $\sigma_{raceway}^2$ is the RMS roughness of raceways.

As it can be seen from Figure 2.3 and Figure 2.4 if Λ is less than the unity the bearing probably will not satisfy the estimated design life because of the surface distress, which can lead to a rapid fatigue failure of the rolling surfaces. The regime of mixed lubrication occurs in the range of Λ between 1 and 3, while the full elastohydrodynamic lubrication regime occurs for $\Lambda > 3$. Typical industrial applications take place in a region between those two regimes.

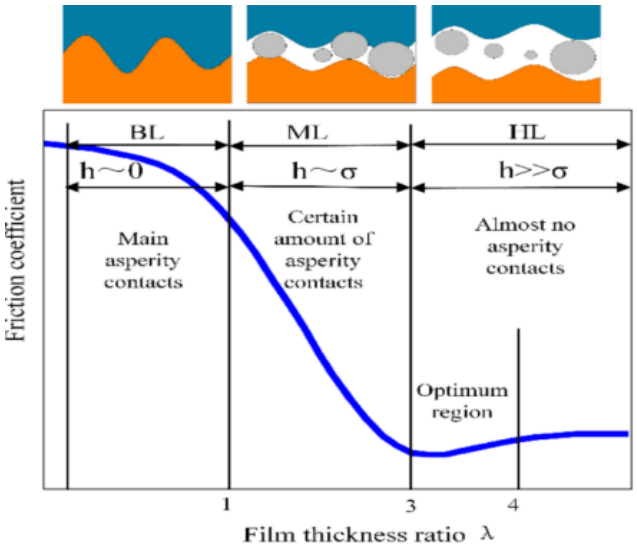

Figure 2.3: Friction coefficient versus Λ [76]

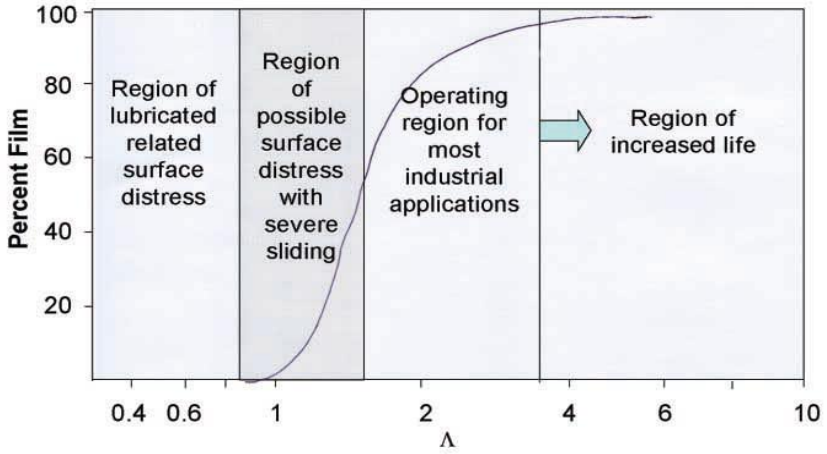

Figure 2.4: Film thickness percentage versus Λ [1]

- *Waviness*: for longer wavelength surface features and peak curvatures, which are low compared with that of the Hertzian contacts, rolling motion is continuous with the rolling elements following the surface contour. The relationship between surface geometry and vibration level is complex, being dependent upon the bearing contact geometry as well as the conditions of load and speed. In the direction of rolling, elastic deformation at contact area attenuates simple harmonic waveforms over the contact width and the level of attenuation increases as the waviness wavelength decreases (Figure 2.5); so, its minimum is reached when it becomes equal to the contact width.
  Even with modern precision machining technology, waviness cannot be eliminated completely; moreover, any geometrical errors on the shaft outside diameter or on the housing bore can be reflected on the bearing raceways, increasing vibration phenomena. [1]

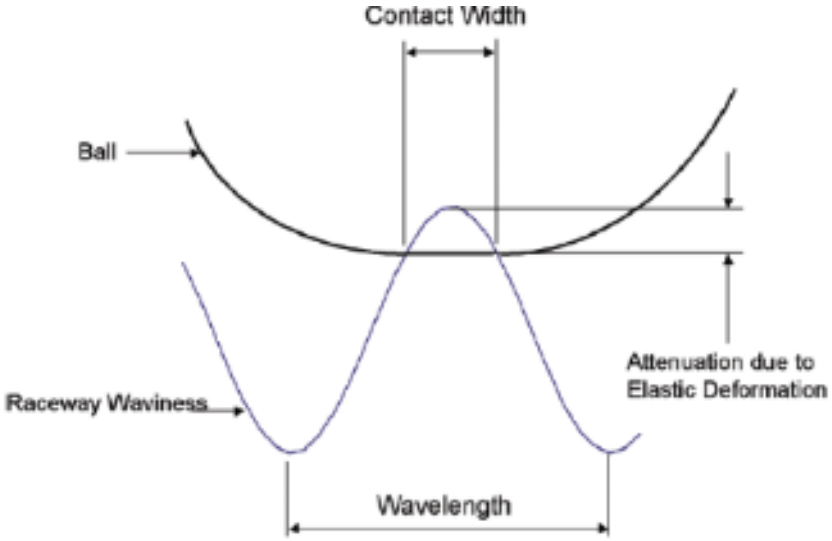

Figure 2.5: Attenuation due to contact width

2.3.3 Discrete defects

If roughness and waviness directly come from the manufacturing process, discrete defects are the consequence of damage on rolling surfaces due to assembly, contamination, operation, mounting and poor maintenance. These defects are generally really small and difficult to be detected, but they have a significant impact on vibration and can reduce the bearing expected life.

These defects can be indentations, scratches, pits, debris and particles in the lubricant. Manufacturers usually do tests on finished products to detect such defects, that are recognizable from characteristic impulsive vibrations with high peak-RMS ratio. [1]

The main types of discrete defects discussed in the present work are those on the inner and outer raceway with point dimensions.

3 ROLLING ELEMENT BEARINGS THEORETICAL BACKGROUND

This chapter provides a theoretical basis for the calculations of all the parameter involved in the dynamic equations of motions of spherical roller bearings. The following topics will be discussed:

- geometry relationships
- loads on bearing elements involving the Hertzian theory of contact
- kinematic relationships
- EHL theory and its application film thickness and damping computation
- lumped parameter model of contact in bearing elements

At a first glance, rolling element bearings appear to be comparatively simple machine elements. However, their design and adequate theoretical description are challenging problems for engineers, physicists, chemists, and mathematicians. The loads acting between the rolling elements and raceways develop only small areas of contact. Consequently, although local loading may be moderate, stresses induced on the surfaces of the rolling elements and raceways are mostly one order of magnitude higher than the maximum permissible stresses for structural components in mechanical engineering. Thus, common stress analysis methods for the design of machine elements cannot be applied to rolling bearings. Furthermore, a rolling element bearing is a nonlinear, statically indeterminate structural system.

In order to attain practically acceptable bearing lifetimes, rolling element bearings need to be lubricated. The major functions of the lubricant, in most cases mineral oils, are the following ones: reduction of friction and wear, force transmission, heat removal, corrosion protection and vibration damping. The lubricant within the highly loaded contact zone between rolling elements and raceways separates contacting surfaces and prevents metallic contact of the surface asperities. In many cases, oil film thickness is less than 1 μm thick and, thus, it is of the same order of magnitude as the surface roughness of raceways. Owing to the extreme contact pressures, the lubricant within a ball-to-raceway contact is transformed from a liquid to a solid with the consistency of a soft metal. The theory describing the lubricant film formation is called elastohydrodynamic lubrication (EHL) theory and a fast solution of the governing, highly nonlinear equations is a challenging numerical problem. [10]

3.1 GEOMETRY

In this section, the principal macro geometric relationships governing the operation of ball and roller bearings will be presented and examined. The formulas reported in this section are taken from Harris [6].

3.1.1 Pitch diameter

Bearing pitch diameter is approximately equivalent to the mean value of the outer and bore diameters, but, more precisely, it is the mean of the inner and outer ring diameters:

$$d_m = \frac{D + d}{2} \tag{3.1}$$

3.1.2 Mean diameter of the raceways

The inner and outer raceways mean diameters are defined as follows:

$$d_{IR} = \frac{d + d_2}{2} \tag{3.2}$$

$$d_{OR} = \frac{D + D_1}{2} \tag{3.3}$$

where d_2 is the inner ring external diameter and D_1 is the outer ring internal diameter

3.1.3 Diametral clearance

Spherical roller bearings are designed with diametral clearance, whose expression is the following one:

$$c_d = D_1 - D_2 - 2D_r \tag{3.4}$$

In this work the nominal clearance from SKF manual was used.

3.1.4 Curvature

The description of the contact between mating surfaces of revolution uses the concept of curvature of each body along a pair of principal axes:

$$\rho = \frac{1}{r} \tag{3.5}$$

Curvature is positive for convex surfaces and negative for concave ones.
Other two important parameters are the curvature sum and difference:

$$\Sigma\rho = \frac{1}{r_{1x}} + \frac{1}{r_{1y}} + \frac{1}{r_{2x}} + \frac{1}{r_{2y}} \tag{3.6}$$

$$F(\rho) = \frac{(\rho_{1x} - \rho_{1y}) + (\rho_{2x} - \rho_{2y})}{\Sigma\rho} \tag{3.7}$$

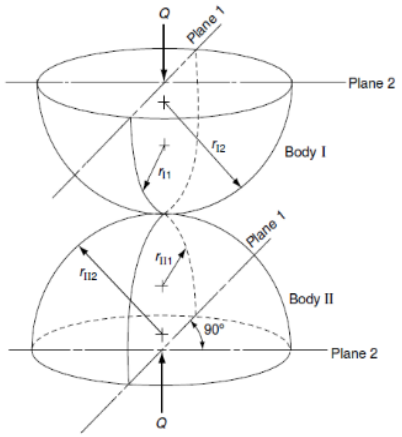

Figure 3.1: Geometry of contact bodies [6]

In case of spherical roller bearings with point contact between rollers and raceways, whose scheme is presented in Figure 3.3 and Figure 3.3, the equations for the contact between the rolling elements and the inner and outer raceways are the following ones:

$$r_{1x}^{in} = \frac{D_r}{2} \tag{3.8}$$

$$r_{1y}^{in} = r_r \tag{3.9}$$

$$r_{2x}^{in} = \frac{d_m - D_r \cos\alpha - \frac{c_d}{2}\cos\alpha}{2\cos\alpha} \tag{3.10}$$

$$r_{2y}^{in} = -r_{in} \tag{3.11}$$

$$r_{1x}^{out} = \frac{D_r}{2} \tag{3.12}$$

$$r_{1y}^{out} = r_r \tag{3.13}$$

$$r_{2x}^{out} = -\frac{d_m + \cos\alpha + \frac{c_d}{2}\cos\alpha}{2\cos\alpha} \tag{3.14}$$

$$r_{2y}^{out} = -r_{out} \tag{3.15}$$

So, denoting:

$$\gamma = \frac{D_r \cos\alpha}{d_m} \tag{3.16}$$

The curvature sum and difference are:

$$\Sigma\rho_{in} = \frac{1}{D_r}\left[\frac{2}{\gamma - 1} + D_r\left(\frac{1}{r_r} - \frac{1}{r_{in}}\right)\right] \tag{3.17}$$

$$F(\rho)_{in} = \frac{\frac{2}{1-\gamma} - D_r\left(\frac{1}{r_r} - \frac{1}{r_{in}}\right)}{\frac{2}{1-\gamma} + D_r\left(\frac{1}{r_r} - \frac{1}{r_{in}}\right)} \tag{3.18}$$

$$\Sigma\rho_{out} = \frac{1}{D_r}\left[\frac{2}{1 + \gamma} + D_r\left(\frac{1}{r_r} - \frac{1}{r_{out}}\right)\right] \tag{3.19}$$

$$F(\rho)_{in} = \frac{\frac{2}{1+\gamma} - D_r\left(\frac{1}{r_r} - \frac{1}{r_{out}}\right)}{\frac{2}{1+\gamma} + D_r\left(\frac{1}{r_r} - \frac{1}{r_{out}}\right)} \tag{3.20}$$

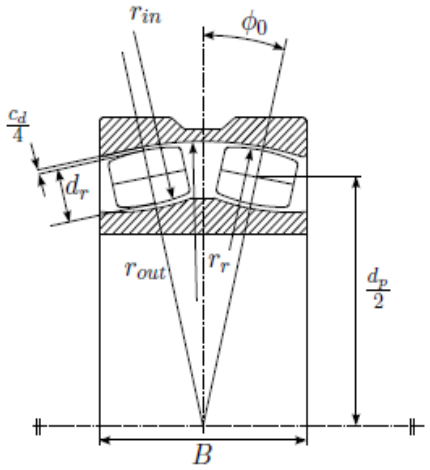

Figure 3.3: Characteristic dimensions of a spherical roller bearing [8]

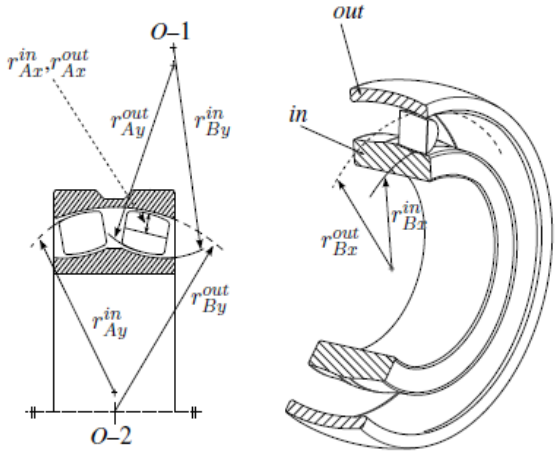

Figure 3.3: Radii of curvature between roller, outer race and inner race [8]

Since it is difficult to get reliable data about curvature radii of spherical rolling elements, in this work they are considered as cylindrical (the scheme is shown in Figure 3.4), therefore:

$$r_r = r_{in} = r_{out} \to \infty \tag{3.21}$$

So:

$$\Sigma\rho_{in} = \frac{1}{D_r}\left[\frac{2}{\gamma - 1}\right] \tag{3.22}$$

$$F(\rho)_{in} = 1 \tag{3.23}$$

$$\Sigma\rho_{out} = \frac{1}{D_r}\left[\frac{2}{1 + \gamma}\right] \tag{3.24}$$

$$F(\rho)_{in} = 1 \tag{3.25}$$

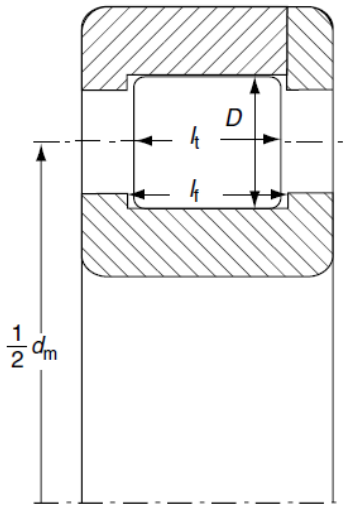

Figure 3.4: Cylindrical roller bearing schematic drawing [6]

3.2 LOADS

In this section a review of the main forces acting on bearing elements is presented.

3.2.1 External loads

Loads applied to bearings can be forces (radial, axial or combined) and moments.

In this works the external load is only a force in the radial direction, as shown in Figure 3.5, and, since only a row of rolling element is considered, the real radial force applied to the bearing assembly is:

$$F_{r,1} = \frac{F_r}{2} \tag{3.26}$$

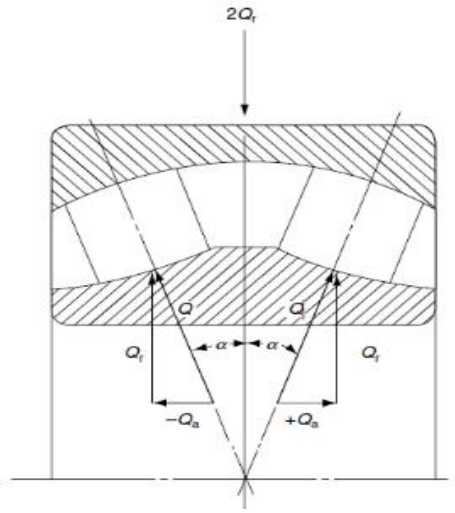

Figure 3.5: Radially loaded spherical roller bearing [6]

3.2.2 Centrifugal forces

The centrifugal force of each rolling element is expressed as:

$$F_c = \frac{1}{2} m_r d_m \omega_m^2 \tag{3.27}$$

For a cylindrical roller:

$$m = \frac{1}{4} \rho_{steel} \pi D_r^2 L_r \tag{3.28}$$

And:

$$\omega_m = \frac{2\pi n_m}{60} \tag{3.29}$$

Substituting this relation in the first one and considering steel rolling elements, the centrifugal force is:

$$F_c = 3.39 \times 10^{-11} D_r^2 L_r d_m n_m^2 \tag{3.30}$$

Since this model has been developed for low-average velocities, centrifugal forces can be neglected.

3.2.3 Contact stresses and deformations

The classical solution for local stresses and deformations of two elastic bodies contacting at a single point was developed by Hertz.

Two solids having different radii of curvature in two principal directions, like x and y, are in point contact when no load is applied to them. On the contrary, when they are pressed together by a force F, the resulting contact area becomes elliptical with a being the major axis and b the minor one, as it can be seen in Figure 3.6. [6]

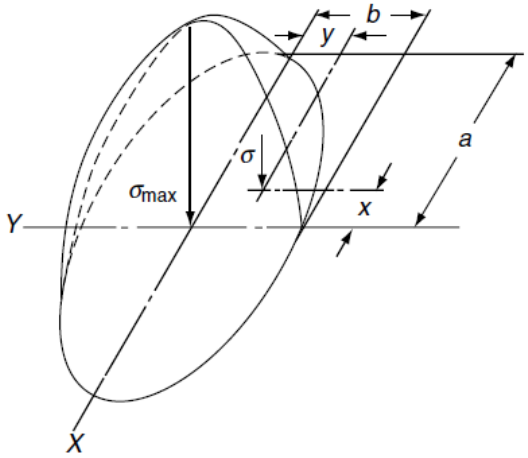

Figure 3.6: Ellipsoidal surface compressive stress distribution in point-contact [6]

The elliptic parameter is defined as:

$$k_e = \frac{a}{b} \tag{3.31}$$

It can be re-defined as a function of the curvature difference $F(\rho)$, which is the curvature difference presented in Equation 3.7, and the elliptic integrals of the first $\bar{E}$ and second $\bar{F}$ kinds:

$$k_e = \left[\frac{2\bar{E} - \bar{F}(1 + F(\rho))}{\bar{F}(1 - F(\rho))}\right]^{1/2} \tag{3.32}$$

Brewe and Hamrock [11] used numerical iteration and curve fitting techniques to find the following approximation formulas for the ellipticity parameter k_e and the elliptical integrals of the first $\bar{E}$ and second $\bar{F}$kinds as shown below:

$$k_e = 1.0339\left(\frac{R_y}{R_x}\right)^{0.636} \tag{3.33}$$

$$\bar{E} = 1.003 + 0.5968\frac{R_x}{R_y} \tag{3.34}$$

$$\bar{F} = 1.5277 + 0.6023\ln\left(\frac{R_y}{R_x}\right) \tag{3.35}$$

Where:

$$\frac{1}{R_x} = \frac{1}{r_{1x}} + \frac{1}{r_{2x}} \tag{3.36}$$

$$\frac{1}{R_y} = \frac{1}{r_{1y}} + \frac{1}{r_{2y}} \tag{3.37}$$

The ellipse axes and deformation can be obtained through the following equations:

$$a = a^*\left(3\frac{Q}{E'\Sigma\rho}\right)^{\frac{1}{3}} \tag{3.38}$$

$$b = b^*\left(3\frac{Q}{E'\Sigma\rho}\right)^{\frac{1}{3}} \tag{3.39}$$

$$\delta = \delta^*\left(3\frac{Q}{E'\Sigma\rho}\right)^{\frac{2}{3}}\frac{\Sigma\rho}{2} \tag{3.40}$$

Where:

$$\frac{1}{E'} = \frac{1}{2}\left(\frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2}\right) \tag{3.41}$$

$$a^* = \left(\frac{2k_e^2\bar{E}}{\pi}\right)^{\frac{1}{3}} \tag{3.42}$$

$$b^* = \left(\frac{2\bar{E}}{\pi k_e}\right)^{\frac{1}{3}} \tag{3.43}$$

$$\delta^* = \frac{2\bar{F}}{\pi}\left(\frac{\pi}{2k_e^2\bar{E}}\right)^{\frac{1}{3}} \tag{3.44}$$

In case of a cylindrical roller bearing (Figure 3.7), the contact between rolling elements and raceways takes place on an ideal line and $k_e \to \infty$, so the contact area degenerates into a semi-cylindrical form with dimensions $2b$ and $L_r = 2a$. The contact minor axis b and deformation δ are computed through:

$$b = \left[\frac{4Q}{\pi L_r \Sigma\rho}\left(\frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2}\right)\right]^{\frac{1}{2}} \tag{3.45}$$

$$\delta = \frac{2Q(1-\nu^2)}{\pi EL}\ln\left[\frac{\pi EL_r}{Q(1-\nu^2)(1 \mp \gamma)}\right] \tag{3.46}$$

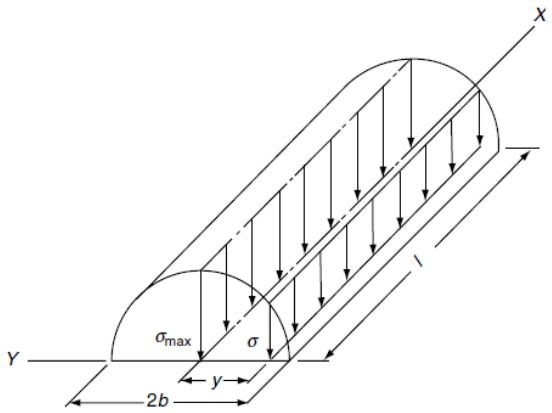

Figure 3.7: Semicylindrical surface compressive stress distribution of ideal line-contact [6]

In practice, rollers are crowned and basing on laboratory testing Palmgren [12] developed an empiric equation for contact deformation:

$$\delta = 3.81\left[\frac{1}{\pi}\left(\frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2}\right)\right]^{0.9}\frac{Q^{0.9}}{L_r^{0.8}} \tag{3.47}$$

Note that in this formula the resulting contact deformation does not depend either on the radii of rolling elements or those of raceways.
The contact force can be obtained from displacement through:

$$Q = \frac{1}{3.81^{\frac{10}{9}}\left[\frac{1}{\pi}\left(\frac{1-\nu_1^2}{E_1}+\frac{1-\nu_2^2}{E_2}\right)\right]} L_r^{\frac{8}{9}} \delta^{\frac{10}{9}} \tag{3.48}$$

So:

$$K_l = \frac{1}{3.81^{\frac{10}{9}}\left[\frac{1}{\pi}\left(\frac{1-\nu_1^2}{E_1}+\frac{1-\nu_2^2}{E_2}\right)\right]} L_r^{\frac{8}{9}} \tag{3.49}$$

3.2.4 Load distribution under combined radial and thrust load for single-row bearings

In static or quasi-static conditions, if a rolling bearing without diametral clearance is subjected simultaneously to a radial load in the central plane of the rollers and a centric thrust load, then the inner and outer rings will remain parallel and will be relatively displaced by a distance δ_a in the axial direction and δ_r in the radial one. At any position ϕ measured from the most heavily loaded rolling element, the approach of the rings is: [6]

$$\delta_\phi = \delta_a \sin\alpha + \delta_r \cos\alpha \cos\phi \tag{3.50}$$

where δ_r is the ring radial shift, δ_a is the axial shift and c_d is the diametral clearance, α is the contact angle and ϕ is the rolling element angular position.

So, the maximum deflection occurs when $\phi = 0$ and is given by:

$$\delta_{max} = \delta_a \sin\alpha + \delta_r \cos\alpha \tag{3.51}$$

Combining the two equations:

$$\delta_\phi = \delta_{max}\left[1 - \frac{1}{\varepsilon}(1-\cos\phi)\right] \tag{3.52}$$

$$\varepsilon = \frac{1}{2}\left(1 + \frac{\delta_a}{\delta_r}\tan\alpha\right) \tag{3.53}$$

The load at a given angle is:

$$Q_\phi = Q_{max}\left[1 - \frac{1}{\varepsilon}(1-\cos\phi)\right]^{n_h} \tag{3.54}$$

The sum of rolling element forces in each direction must equal the applied load in that direction to satisfy the static equilibrium conditions:

$$F_r = \sum_{\phi=-\phi_l}^{\phi=\phi_l} Q_\phi \cos\alpha \cos\phi \tag{3.55}$$

$$F_a = \sum_{\phi=-\phi_l}^{\phi=\phi_l} Q_\phi \sin\alpha \tag{3.56}$$

Where:

$$\phi_l = \cos^{-1}\left(-\frac{\delta_a}{\delta_r}\tan\alpha\right) = \cos^{-1}(1-2\varepsilon) \tag{3.57}$$

The static equilibrium equations can be rewritten respectively as functions of radial and thrust integrals:

$$F_r = ZQ_{max}\, J_r(\varepsilon)\cos\alpha \tag{3.58}$$

$$F_a = ZQ_{max}\, J_a(\varepsilon)\cos\alpha \tag{3.59}$$

Where:

$$J_r(\varepsilon) = \frac{1}{2\pi}\int_{-\phi_l}^{\phi_l}\left[1-\frac{1}{2\varepsilon}(1-\cos\phi)\right]^{n_h}\cos\phi\, d\phi \tag{3.60}$$

$$J_a(\varepsilon) = \frac{1}{2\pi}\int_{-\phi_l}^{\phi_l}\left[1-\frac{1}{2\varepsilon}(1-\cos\phi)\right]^{n_h} d\phi \tag{3.61}$$

Those integrals were introduced by Sjovall [13] and Table 3.1 gives their values for point and line contact as function of $F_r\tan\alpha\,/F_a$, assuming that the contact angle α is constant for all the loaded rolling elements.

$J_r(\varepsilon)$ and $J_a(\varepsilon)$ for Single-Row Bearings

	Point Contact			Line Contact		
ε	$\frac{F_r\tan\alpha}{F_a}$	$J_r(\varepsilon)$	$J_a(\varepsilon)$	$\frac{F_r\tan\alpha}{F_a}$	$J_r(\varepsilon)$	$J_a(\varepsilon)$
0	1	1/Z	1/Z	1	1/Z	1/Z
0.2	0.9318	0.1590	0.1707	0.9215	0.1737	0.1885
0.3	0.8964	0.1892	0.2110	0.8805	0.2055	0.2334
0.4	0.8601	0.2117	0.2462	0.8380	0.2286	0.2728
0.5	0.8225	0.2288	0.2782	0.7939	0.2453	0.3090
0.6	0.7835	0.2416	0.3084	0.7480	0.2568	0.3433
0.7	0.7427	0.2505	0.3374	0.6999	0.2636	0.3766
0.8	0.6995	0.2559	0.3658	0.6486	0.2658	0.4098
0.9	0.6529	0.2576	0.3945	0.5920	0.2628	0.4439
1	0.6000	0.2546	0.4244	0.5238	0.2523	0.4817
1.25	0.4338	0.2289	0.5044	0.3598	0.2078	0.5775
1.67	0.3088	0.1871	0.6060	0.2340	0.1589	0.6790
2.5	0.1850	0.1339	0.7240	0.1372	0.1075	0.7837
5	0.0831	0.0711	0.8558	0.0611	0.0544	0.8909
∞	0	0	1	0	0	1

Table 3.1: Values of eccentricity, thrust and radial integrals for different values of external loads [6]

If axial load is not present, the static model allows the analysis of the effect of clearance on the load distribution. For a rigidly supported bearing, the radial deflection at any rolling element angular position is given by:

$$\delta_\phi = \delta_r \cos\phi - \frac{1}{2} c_d \tag{3.62}$$

where δ_r is the ring radial shift, and c_d is the diametral clearance.

If the equation is rearranged in terms of maximum deformation:

$$\delta_\phi = \delta_{max}\left[1 - \frac{1}{2\varepsilon}(1 - \cos\phi)\right] \tag{3.63}$$

$$\varepsilon = \frac{1}{2}\left(1 - \frac{c_d}{2\delta_r}\right) \tag{3.64}$$

The angular extension of the load zone is determined by the diametral clearance:

$$\phi_l = \cos^{-1}\left(\frac{c_d}{2\delta_r}\right) \tag{3.65}$$

For static equilibrium, the applied radial load must equal the sum of the vertical components of the rolling element loads:

$$F_r = \sum_{\phi=-\phi_l}^{\phi=\phi_l} Q_\phi \cos\phi \tag{3.66}$$

or

$$F_r = Q_{max} \sum_{\phi=0}^{\phi=\pm\phi_l} \left[1 - \frac{1}{2\varepsilon}(1 - \cos\phi)\right]^{n_h} \cos\phi \tag{3.67}$$

This equation can also be rewritten in integral form:

$$F_r = \mathrm{Z}\, \mathrm{Q}_{max} \frac{1}{2\pi} \int_{-\phi_l}^{\phi_l} \left[1 - \frac{1}{2\varepsilon}(1 - \cos\phi)\right]^{n_h} \cos\phi \, d\phi \tag{3.68}$$

or

$$F_r = Z Q_{max} J_r(\varepsilon) \tag{3.69}$$

where the radial integral is defined as:

$$J_r(\varepsilon) = \frac{1}{2\pi} \int_{-\phi_l}^{\phi_l} \left[1 - \frac{1}{2\varepsilon}(1 - \cos\phi)\right]^{n_h} \cos\phi \, d\phi \tag{3.70}$$

In Table 3.2 the radial integral $J_r(\varepsilon)$ has been evaluated numerically for various values of ε.

Load Distribution Integral $J_r(\varepsilon)$

ε	Point Contact	Line Contact	ε	Point Contact	Line Contact
0	1/Z	1/Z	0.8	0.2559	0.2658
0.1	0.1156	0.1268	0.9	0.2576	0.2628
0.2	0.1590	0.1737	1.0	0.2546	0.2523
0.3	0.1892	0.2055	1.25	0.2289	0.2078
0.4	0.2117	0.2286	1.67	0.1871	0.1589
0.5	0.2288	0.2453	2.5	0.1339	0.1075
0.6	0.2416	0.2568	5.0	0.0711	0.0544
0.7	0.2505	0.2636	∞	0	0

Table 3.2: Values of radial integral for various values of ε

For a radial roller bearing with zero internal radial clearance and subjected to a simple radial load, it holds that

$$Q_{max} = \frac{4.08F_r}{Z\cos\alpha} \tag{3.71}$$

The effect of radial clearance on the load distribution among the rolling elements can be seen in Figure 3.8

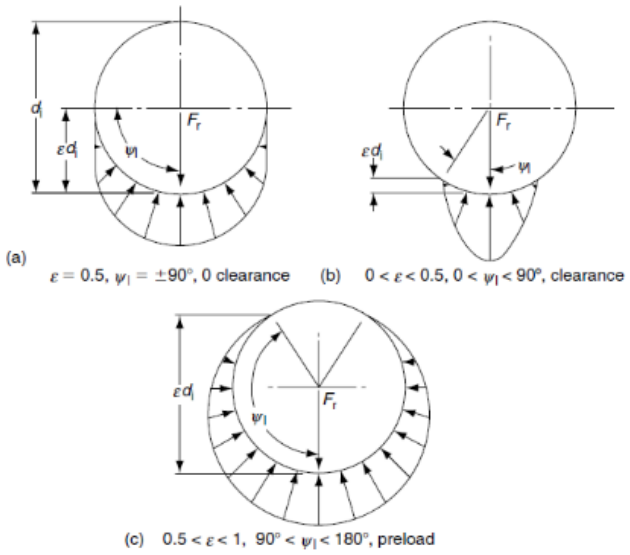

Figure 3.8: Rolling element load distribution for different amount of clearance [6]

3.3 KINEMATICS

In most applications of rolling bearings and particularly those operating at relatively slow shaft or outer-ring speeds, the velocities of all the internal element of bearing can be calculated with good accuracy using simple kinematical relationships and assuming that rolling elements roll on the raceways without sliding. [14]

3.3.1 Cage speed

Assuming that both inner and outer rings rotate in a bearing having common contact angle α (Figure 3.9), given the shaft rotational speed, its angular speed and so the one of the inner ring is [6]:

$$\omega_{IR} = \frac{2\pi n}{60} \tag{3.72}$$

Consequently:

$$v_{IR} = \frac{1}{2}\omega_{IR}(d_m - D_r \cos\alpha) = \frac{1}{2}\omega_{IR} d_m (1-\gamma) \tag{3.73}$$

$$v_{OR} = \frac{1}{2}\omega_{OR} d_m (1+\gamma) \tag{3.74}$$

where γ is defined in Equation 3.16.

If there is no slip at the rolling element-raceway contact, the cage velocity is the mean of the inner and outer raceways velocities, so:

$$v_m = \frac{1}{2}(v_{IR} + v_{OR}) = \frac{\pi d_m}{120}[n_{IR}(1-\gamma) + n_{OR}(1+\gamma)] \tag{3.75}$$

If the outer ring does not move, as in the analysed case, the nominal angular speed of the cage is:

$$\omega_m = \frac{\omega_{IR}}{2}(1-\gamma) \tag{3.76}$$

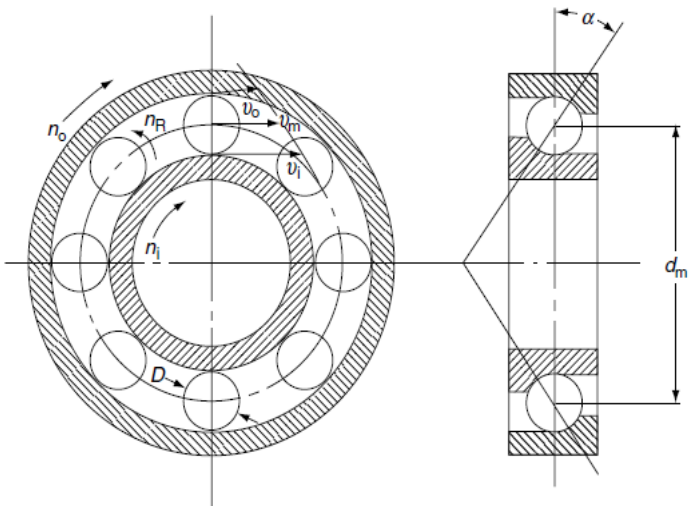

Figure 3.9: Rolling speeds and velocities in a bearing [6]

3.3.2 Rolling element speed

Assuming that there is no slipping between rolling element and raceways, the linear velocity of the rolling element is equal to that of the inner race at the contact point, thus [6]

$$\frac{1}{2}\omega_m d_m(1-\gamma) = \frac{1}{2}\omega_r D_r \tag{3.77}$$

Since the rotational speed of the cage relative to the inner race is

$$n_{m,IR} = n_m - n_{IR} \tag{3.78}$$

substituting the expression of n_m, the first equation becomes:

$$n_r = \frac{d_m}{2D_r}(1-\gamma)(1+\gamma)(n_{OR} - n_{IR}) \tag{3.79}$$

In the analyzed case only the inner ring rotates, so:

$$n_R = \frac{d_m n_{IR}}{2D_r}(1-\gamma^2) \tag{3.80}$$

3.4 EHL THEORY

Ball and roller bearings require fluid lubrication if they are to perform satisfactorily for long periods of time. The primary function of a lubricant is to prevent wear of the rolling and sliding contacts of a bearing under hydrodynamic, elastohydrodynamic (EHL), or boundary lubrication conditions; however, there are many other vital functions, such as: [6]

- minimizing the frictional power loss of the bearing
- acting as a heat transfer medium to remove heat from the bearing or to redistribute the heat energy within the bearing assembly to minimize differential thermal expansions
- protecting the precise surfaces of bearing components from corrosion
- removing wear debris from the roller contact paths
- minimizing the amount of extraneous dirt entering the roller contact paths
- providing a damping medium for cage dynamic motions

In the 1880s Reynolds established the fundamental theory, called Hydrodynamic lubrication theory (HL-theory) [15] [14], describing the lubricant flow between conformal surfaces. Those kinds of surfaces are those of journal bearings, in which contacting bodies have almost the same radii of curvature, the surface area, through which the external forces are transmitted, is large and consequently the pressure within the lubricant film is comparatively low. The corresponding basic differential equation, which relates the pressure in the lubricant film to its geometry and the velocities of the moving surfaces, is called Reynolds equation. [16]

In rolling element bearings, contact areas are non-conformal surfaces, since radii of curvature might differ a lot and only small areas of contact with extremely high pressures are developed. For that reason, it took a long time to recognise that the fluid film could separate the metallic surfaces.

The first breakthrough in this field of research occurred in 1949 when Grubin [17] managed to incorporate in his theory both the elastic deformation of the solids and the viscosity-pressure characteristics of the lubricant to analyse the lubricant flow between non-conformal surfaces. EHL-analysis was first performed for line-contact (two-dimensional) problems in isothermal conditions and without transient normal motion of the contacting bodies (steady-state).

As computer power increased and numerical solution techniques were improved, these assumptions could be removed step by step. The modern field of research is the transient EHL-problem in which the time-dependent Reynolds equation has to be solved. [10]

3.4.1 EHL system of equations

The elastohydrodynamic model is composed of the Reynolds equation, the lubricant film thickness equation and the equation of motion of EHL contact. Since the properties of the lubricant fluid change due to large contact forces, the viscosity and density relations as function of pressure are also added to the EHL model:

- *Reynolds equation*:

$$\frac{\partial}{\partial x}\left(\frac{\rho h^3}{\eta}\frac{\partial p}{\partial x}\right)+\frac{\partial}{\partial y}\left(\frac{\rho h^3}{\eta}\frac{\partial p}{\partial y}\right)=6u_s\frac{\partial(\rho h)}{\partial x}+12\frac{\partial(\rho h)}{\partial t} \tag{3.81}$$

$$p(x,y,t)\geq 0;\ \ \forall(x,y)\in\Omega \tag{3.82}$$

  where h is the lubricant film thickness, η is the absolute viscosity of the fluid, ρ is its density, u_s is the entrainment velocity

- *Film thickness equation*:

$$h(x,y,t)=-\delta(t)+\frac{x^2}{2R}+\frac{2}{\pi E'}\iint\frac{p(x',y',t)}{\sqrt{(y-y')^2+(x-x')^2}}dx'dy \tag{3.83}$$

  where δ is the mutual approach between both surfaces in contact, R is the equivalent radius of curvature, E' is the reduced modulus of elasticity defined in Equation 3.41

$$\frac{1}{R}=\frac{1}{R_1}+\frac{1}{R_2} \tag{3.84}$$

- *Viscosity-pressure relationship of Roelands*:

$$\eta(p)=\eta_0\exp\left\{\frac{\alpha_{PV}\,p_0}{z_{PV}}\left[-1+\left(1+\frac{p}{p_0}\right)^z\right]\right\} \tag{3.85}$$

  Where η_0 is the kinematic viscosity at atmospheric pressure p_0, α_{PV} is the pressure-viscosity coefficient, z_{PV} is the viscosity-pressure ratio and p is the manometric pressure in $[Pa]$.

- *Density-pressure relationship of Dowson and Higginson*:

$$\rho(p) = \rho_0 \left(\frac{5.9 * 10^8 + 1.34\, p}{5.9 * 10^8 + p} \right) \tag{3.86}$$

- *Equilibrium equation*:

$$f = \iint p(x, y, t) dx dy \tag{3.87}$$

- *Motion equation*:

$$m_r \ddot{\delta}(t) + \iint p(x, y, t) dx dy = f(t) \tag{3.88}$$

$$m_r = \pi R^2 L \rho_{steel} \tag{3.89}$$

Where $\ddot{\delta}$ is the acceleration between bodies in contact, m_r is the roller mass in the reduced contact, L is the effective length of the contact surface and ρ_{steel} is the material density of the contact bodies. [18] [19]

3.4.2 Steady-state EHL-contact and lubricant film thickness

Numerous publications, such as Harris [14], Hamrock [15] and Venner [20], present the theory of steady-state, isothermal EHL-contacts.

Figure 3.10 shows a qualitative representation of the lubricant film thickness and the pressure distribution profile within the EHL-contact between a rolling element and raceway. The contacting bodies are deformed due to the normal load Q and the viscous lubricant fluid is dragged into the high-pressure zone of the contact, adhering to the surface of the rolling bodies and separating the two mating surfaces. Since the lubricant viscosity strongly increases with increasing pressure, the film thickness is almost constant across the contact area and the minimum value h_{min} is found at the trailing edge of the contact.

The pressure distribution in EHL-contacts is similar to the elliptical Hertzian pressure distribution: from the inlet zone pressure increases, reaching the maximum value p_0 equal to the Hertzian pressure; in correspondence the minimum film thickness h_{min}, there is a spike in pressure distribution.

For high loads the pressure profile is nearly similar to the Hertzian one, whereas for high speeds and low loads the hydrodynamic effects become more important, since the pressure spike becomes bigger and shifts towards the centre of the contact area. [10]

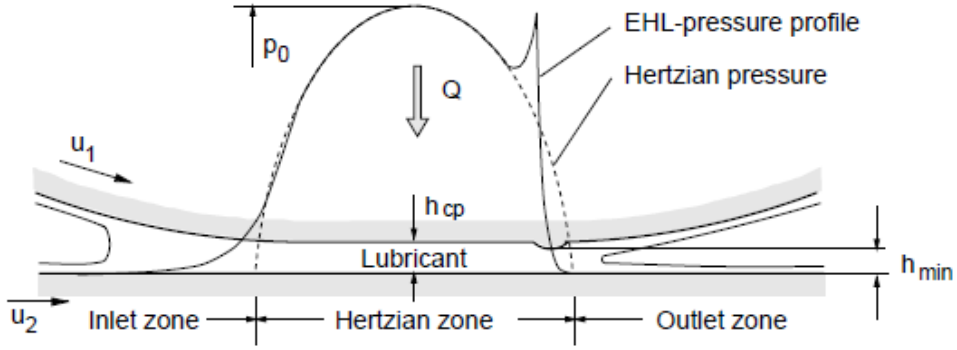

Figure 3.10: EHL-contact scheme showing a qualitative shape of lubricant film and pressure profile [10]

In order to obtain the oil film pressure profile and thickness in EHL lubrication, the Reynolds equation should be solved together with the surfaces deformations. When surfaces are rough, Reynolds equation should be modified once more since roughness changes the load balance: the external load is shared between the fluid and the asperities, which can deform elastically, elasto-plastically or plastically, while smooth surfaces deform only in an elastic way. [21]

Therefore, the total pressure is the sum of the hydrodynamic and asperity pressure:

$$p = p_{ehl} + p_a \tag{3.90}$$

In line contact EHL, the steady-state Reynolds equation formulated by Patir and Cheng [22], that includes the effects of roughness, is written as follows:

$$\frac{\partial}{\partial x}\left(\phi_x \frac{\rho h^3}{12\mu}\frac{\partial p}{\partial x}\right) = u\frac{\partial(\rho h_T)}{\partial x} \tag{3.91}$$

where h is the film thickness, μ is the fluid viscosity, ρ is its density, u is the rolling speed, ϕ_x is the pressure flow factor in x direction and h_T is the average gap between the two mating surfaces, computed for a Gaussian distribution of surface heights.

Reynolds equation can be solved through numerical methods, but Masjedi and Khonsari [23] [24], derived curve-fitting equations for the minimum and central film thickness, after having simulated hundreds of cases with a wide range of numerical values for the main parameters.

The central film thickness equation is:

$$H_c = \frac{h_{ehl,c}}{R_x} = 2.691\, W^{-0.135} U^{0.705} G^{0.556}\, (1 + \bar{\sigma}^{1.222} V^{0.223} U^{-0.748} G^{-0.842}) \tag{3.92}$$

While the minimum film thickness can be computed through:

$$H_{min} = \frac{h_{ehl,min}}{R_x} = 1.652\, W^{-0.077} U^{0.716} G^{0.695}\, (1 + 0.026\, \bar{\sigma}^{1.120} V^{0.185} W^{-0.312} G^{-0.977}) \tag{3.93}$$

The dimensionless parameters involved in those equations are defined as follows:

$$W = \frac{q}{E'R_x} \tag{3.94}$$

$$U = \frac{\mu_0 u}{2E'R} \tag{3.95}$$

$$G = \alpha_{PV} E' \tag{3.96}$$

$$\bar{\sigma} = \frac{\sigma}{R_x} \tag{3.97}$$

$$V = \frac{hd}{E'} \tag{3.98}$$

Where w is the load per contact length, E' is the effective modulus of elasticity defined in Equation 3.41, R_x is the equivalent contact radius, μ_0 is the lubricant viscosity at atmospheric pressure, u is the mean rolling speed, α is the fluid pressure-viscosity coefficient, σ is the standard deviation of surface height, hd is the Vickers hardness of the softest material.

$$q = \frac{Q}{L_r} \tag{3.99}$$

$$\frac{1}{R_x} = \frac{1}{r_{1x}} + \frac{1}{r_{2x}} \tag{3.100}$$

$$u = \frac{u_{roller} + u_{raceway}}{2} \tag{3.101}$$

$$\sigma = \sqrt{\sigma_{roller}^2 + \sigma_{raceway}^2} \tag{3.102}$$

The mean velocities for the inner and outer raceways are:

$$u_{IR} = \frac{d_m}{2}[(1-\gamma)(\omega - \omega_m) + \gamma\omega_r) \tag{3.103}$$

$$u_{OR} = \frac{d_m}{2}[(1+\gamma)\omega_m + \gamma\omega_r) \tag{3.104}$$

The empirical formula for the film-thickness calculation is valid for the dimensionless parameters in the ranges presented in Table 3.3: Dimensionless parameters ranges Table 3.3.

PARAMETER	W	U	G	$\bar{\sigma}$	$\bar{\sigma}$
Min	$2 * 10^{-5}$	$1 * 10^{-12}$	2500	0	0.005
Max	$5 * 10^{-4}$	$1 * 10^{-10}$	7500	$5 * 10^{-5}$	0.03

Table 3.3: Dimensionless parameters ranges

3.5 DAMPING

Fluid film damping, coming from the variation of lubricant film thickness, and structural film damping, which arises from hysteresis losses within the deformed material, are associated with the variation of the elastic displacement δ of the contacting bodies, so the contact damping force follows the law:

$$F_D = c\,\dot{\delta} \tag{3.105}$$

where the damping coefficient c includes all the viscous loss effects.

The evaluation of the damping coefficient has been carried by many researchers, but it seems to be a very difficult field. The most used relation is the one coming form Kramer [25], that relates the bearing damping coefficient to its linearized stiffness, obtained in static or quasi-static conditions:

$$c = (0.25 - 2.5)\ 10^{-5}\ k \tag{3.106}$$

The main sources of damping in rolling element bearings are summarized in Figure 3.11. Since damping effects are important only in case of notable relative motions only those happening in the Hertzian contact areas between rolling elements and raceways, housing and bearing, shaft and bearing are taken into account. The most important damping sources in a bearing according to Zeillinger [26] are:

- material damping due to Hertzian deformation of the rolling bodies
- oil-fil damping due to squeeze effects within the inlet zone of the EHL-contact

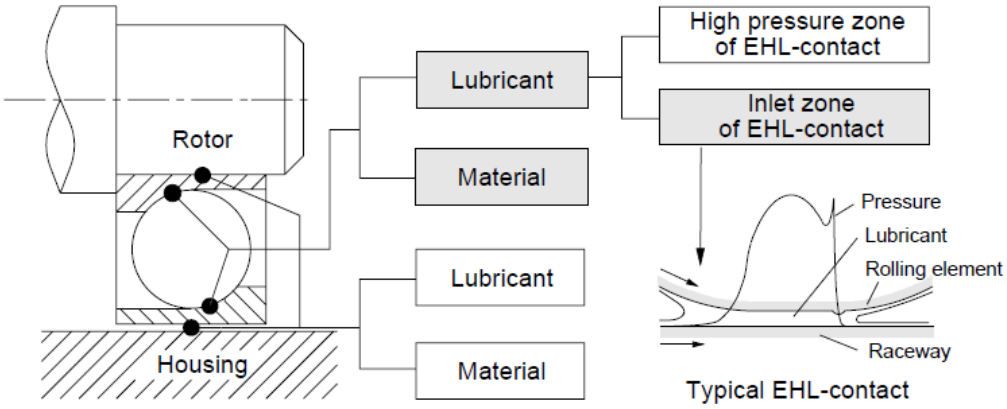

Figure 3.11: Damping sources in a rolling bearing joint [10]

3.5.1 Dry Hertzian contact damping

When an elastic structure is deformed, the major portion of the supplied energy is stored in the structure as potential energy, while the rest is absorbed and transformed into thermal energy due to dissipative mechanisms, such as thermo-elasticity, magneto-elastic effects, relaxation processes in crystal structures and interactions between molecular forces. Regarding metal materials, the friction between grain boundaries is the main dissipative mechanism.

If an elastic body is subjected to cyclic stress, the energy loss in the material causes a hysteresis loop in the stress-strain diagram and the area within it is proportional to the absorbed damping energy per load cycle. [27]

In common metals energy loss is low, so the total strain energy stored in deformed bodies is barely equal to the energy needed for the deformation and the hysteresis loop is slim. Being E_s the total strain energy stored in the deformed structure of volume V and E_D the energy dissipated per load cycle, the loss factor is defined as (Figure 3.12):

$$\psi_L = \frac{E_D}{2\pi E_S} \tag{3.107}$$

Through the analysis of experimentally determined values, the dissipated energy per load cycle can be written in the form:

$$E_D = \psi_L \pi k X_a^2 \tag{3.108}$$

Where k is the stiffness of the structure, and X_a is the maximum deflection amplitude.

According to the study of Lazan [28], the loss factor for common metals is lower than 1%, while for ball bearings is between $0.02 - 0.07\%$. Experimental investigations on single ball-to-raceways contact were conducted by Zeillinger [26], who reported $\psi_L = 1.5 - 3\%$, Krempf and Sabot ($\psi_L = 1.5\%$). In the work written by Dietl [10] ψ_L values between $1.7 - 2\%$ were found for dry rolling bearing contacts.

The energy dissipated through structural damping is independent of the frequency of deformation, while viscous damping forces lead to an enlargement of the hysteresis loop with increasing frequency. Since a viscous damping model is easier to handle in theoretical analysis, the loss factor is transformed into an equivalent viscous damping coefficient c_{ψ_L}. For harmonic vibrations of frequency f:

$$c_{\psi_L} = \psi_L \frac{k}{2\pi f} \tag{3.109}$$

The natural frequency of a cylinder with mass per unit width $\hat{m}$ and stiffness per unit length $\hat{k}_h$, oscillating between two raceways, is:

$$f = \frac{1}{2\pi}\sqrt{\frac{2\hat{k}_h}{\hat{m}}} \tag{3.110}$$

$$\hat{k}_h = \frac{\pi E'}{2[2\ln 2 - \ln(\gamma q)} \tag{3.111}$$

$$\hat{m} = \frac{m}{L_r} = \frac{E'}{4\pi[2\ln 2 - \ln(\gamma_h q)} \tag{3.112}$$

$$\gamma_h = \frac{8e^{\frac{1}{2(1-\nu)}}}{\pi E' L_r} \tag{3.113}$$

The equivalent material damping coefficient of a single Hertzian contact in a rolling element bearing can be written as:

$$c_{h,i} = \psi_L \frac{k_{h,i}}{2\pi f} \tag{3.114}$$

If ψ_L is assumed to be equal for all rolling elements, the material damping matrix can be obtained directly from the bearing stiffness matrix:

$$\boldsymbol{C}_{B,h} = \frac{\psi_L}{2\pi f} \boldsymbol{K}_B \tag{3.115}$$

Experimental validations of the viscous damping model for the dry Hertzian contact were successfully conducted by Hunt and Crossley [29] and by Kremps and Sabot [30].

Summing up, the loss factor approach can be used to make quick estimations of the minimum damping capability of a rolling element bearing.

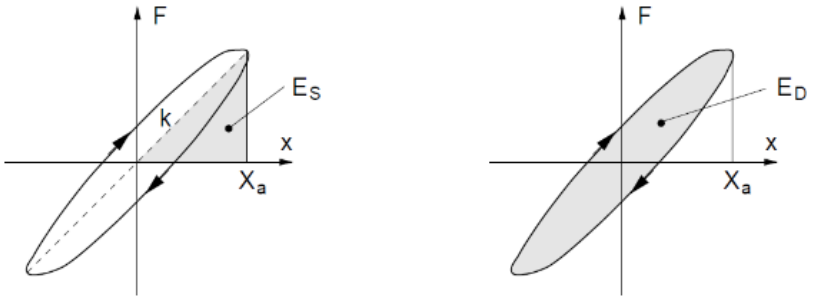

Figure 3.12: Force-deflection diagram with hysteresis loop caused by internal damping; E_s is the maximum strain energy, E_D is the dissipated energy [10]

3.5.2 Lubricant film damping

The lubricant film damping can be computed through numerical methods, but here a curve-fitted formula taken from the work of Dietl [10] is presented. Oil-film damping coefficient were obtained solving the complete transient EHL line-contact problem under isothermal conditions, using a code developed by Van Nijen [31].

The equivalent oil-film damping coefficient of a single line-contact can be estimated through the following empirical law:

$$c_{ehl} = 0.1963\, R_x^{0.781}\, L_r^{0.769}\, E'^{1.069}\, \mu_0^{0.531}\, \alpha_{PV}^{0.424}\, q^{-0.136}\, u_s^{-0.434}\, F_{inlet}^{-0.563} \tag{3.116}$$

where R_x is the equivalent radius of curvature of Equation 3.99, L_r is the roller length, E' is the equivalent modulus of elasticity defined in Equation 3.41, μ_0 is the lubricant viscosity at ambient pressure p_0, α_{PV} is the pressure-viscosity coefficient, q is the rolling element load per unit length (Equation 3.98), u_s is the sum of surface velocities, F_{inlet} is the inlet-zone length factor (Figure 3.13).

$$u_s = u_1 + u_2 \tag{3.117}$$

$$u_1 = u_2 = \frac{\pi D_r}{60}\left(\frac{d_m}{D_r} - \frac{D_r}{d_m}\cos^2\alpha\right)\frac{n_{IR} - n_{OR}}{2} \tag{3.118}$$

$$F_{inlet} = \frac{l_{inlet}}{b} \tag{3.119}$$

$$b = \sqrt{\frac{8QR_x}{\pi E' L_r}} \tag{3.120}$$

l_{inlet} is the length of the lubricant inlet zone and b is the semi-minor axis of the line-contact zone. For $F_{inlet} > 4$ the equivalent damping coefficient does not change and $F_{inlet} = 4$ can be considered when in starved lubrication conditions. Higher values of F_{inlet} give lower damping coefficients.

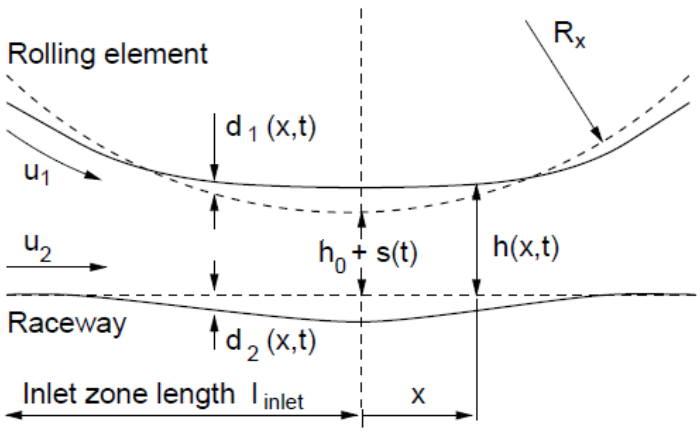

Figure 3.13: EHL-contact profile and parameters

Dietl's empirical formula for the damping factor computation is valid for the following ranges of input parameters (Table 3.4):

SYMBOL	UNIT	DESIGNATION	RANGE
R_x	m	equivalent radius of curvature	$5 - 20 * 10^{-3}$
L_r	m	contact length of line-contact	$1 - 16 * 10^{-3}$
E'	Pa	equivalent modulus of elasticity	$1.95 - 2.55 * 10^{11}$
μ_0	Pa s	lubricant viscosity at ambient pressure	$0.01 - 0.28$
α	Pa^{-1}	Pressure-viscosity coefficient	$1.8 - 2.3 * 10^{-8}$
q	N/m	rolling element load per unit length	$70000 - 260000$
p_0	Pa	maximum Hertzian pressure	$600 - 1000 * 10^6$
u_s	m/s	sum of surface velocities	$1.8 - 11.0$
F_{inlet}		inlet zone length factor	$1.5 - 4.0$

Table 3.4: Input parameters' validity ranges

3.6 EQUIVALENT STIFFNESS AND DAMPING OF A SINGLE LUBRICATED ROLLING CONTACT

As previously discussed, the damping capability of a roller bearing mainly comes from the dissipative processes within the contact zones between rolling elements and raceways. due to viscous losses within the lubricant film, material damping losses and dry friction losses.

Figure 3.14 shows the scheme of a single lubricated rolling contact and the corresponding, dynamically equivalent spring-damper model. Numerical calculations of EHL non-linear equations show that for sufficiently high contact preloads Q and small amplitudes of lateral vibrations of the rolling bodies, the EHL-contact has got approximately the characteristics of a linear, viscoelastic system, that can be linearised for a certain reference load and modelled by a combination of linear springs and viscous dampers.

The hertzian stiffness is represented by the spring coefficient k_h, while the lubricant film stiffness k_{ehl} can be set to infinite, since it is some orders of magnitude higher than the hertzian one. The energy dissipation within the deformed material and the frictional losses in case of metallic contact of rolling bodies are represented by the damping coefficient c_h and c_{ehl} is the damping capability of the lubricant film.

Since the film stiffness is very high in the contact zone, the damping capability of the lubricant is caused by viscous losses within the inlet zone of the EHL contact; for that reason, the dashpot c_{ehl} is modelled in parallel with the other lumped parameters. [10]

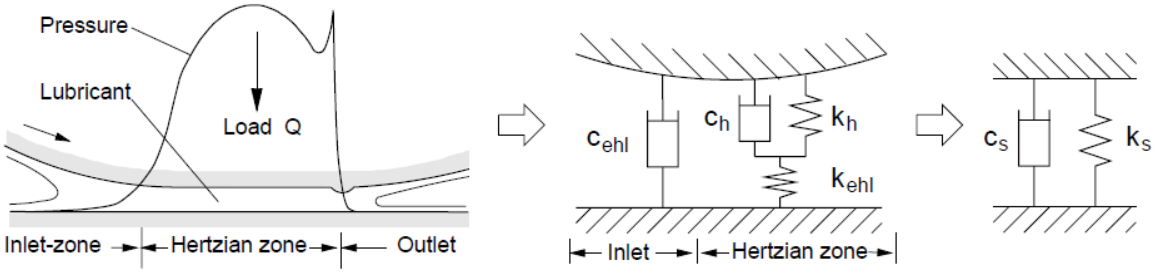

Figure 3.14: Lubricated rolling contact (left), spring-damper model of the contact (centre), equivalent Kelvin-Voigt element (right) [10]

3.6.1 Equivalent stiffness and damping of a single rolling element

In common rolling bearing applications, the vibration frequency of the inner ring with respect to the outer one is much lower than the natural frequency of a single rolling element oscillating between the raceways. Therefore, the mass of the rolling elements can be neglected and the scheme of the system is shown in Figure 3.15. [32]

c_s and k_s for both the inner and outer ring can be evaluated through:

$$k_s = k_h \tag{3.121}$$

$$c_s = c_h + c_{ehl} \tag{3.122}$$

So:

$$\frac{1}{c_j} = \frac{1}{c_{sj,IR}} + \frac{1}{c_{sj,OR}} \tag{3.123}$$

$$\frac{1}{k_j} = \frac{1}{k_{sj,IR}} + \frac{1}{k_{sj,OR}} \tag{3.124}$$

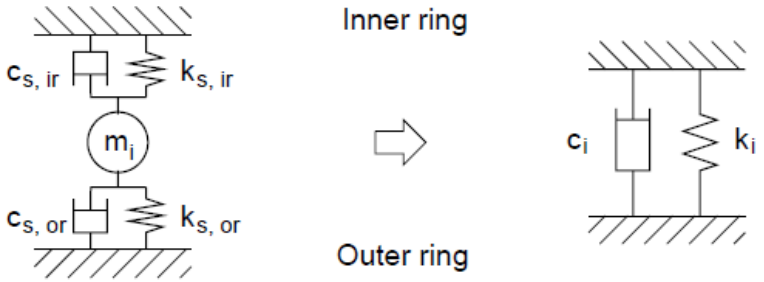

Figure 3.15: Spring-damper model for a single rolling element and equivalent Kelvin-Voigt element [10]

4 SPHERICAL ROLLER BEARING DYNAMIC MODEL

The first step in order to setup an AI algorithm able to predict the health status of a spherical rolling element bearing is the elaboration of a suitable bearing mathematical model that can generate a wide database of vibration data to be used in the training phase. Such a model should be able to simulate the behavior of different types of bearings, in various operating conditions and with different kinds and dimensions of defects.

In this chapter a brief review of the main types of analytical models, used to simulate bearing performances, and the development of the full 4 DoF multi-body non-linear dynamic model of the spherical roller bearing are described.

4.1 ANALYTICAL MODELS

The simulation of rolling element bearings performance through an analytical formulation can be done following two different approaches: the quasi-static model is based on static force and moment equilibrium equations, while dynamic one involves the integration of the differential equations of motion.

4.1.1 Quasi-static model

In this method, the force and moment equilibrium equations are written for each bearing element for the prescribed loads, including the effects of centrifugal forces and gyroscopic moments, together with the external forces and moments.

The rolling element angular velocities are computed from the prescribed angular velocities of the races, assuming pure rolling condition and the rotation of each roller about its axis of symmetry. In angular contact ball bearings, kinematic relationships are more complicated due to the possible spin velocity, that is an angular velocity along an axis normal to the contact plane. For that reason, a “race control” hypothesis, which restricts the ball to spin either on the outer or on the inner race, depending on the race that gives the smaller spin moment for the known contact loads, is added. However, it has been shown that in well-lubricated bearings, the race control hypothesis is not valid and rolling elements spin on both the raceways, so the incorporation of realistic lubricant behavior in this kind of model is impossible.

Other limitations of the quasi-static model are the simulation of the cage interaction with rolling elements, that implies short duration collisions (highly dynamic), roller skidding and skewing and the influence of time-varying loads.

Beside those limitations, quasi-static models are very useful for the design of rolling element bearings, because they provide a realistic load distribution on the rolling elements and well accepted prediction of fatigue life and bearing stiffness. Moreover, the static equilibrium solution provides favourable initial conditions for the integration of the differential equations of motion in dynamic simulations.

The most important features of the quasi-static model can be summarized as follows: [5]

- the solution to a set of nonlinear algebraic equations of static equilibrium is obtained
- kinematic constraints are employed for computing angular velocities of rolling elements
- it is very effective in the estimation of the overall load distribution, bearing fatigue and stiffness
- since a single solution provides all the performance parameters, the model requires only moderate computational effort
- Convergence problems can be found when simulating the realistic lubricant traction behavior of a lubricant
- none of the bearing elements accelerate and all velocities are assumed to be constant
- the cage motion may not be considered
- roller instabilities, such as roller skidding and skewing, time-varying loads and race speeds cannot be treated

4.1.2 Dynamic model

In the dynamic model the equilibrium equations of the quasi-static model are replaced by the differential equation of motion of each bearing element and the applied forces and moments are computed as a consequence of the various interaction between the elements. Given the initial conditions, the differential equations are integrated numerically.

The dynamic model provides a real time simulation of the performance of the bearing assembly and eliminates most of the problems of the quasi-static model: for example, angular velocities are now computed by integrating the corresponding accelerations, so no kinematics constraints are necessary. The model also allows the treatment of all the internal transient forces, resulting from the interaction between the different elements of the bearing, and the lubrication effects. It also provides an improved steady-state solution, that can be presented as a time-cycle solution rather than a one-point solution of the equilibrium problem.

Despite the increased computational effort for the numerical integration, the dynamic model provides an overall simplification of the design procedure and costs, because the real-time simulation helps in the replacement of experimental investigations.

So, the main features can be summarized as follows: [5]

- the model consists of the integration of the differential equations of motion of the bearing elements, providing a real-time simulation of the bearing performance
- no kinematic constraints are necessary
- the lubricant behavior can be modelled
- the accelerations of any bearing element are determined by applied forces and moments, which are computed from the interactions between the elements themselves
- all external interactions may vary with time
- instabilities can be simulated

4.1.3 Multi-body dynamic model

The nonlinear multi-body dynamic analytical model of rolling element bearings and associated systems are lumped parameter (masses, springs and dampers) models and generally consider outer and inner rings as rigid masses and the rolling elements-to-raceways contact interfaces as nonlinear springs. They predict the vibration response of bearings, bearing-pedestal and rotor-bearing system due to the presence of localized (point, circular, elliptical and rectangular spalls) and extended defects.

Dynamics of rolling element bearings are addressed to be highly nonlinear and time-variant. Non-linearity comes from the nonlinear effects due to the Hertzian force-deformation relationship, the varying stiffness resulting from load transmission through a finite number of rolling elements, the presence of radial clearance in the rolling element-raceway contacts and the effect of lubricant film. Time-variant characteristics are the result of the orbital motion of rolling elements: varying compliance vibration is due to the bearing rotation movement itself, that leads to variations in its stiffness.

In order to simplify the analysis, the majority of the multi-body models uses the following assumptions: [33]

- the outer and inner rings are rigidly connected to the housing and the shaft respectively
- the rolling elements are excluded or considered massless
- the inertial and centrifugal effects on rolling elements are neglected
- the slippage of the rolling elements is ignored
- the EHL fluid film in rolling contacts is not considered
- the bearing stiffness is considered to be linear

Prior to investigating the vibration response of rolling element bearings due to the presence of defects, the research was primarily focused on understanding the characteristics of the vibration response of non-defecting ones. The first systematic investigations were conducted by Perret and Meldau in the early 1950s and concluded that rolling element bearings generate cyclic vibrations even in absence of manufacturing or geometrical imperfections, such as surface roughness, waviness, misaligned raceways, off-sized rolling elements, and out-of-round components; those vibrations are called variable compliance vibrations. Later, several researchers reported on the development of analytical models to predict the vibration response of rolling element bearings due to the presence of distributed defects. The first nonlinear multibody dynamic model was reported in 2002 by Feng et al [34]., that was based on the model developed by Fukata et al. [35], which describes the vibration response of an ideal ball bearing using a 2-DoF model.

Since that time, many non-linear multi-body dynamic models have been presented, each one with different features and levels of complexity, because the simulated vibration time-traces can easily highlight the defect-related frequency components and corresponding sidebands. The most important problem linked to those models is the amplitude mismatch between modelled and measured vibration frequencies, so in many researchers corrected the predicted amplitudes relying on experimental results.

4.2 4 DoF NON-LINEAR MULTI-BODY DYNAMIC MODEL

Although spherical roller bearings are extensively applied thanks to their high load-supporting capabilities, there is a lack of models of this kind of bearings; therefore, this work is based on ball or roller bearings models, that can easily be adapted to spherical roller bearings, assuming some simplifications.

The first major works on rolling element bearings dynamic modelling were performed by Lundberg and Palmgren [36] and Harris [14], who introduced the concept of nonlinear stiffness, but did not exploit the total non-linearity and time-varying characteristics of bearings. The first complete dynamic model for angular contact ball bearings was developed by Gupta [37], who introduced and solved the generalised differential equations of motion of the complete system. Tiwari et al. [38] [39] studied the effect of clearance on the dynamic response of a rigid rotor and reported that as clearance decreases, the bearing stiffness and its non-linear characteristics increase.

Elastohydrodynamic lubrication theory (EHL) was introduced in modelling in 1960s and Wijnant et al. [40] reported the studies and the computational model used to compute the effects of EHL influence on bearing dynamics: the interaction between bearing elements was described through non-linear spring-damper models, based on numerical solution of the EHL problem.

One of the most comprehensive models, even if it does not include the effects of the EHL theory, was presented by Fukata et al. [35] and included all the non-linear and time-variant characteristics of rolling element bearings. This was a 2 DoF model, which is based on the Hertzian load-displacement relationship. This model was then further developed by Feng et al. [34], including the effects of cage and rolling element slipping and of defects on raceways.

A more recent model is addressed to Sopanen and Mikkola [41] [42] and includes the effects of different geometrical faults and extended defects, such as surface roughness, waviness and localised defect on raceways and rolling elements, non-linear Hertzian contact deformation and the effects of EH fluid film, using a 6-DoF model for deep-groove ball bearings. They showed that radial clearance has a significant effect on natural frequencies and vibration response of the whole system, variable compliance vibration is generated even if the bearing is geometrically perfect and localised defects generate vibration at bearing defect frequencies.

Basing on the work by Feng et al., Sawalhi and Randall [43] [44] proposed a 5-DoF model, adding one degree of freedom through a mass-spring-damper system resonator to simulate the bearing high resonant mode and introducing a new defect profile that reduced the vibration amplitude peak to obtain results similar to experimental data. Petersen [45] [46] used two resonators, one along x and one along y-axis, and developed his model for two-rows spherical roller bearings, including rolling element slippage, EHL theory effect for lubricant damping, extended and localized defects effects.

Working on the above-mentioned model, Mohazen Ahmadi [47] developed a dynamic model of deep groove ball-bearings considering the finite size and inertia effects of rolling elements, effects of lubricant and extended defects. Recent models of ball bearing, including more or less the same approach to simulate the non-linear and time-variant behavior of rolling element bearings are those of Singh et al. [2], Shah and Patel [48], Liu and Shao [33].

The model from Yan et al. [49] [50] [51] is a 5 DoF nonlinear dynamic model for ball bearings that takes into account the effects of lubricant film damping and height: a new computational fluid dynamics method for modeling elastohydrodynamic contacts was proposed, in which fluid-solid interaction and elastic deformation of the solid as well as thermal effects were taken into account. This work presents a non-linear multi-body dynamic model that can be used to predict the time-domain vibration response of a radially loaded two-rows spherical roller element bearing due to a localized defect on a raceway. The described model is based on the ones of Petersen [45] [46] and Yan [49], that used the concept of variable compliance to describe the source of vibration in loaded rolling element bearings.

In the first formulation of the model presented in this work the high frequency resonator was not added, but it is one of the topics for future developments as well as the inclusion of the two rows of rolling elements. The defect profile model comes from Wen's dynamic model [52], that simulates the effects of a defected raceway with a change in the rolling element displacement.

The aforementioned models account for the effects of lubricating fluid, but not rigorously; in this research the fluid thickness and its damping coefficient are computed real-time at every instant of the simulation and for each rolling element, using empirical curve-fitted formulas coming from the EHL-theory, presented in the works of Masjedi [23] [24] and Dietl [10].

The non-linear multi-body dynamic model presented in this work has got 4 DoF (Figure 4.1), two for the inner (x_{IR} and y_{IR} displacements) and two for the outer ring (x_{OR} and y_{OR} displacements), can simulate the vibration of the bearing for average radial loads applied to the outer raceway and rotational velocities and with the presence of localized or extended defects on raceways. Rolling elements are modelled as massless bodies (since it has been proven that their effect is minimal [53]) with time-varying non-linear stiffness, while raceways are massive bodies with intrinsic stiffness and damping capabilities: the inner raceway has got mass m_{IR}, stiffness k_{IR} and damping c_{IR}, while the outer raceway has got mass m_{OR}, stiffness k_{OR} and damping c_{OR}. The model allows the simulation of all those bearing, whose internal contacts between their elements are assumed to be line-contacts, since it is really versatile and its input data can be easily changed.

The lubricant is assumed to be a Newtonian incompressible fluid, whose viscosity depends on temperature, working under isothermal conditions. The EHL effects also account for the bearing elements roughness and hardness.

The output data of the model include accelerations, velocities and positions relative to the centre of the bearing of the two raceways, rolling elements positions, Hertzian and damping loads, that allows the analysis of the load distribution and its comparation with static models if the rotating speed of the shaft is set to zero (Figure 4.2).

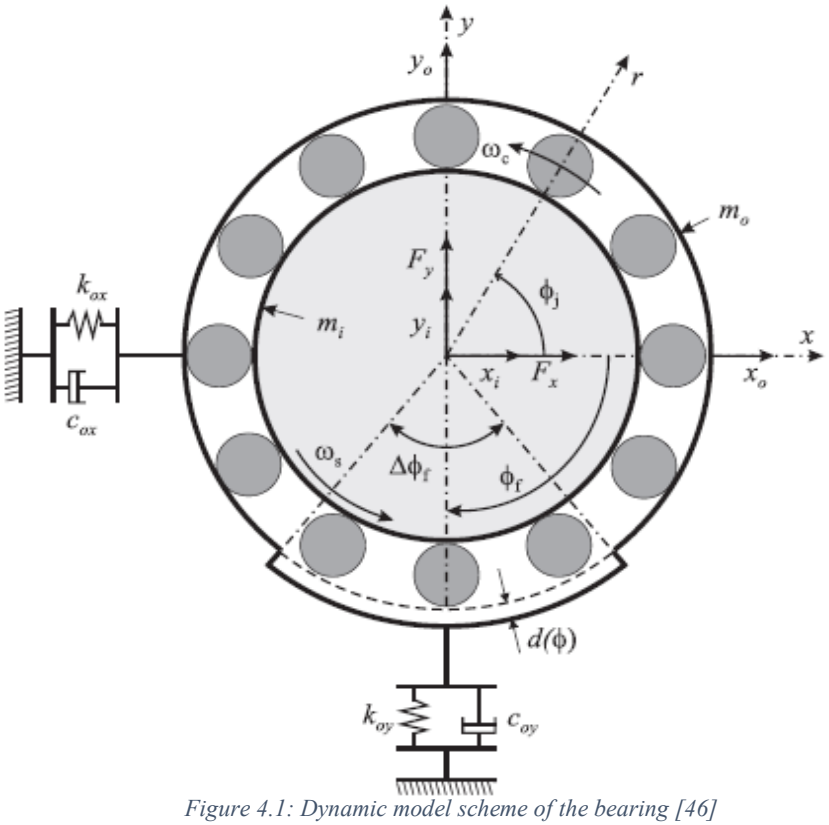

Figure 4.1: Dynamic model scheme of the bearing [46]

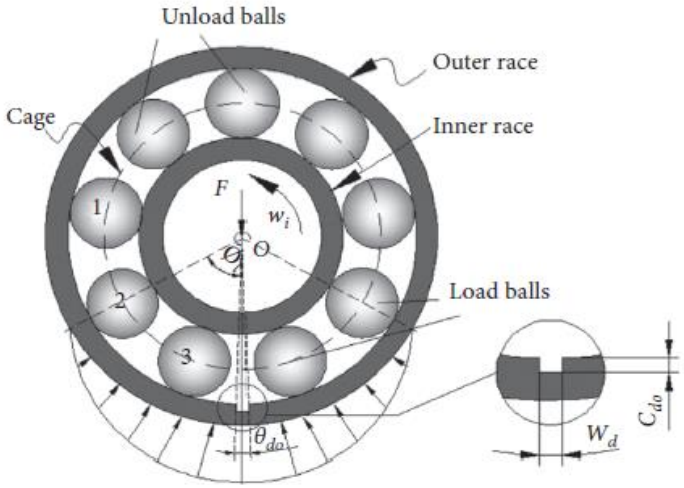

Figure 4.2: Schematic of bearing in load and unload zone

4.2.1 Assumptions and simplifications

The dynamic model is based on the Hertzian and EHL theories, considers the time-varying displacement excitation, the effects of radial clearance, lubricant oil film and the following assumptions:

- only one of the two rows of rolling elements is considered
- the external load is equally divided between the two rows of rolling elements
- the outer ring is fixed, while the inner rotates with the shaft
- the cage movement is based on the bearing geometric dimensions; therefore, no slipping occurs between the components of the bearing itself
- all the rollers are equally distributed around the inner race, move around the raceways with equal velocity and the centrifugal forces are neglected
- rollers are considered massless
- only nonlinear Hertzian contacts occur between rolling elements and raceways
- Hertzian contacts are assumed to be line-contacts
- all translational movements are limited to the $x - y$ plane and the rolling elements rotate around z-axis
- the lubricant oil is a Newtonian incompressible fluid, whose viscosity is constant along the thickness of the film
- the bearing is assumed to operate under isothermal working conditions
- the variation of lubricant oil film in the vertical direction is ignored
- the contact joint between the shaft and the inner ring is considered infinitely rigid and without damping capabilities
- the defect covers all the raceway width

4.3 DYNAMIC MODEL DEVELOPMENT

The modelling of the vibration response of the rolling element bearing with defects is obtained through a multi-body nonlinear dynamic model, based on the one developed by Sawalhi and Randall, which in includes 4 DoF: 2 translational DoF for the inner raceway and 2 for the outer one, that have respectively mass m_{IR} and m_{OR}. The spring and damper constant k_{IR}, k_{OR}, c_{IR} and c_{OR} represent the stiffness and damping of the bearing support structure (Figure 4.1).
The Hertzian contacts between the rolling elements and raceways are modelled by time-varying nonlinear contact springs and dampers.

4.3.1 Rolling element angular position

The position $\phi_j(t)$ of each rolling element is defined as:

$$\phi_j(t) = \phi_m(t) + \frac{2\pi(j-1)}{Z} + \phi_0, \quad j = 1, \dots, Z \tag{4.1}$$

where Z is the number of rolling elements, ϕ_m is the current cage angular position and ϕ_0 is the initial cage angular position.

$$\phi_m(t) = \omega_m t \tag{4.2}$$

where

$$\omega_m = \frac{\omega_{IR}}{2}(1-\gamma) \tag{4.3}$$

D_r is the rolling elements diameter, α is the contact angle and d_m is the pitch diameter.

If the slippage of rolling elements is also taken into account:

$$\phi_m(t) = \omega_m t + v_{rand}(t) \tag{4.4}$$

where $v_{rand}(t)$ is a random process uniformly distributed in the range $[-\phi_{slip}.\phi_{slip}]$. Typical maximum values for the phase variation ϕ_{slip} are in the order of $0.01 - 0.02$ rad

4.3.2 Hertzian contact model

The contact deformation $\delta_j(t)$ for each rolling element is a function of the relative displacement $\delta_x(t)$ and $\delta_y(t)$ of the inner and outer raceways, the position $\phi_j(t)$ of the rolling element itself, the defect profile $H_d\left(\phi_j(t)\right)$ at the rolling element position, which will be described in Section 4.3.6, the radial clearance c_d and the film thickness $h_{ehl,j}$ at rolling each element location: [49]

$$\delta_j(t) = \delta_x(t)\cos\phi_j(t) + \delta_y(t)\sin\phi_j(t) - h_{ehl,j} - c_d - H_d\left(\phi_j(t)\right) \tag{4.5}$$

$$\delta_x(t) = x_{IR}(t) - x_{OR}(t) \tag{4.6}$$

$$\delta_y(t) = y_{IR}(t) - y_{OR}(t) \tag{4.7}$$

The lubricant film thickness can be evaluated through the empirical formula illustrated in Section 3.4.2:

$$h_{ehl,j} = \frac{h_{min}}{R_x} = 2.691\, W_j^{-0.135} U^{0.705} G^{0.556}\,(1 + \bar{\sigma}^{1.222} V^{0.223} U^{-0.748} G^{-0.842}) \tag{4.8}$$

In this formula the load parameter W_j is computed through the current load on each element Q_j.

The Hertzian contact force $f_{h,j}(t)$ associated to the deformation $\delta_j(t)$ exists only for positive values of deformation and can be written as

$$f_{h,j}(t) = K_l \delta_j(t)^{n_h} \zeta_j(t) \qquad \zeta_j(t) = \begin{cases} 1 & if\ \delta_j(t) > 0 \\ 0 & if\ \delta_j(t) \le 0 \end{cases} \tag{4.9}$$

K_l depends on the curvature and material properties of the surfaces in contact and $n = \frac{10}{9}$ for roller bearings. For line-contact, as discussed in section 3.2.3, the stiffness can be evaluated through:

$$K_l = \frac{1}{3.81^{\frac{10}{9}}\left[\frac{1}{\pi}\left(\frac{1-\nu_{roller}^2}{E_{roller}} + \frac{1-\nu_{raceway}^2}{E_{raceway}}\right)\right]} L_r^{\frac{8}{9}} \tag{4.10}$$

where L_r is the rolling element length, E and ν are the Young modulus and the Poisson coefficient of the rolling element and raceway.

The total radial contact forces acting on the inner and outer raceways and computed over all rolling elements on both rows are:

$$\begin{bmatrix} f_{h,x}(t) \\ f_{h,y}(t) \end{bmatrix} = \sum_{j=1}^{Z} K_{in-ou}\delta_j(t)^{n_h}\zeta_j(t) \begin{bmatrix} \cos\phi_j(t) \\ \sin\phi_j(t) \end{bmatrix} \tag{4.11}$$

or

$$\begin{bmatrix} f_{h,x} \\ f_{h,y} \end{bmatrix} = \sum_{i=1}^{Z} K_{in-ou}\,\zeta_j \left[(x_{IR} - x_{OR})\sin\phi_j + (y_{IR} - y_{OR})\cos\phi_j - H_d - c - h_{EHL}\right]^{1.11} \begin{bmatrix} \cos\theta_j \\ \sin\theta_j \end{bmatrix} \tag{4.12}$$

where:

$$K_{in-ou} = \left(\frac{1}{\left(K_{l,IR}^{-n_h} + K_{l,OR}^{-n_h}\right)}\right)^{n_h} \tag{4.13}$$

is the resultant stiffness from the contact between each rolling element and both the raceways, since the springs K_{IR} and K_{OR}, that model the contact stiffness, are in series.

Taking into account the rolling elements contact angle:

$$\begin{bmatrix} F_{h,x} \\ F_{h,y} \end{bmatrix} = \begin{bmatrix} f_{h,x} \\ f_{h,y} \end{bmatrix} \cos\alpha \tag{4.14}$$

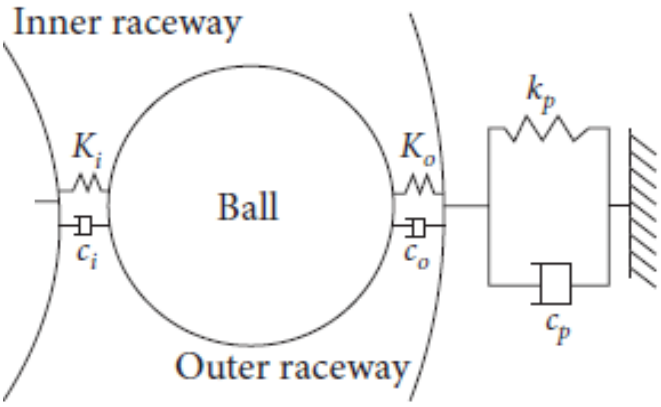

Figure 4.3: Scheme of the contact between a rolling element and raceways [49]

4.3.3 Damping in rolling elements contacts

To account for lubricant film damping effect in contacts between rolling elements and raceways, dampers $c_j(t)$ are included in the model. The force $f_{d,j}(t)$ associated with a rolling element located at $\phi_j(t)$ acts on the inner and outer raceways in the radial direction and is given by:

$$f_{d,j}(t) = c_j\ \dot{\delta}_j(t)^n \zeta_j(t) = c_j(t)\ \dot{\delta}_j(t) \qquad \zeta_j(t) = \begin{cases} 1 & if\ \delta_j(t) > 0 \\ 0 & if\ \delta_j(t) \leq 0 \end{cases} \tag{4.15}$$

The total contact damping forces acting on the inner and outer raceways in x and y directions are:

$$\begin{bmatrix} f_{d,x}(t) \\ f_{d,y}(t) \end{bmatrix} = c_j \sum_{j=1}^{\mathrm{Z}} \dot{\delta}_j(t) \zeta_j(t) \begin{bmatrix} \cos\phi_j \\ \sin\phi_j \end{bmatrix} \tag{4.16}$$

The damping coefficient c_j of each rolling element is the resultant damping coefficient of the inner race-roller and outer race-roller contacts. As reported in Section 3.6.1, the total damping coefficient for each rolling element is

$$\frac{1}{c_j} = \frac{1}{c_{sj,IR}} + \frac{1}{c_{sj,OR}} \tag{4.17}$$

where:

$$c_{s,j} = c_{h,j} + c_{ehl,j} \tag{4.18}$$

c_{ehl} is the damping coefficient of the lubricant film, that can be computed through the empirical formula reported in Section 3.5.2:

$$c_{ehl,j_{IR}} = 0.1963\ R_{x,IR}^{0.781}\ L_r^{0.769}\ E'^{1.069}\ \eta_0^{0.531}\ \alpha_{PV}^{0.424}\ q_j^{-0.136}\ u_{s,j}^{-0.434}\ F_{inlet}^{-0.563} \tag{4.19}$$

$$c_{ehl,j_{OR}} = 0.1963\ R_{x,OR}^{0.781}\ L_r^{0.769}\ E'^{1.069}\ \eta_0^{0.531}\ \alpha_{PV}^{0.424}\ q_j^{-0.136}\ u_{s,j}^{-0.434}\ F_{inlet}^{-0.563} \tag{4.20}$$

The description of all the parameters involved in the equation can be found in the afore mentioned paragraph; the load Q_j, used for the computation of the distributed load q_j, is the current load on rolling element j.

c_h is the structural damping coefficient of metal parts single Hertzian contact in a rolling element bearing, as it is discussed in Section 3.5.1, and can be written as:

$$c_{h,j} = \psi_L \frac{k_{h,j}}{2\pi f} \tag{4.21}$$

where ψ_L is the loss factor, that for steel bodies ranges between $0.2 - 3 * 10^{-4}$, $k_{h,i}$ is the contact stiffness and f is the excitation frequency.

Taking into account the rolling elements contact angle:

$$\begin{bmatrix} F_{d,x} \\ F_{d,y} \end{bmatrix} = \begin{bmatrix} f_{d,x} \\ f_{d,y} \end{bmatrix} \cos\alpha \tag{4.22}$$

4.3.4 Model of raceways stiffness

Whenever a rolling element hits a defect in the loaded area, the ring containing the defect is driven into a vibration flexural motion, that can occur in several different modes. However, to make the characterization of parameters easier, only the first flexural mode, which has an elliptic shape, is considered. The bearing rings are also excited by the variable compliance effect, that generates vibration even in absence of defects.

Thus, the external and internal ring are represented through a mass-spring system with one-degree of freedom in radial direction and the natural frequency for the flexural vibration mode n is given by [54]:

$$\omega_n = \frac{n(n^2-1)}{\sqrt{1+n^2}} \sqrt{\frac{EI}{\mu_m r^4}} \tag{4.23}$$

Where E is the modulus of longitudinal elasticity, I is the moment of inertia of the cross-section of the ring, μ_m is the mass per unit length, r is the radius of neutral axis of the ring.

Depending on the rings dimensions and characteristics and imposing $n = 2$ for the first flexural mode, we can evaluate the rings stiffness:

$$k_{IR} = m_{IR}\, \omega_{IR}^2 \tag{4.24}$$

$$k_{OR} = m_{OR}\, \omega_{OR}^2 \tag{4.25}$$

In this model, as mentioned in the assumptions, the inner ring is considered rigid, to obtain quasi-static load distribution for the defect-free case that agree with the well-known Stribeck results [45], and K_{OR} represent the stiffness of the external structure.

4.3.5 Lubricant properties

As mentioned in the assumptions, the lubricant is an incompressible Newtonian fluid, whose properties depend on temperature.

- lubricant density ρ can be read on manufacturer's datasheet.
- kinematic viscosity ν_0 depends on temperature, so, knowing the working temperature of the bearing, it can be derived through interpolation.
- dynamic viscosity is

$$\mu_0 = \eta_0\, \rho \tag{4.26}$$

- viscosity-pressure coefficient α_{PV} can be derived using the ASME relation [55]

$$\alpha_{PV} = 0.1122\, \eta_0^{0.163} \tag{4.27}$$

  that returns the viscosity-pressure coefficient in mm^2/N

4.3.6 Defect model

Bearing fail as a consequence of multiple causes: corrosion, wear, plastic deformation, fatigue, lubrication failure, electrical damage, fracture and incorrect design or setup operations. Most common spalling fatigue leaves pits on races or rollers due to periodic contact stress.

In this work only localised defects on the races are analysed. Many researches report the displacement excitation function, which is used, like in this work, to model the defect effect. Simplest models assume the defect as cube-shaped, but this does not model the real path the rollers follow when they fall into the defect itself. Thus, the geometrical relationship between the roller and the outer raceway centres is described through the following model, introduced by Wen [52]:

$$H_d\left(\phi_j(t)\right) = \begin{cases} \min\left(h_{def}, r_r + r_{OR}\left(\cos\beta_j - 1\right) - \sqrt{r_r^2 - r_{OR}^2\left(\sin\beta_j\right)^2}\right) & if\ \phi_j \in [\theta_i - 0.5\Delta\theta_{def}, \theta_i + 0.5\Delta\theta_{def}] \\ 0 & otherwise \end{cases} \quad (4.28)$$

where h_{def} is the defect height, R_r is the roller radius, R_{OR} is the outer raceways mean radius, θ_i is the angular position of the centre of the defect, $\Delta\theta_{def}$ is the angular extension of the defect, β_j is the current angle between the roller centre and the edges of the defect, which is defined as follows:

$$\beta_j = \begin{cases} \phi_j - \theta_i + 0.5\,\Delta\theta_{def} & if\ \theta_i - 0.5\,\Delta\theta_{def} \le \phi_j \le \theta_i \\ \theta_i - \phi_j + 0.5\,\Delta\theta_{def} & if\ \theta_i \le \phi_j \le \theta_i + 0.5\,\Delta\theta_{def} \end{cases} \quad (4.29)$$

$$\Delta\theta_{def} = \frac{L_{def}}{r_{race}} \quad (4.30)$$

In the outer raceway case $r_{raceway} = r_{OR} = \frac{d_{OR}}{2}$

The equation for the inner race is:

$$H_d\left(\phi_j(t)\right) = \begin{cases} \min\left(h_{def}, r_r + r_{IR}\left(\cos\beta_j - 1\right) - \sqrt{r_r^2 - r_{IR}^2\left(\sin\beta_j\right)^2}\right) & if\ \phi_j \in [\theta_i - 0.5\Delta\theta_{def}, \theta_i + 0.5\Delta\theta_{def}] \\ 0 & otherwise \end{cases} \quad (4.31)$$

In this case $r_{raceway} = r_{IR} = \frac{d_{IR}}{2}$

The comparation of the depth of defect used in this work compared to the one that is most frequently used in dynamic models of defective bearings can be seen in Figure 4.4.

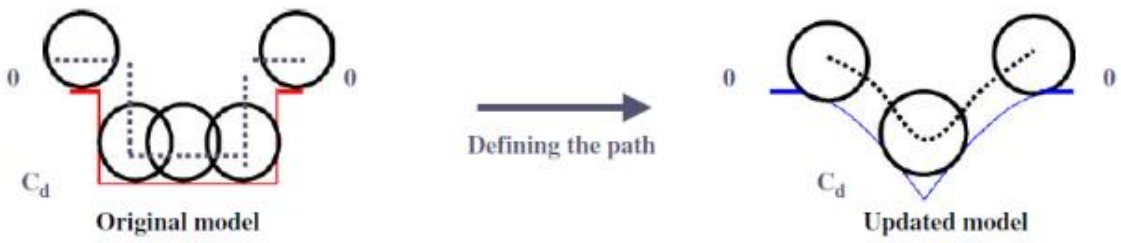

Figure 4.4: Qualitative comparison of spall profile

4.3.7 Nonlinear dynamic equations of motions

Using the equations presented in Paragraph 4.3, the nonlinear equations of motion for the inner and outer raceways are given by

$$[M]\{\ddot{s}\} + [C]\{\dot{s}\} + [K]\{s\} + \{F_{hertz}\} + \{F_{damp}\} = \{F_{ext}\} \tag{4.32}$$

where:

$$\{s\} = \begin{Bmatrix} x_{IR} \\ y_{IR} \\ x_{OR} \\ y_{OR} \end{Bmatrix} \tag{4.33}$$

$$\{\dot{s}\} = \begin{Bmatrix} \dot{x}_{IR} \\ \dot{y}_{IR} \\ \dot{x}_{OR} \\ \dot{y}_{OR} \end{Bmatrix} \tag{4.34}$$

$$\{\ddot{s}\} = \begin{Bmatrix} \ddot{x}_{IR} \\ \ddot{y}_{IR} \\ \ddot{x}_{OR} \\ \ddot{y}_{OR} \end{Bmatrix} \tag{4.35}$$

$$[M] = \begin{bmatrix} m_{IR} & 0 & 0 & 0 \\ 0 & m_{IR} & 0 & 0 \\ 0 & 0 & m_{OR} & 0 \\ 0 & 0 & 0 & m_{OR} \end{bmatrix} \tag{4.36}$$

$$[C] = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & c_{OR} & 0 \\ 0 & 0 & 0 & c_{OR} \end{bmatrix} \tag{4.37}$$

$$[K] = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & k_{OR} & 0 \\ 0 & 0 & 0 & k_{OR} \end{bmatrix} \tag{4.38}$$

$$\{F_h\} = \begin{Bmatrix} F_{h,x} \\ F_{h,y} \\ -F_{h,x} \\ -F_{h,y} \end{Bmatrix} \tag{4.39}$$

$$\{F_d\} = \begin{Bmatrix} F_{d,x} \\ F_{d,y} \\ -F_{d,x} \\ -F_{d,y} \end{Bmatrix} \tag{4.40}$$

$$\{F_{ext}\} = \begin{Bmatrix} 0 \\ -\frac{F_r}{2} \\ 0 \\ 0 \end{Bmatrix} \tag{4.41}$$

k_{OR} and c_{OR} are respectively the stiffness and damping of outer ring, that are presented in Sections 4.3.4 and 3.5.1; F_r is the external radial load applied to the outer ring of the bearing and it is divided by two, because only a row of rolling elements is considered and we assumed that load is equally supported by the two rows.

5 SIMULATION SETUP AND RESULTS

The multi-body nonlinear dynamic model of the spherical roller bearing presented in Chapter 4 was implemented in Matlab and Simulink and the equations of motion were solved using the differential equation solver Ode 4, that is based on Runge-Kutta method with fixed step size.

The developed model allows the simulation of bearing performance under different conditions of shaft rotational speed, radial load and operating temperature. The presence of a defect on the inner, outer or both the raceways and the effects of lubrification and rolling element slipping can be easily included in the simulation setup.

Since the initial conditions of simulations regarding the position and velocities of inner and outer races are set to zero, the model simulates also a transient; therefore, in the simulation setup the starting revolution or the starting time from which data are recorded for further analysis can be set.

In this chapter the data used for simulation and the main results are presented.

5.1.1 Simulation data

All the parameters involved in the dynamic equations of motion can be derived using the equations presented in the previous sections and using the following data and diagrams, relative to the SKF 22240 CCK/W33 spherical roller bearing (

Table 5.1) and the ISO VG 150 lubricant (Figure 5.1).

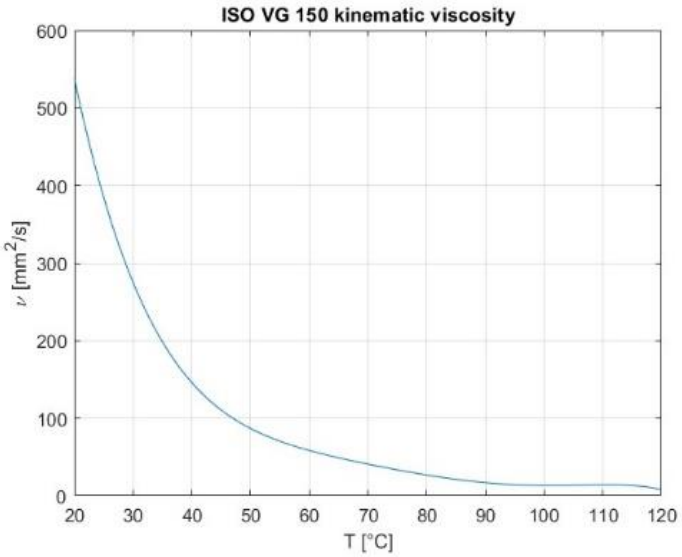

Figure 5.1: Diagram of ISO VG 150 lubricant kinematic viscosity vs. temperature

Parameter	Value	Unit
Number of rollers ($\boldsymbol{Z}$)	19	/
Roller diameter ($\boldsymbol{D_r}$)	38.26	mm
Bearing width ($\boldsymbol{L_r}$)	33.41	mm
Radial clearance ($\boldsymbol{c_d}$)	0	µm
Inner raceway diameter ($\boldsymbol{d}$)	200	mm
Outer raceway diameter ($\boldsymbol{D}$)	360	mm
Inner ring external contact diameter ($\boldsymbol{d_2}$)	238	mm
Outer ring internal contact diameter ($\boldsymbol{D_1}$)	313	mm
Contact angle ($\boldsymbol{\alpha}$)	11.27	°
Bearing exponential coefficient ($\boldsymbol{n_h}$)	1.11	/
Mass of inner ring ($\boldsymbol{m_{IR}}$)	10.5	kg
Mass of outer ring ($\boldsymbol{m_{OR}}$)	15.3	kg
Mass of rollers ($\boldsymbol{m_{roller}}$)	0.349	kg
Material of rings and rollers	steel	/
Elastic modulus of raceways ($\boldsymbol{E_{race}}$)	210	GPa
Elastic modulus of rollers ($\boldsymbol{E_{roller}}$)	207	GPa
Poisson's ratio of rings ($\boldsymbol{\nu_{ring}}$)	0.3	/
Poisson's ratio of rollers ($\boldsymbol{\nu_{roller}}$)	0.3	/
Roughness of raceways ($\boldsymbol{\sigma_{race}}$)	0.1	µm
Roughness of rollers ($\boldsymbol{\sigma_{roller}}$)	0.03	µm
Steel loss factor ($\boldsymbol{\psi_L}$)	0.0002	/
Bearing hardness ($\boldsymbol{hd}$)	$7.65 \cdot 10^9$	/
Cage starting angular position ($\boldsymbol{\phi_0}$)	1.5π	rad
Lubricant density ($\boldsymbol{\rho}$)	872	kg/m^3
Defect angular position ($\boldsymbol{\theta_i}$)	1.5π	rad
Defect linear length ($\boldsymbol{L_{def}}$)	2	mm
Defect height ($\boldsymbol{h_{def}}$)	1	mm

Table 5.1: Simulation data

5.2 SIMULATION RESULTS

The discussion on the spherical roller bearing simulation output that will be presented in this section are based on roller bearing data of Table 5.1 and the tuneable parameters are set to

$$n_{IR} = 127 \text{ rpm}$$

$$F_r = 124.8 \text{ kN}$$

$$T = 30 \text{ °C}$$

The presence of defects and the effect of lubrication are included or excluded, depending on the simulation needs.

5.2.1 Rolling element load

The static load distribution as a function of the cage position is obtained from the defect-free simulation and is presented in Figure 5.2.

In case of dry contact, the load distribution for a rolling element as a function of the cage angle is in accordance with theory: the maximum load is in correspondence to $\phi = 270°$, that is the centre of the load zone, whose angular extension is $\Delta\phi = 180°$ because the diametral clearance $c_d = 0$.

If the effect of lubricant film thickness is considered, there is a reduction in the load zone extension, since it affects the deformation, as presented in Equation 4.5. The maximum load is also higher than in the case of dry contact, because the rolling element crossing the centre of the load zone has to sustain the load of those that have not been deformed yet due to the film thickness effect, even if they have already entered the dry load zone.

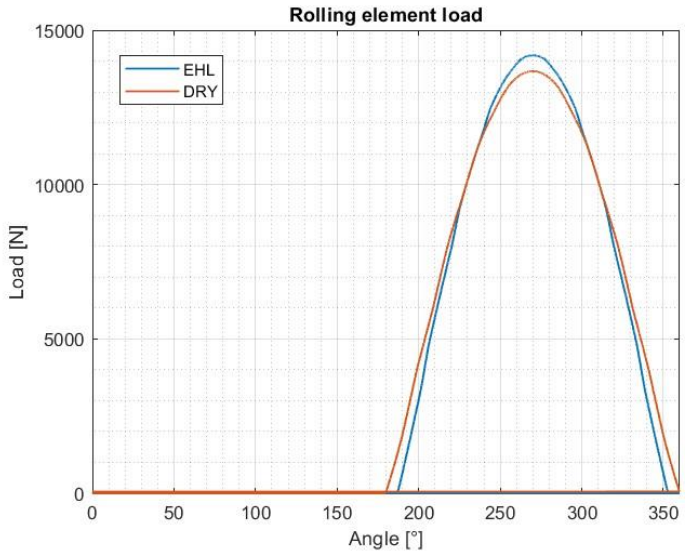

Figure 5.2: Plot of rolling element load distribution in case of dry (orange) or lubricated (blue) contact

5.2.2 Fluid film thickness

The fluid film thickness profile as function of the cage angular position is shown in Figure 5.3, while Figure 5.4 relates it to the rolling element load distribution. In this model the effect of lubricant film thickness is considered only when the rolling element is in the load zone, because the empirical formula presented in Equation 4.8 for the film thickness computation includes the load parameter, that accounts for the current Hertzian contact force sensed by the rolling element. This makes sense, because outside the load zone, the effect of film thickness does not have effect, since the rolling element deformation is negative and it does not produce a Hertzian contact force. The computation of the film thickness is performed using the dry Hertzian contact force, then it is subtracted to the current dry deformation and finally Hertzian contact force is obtained, as described in Section 4.3.2.

As the rolling element reaches the maximum load point, the lubricant is squeezed apart and compressed, reaching conditions similar to solid state, so the film thickness becomes thinner.

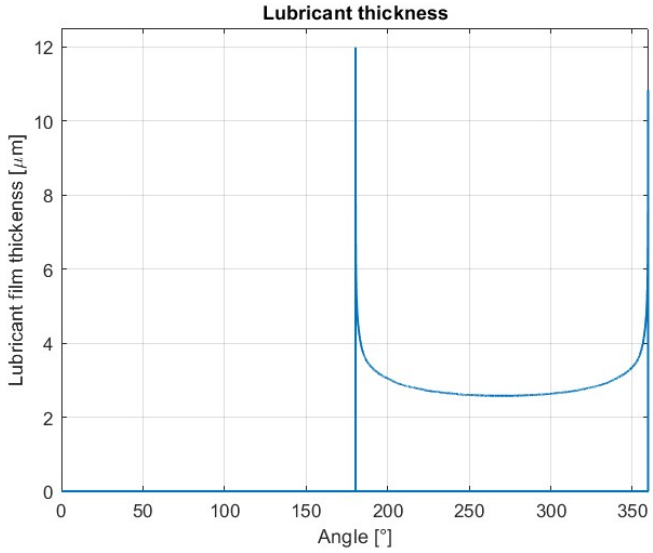

Figure 5.3: Film thickness profile

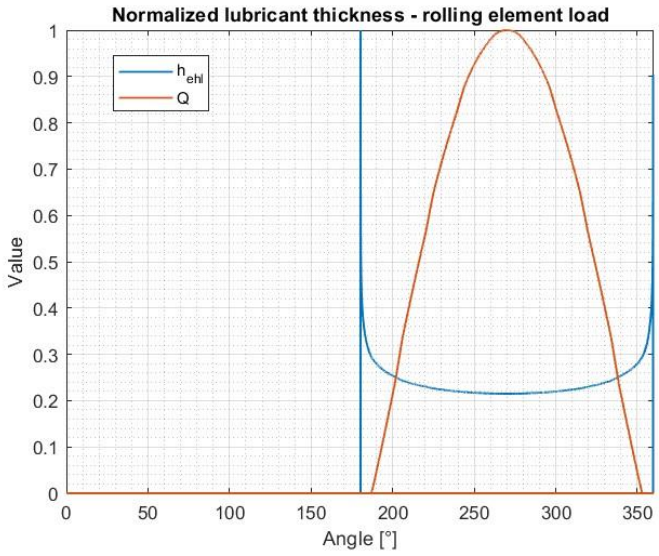

Figure 5.4: Normalized lubricant film thickness (blue) and rolling element load as function of the cage angular position

5.2.3 Lubricant film damping coefficient

The lubricant film damping coefficient is computed through the empirical formula presented in Equation 3.116 and with the assumption regarding the rolling element load made in the previous section. Its maximum value is reached and the entrance of the load zone and its minimum is in correspondence of the maximum load point, where lubricant reaches its maximum pressure (Figure 5.5).

Figure 5.6 shows the relationship between damping coefficient and load in case of the presence of a defect on the outer race.

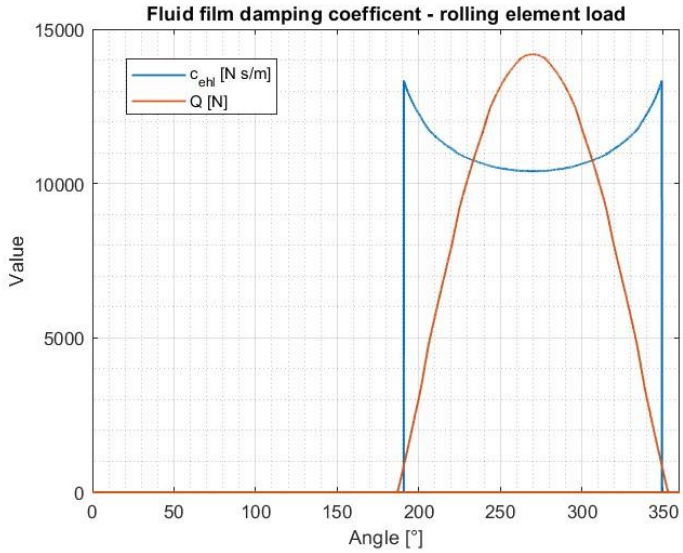

Figure 5.5: Lubricant film damping coefficient (blue) and rolling element load (orange) as function of cage angle

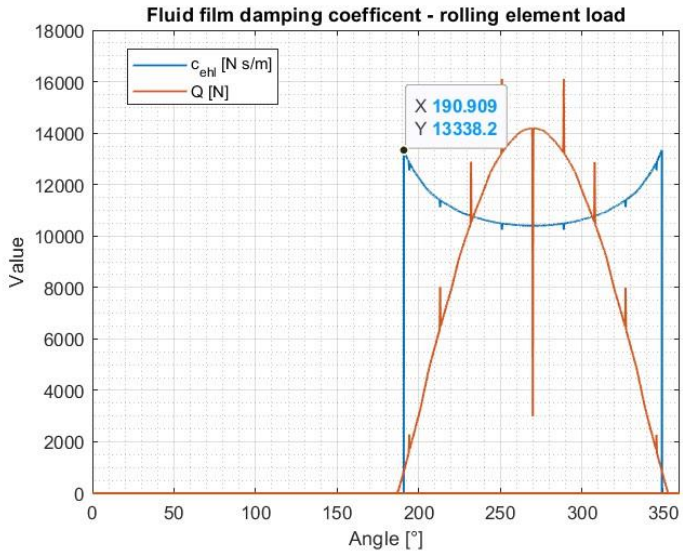

Figure 5.6: Lubricant film damping coefficient (blue) and rolling element load (orange) as function of cage angle for a spherical roller bearing with a localized defect on the outer race

5.2.4 Defect induced effect

The defect-induced reduction in rolling element deformation as function of cage angular position is shown in Figure 5.7.

The modified profile presented in Section 4.3.6 simulates the deviation of the rolling element centre from the circular trajectory around the bearing centre. The original profile assumed a constant value $H_d = h_{def}$ for the whole defect angular extension: in the case analysed in this work the defect heigh is $h_{def} = 1$ mm and it produces too large vibration. From the plot it can be seen that the maximum defect-induced deformation reduction is only about $23\ \mu\text{m}$, because it considers only the true penetration of the rolling element into the defect, depending on their geometries.

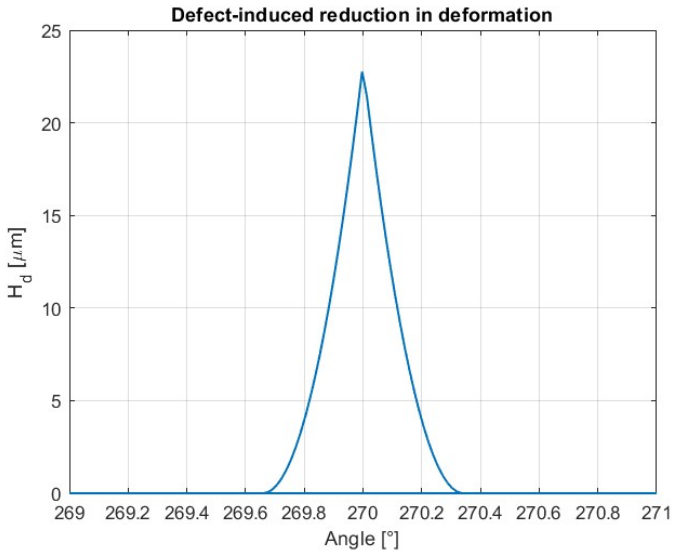

Figure 5.7: Defect-induced reduction in rolling element deformation as function of the cage position

Figure 5.8 shows the normalized defect-induced reduction in deformation and the rolling element load as function of the cage angular position for a rolling bearing with a localized defect on the outer race (left) and on the inner one (right).

In the former case, it can be seen that whenever a rolling element enters into the defect ($\phi = 270°$), it senses an abrupt reduction in its contact deformation and, therefore, in its Hertzian load. This effect induces a peak in the load of nearby rolling elements that have already entered the load zone. The impulsive effect on load is greater as rolling element are closer to the one that encountered the defect.

A defect on the inner race rotates with the same rotational speed of the shaft n_{IR}, so it has an effect only if it is in the load zone, as shown in Figure 5.8 and in Figure 5.9.

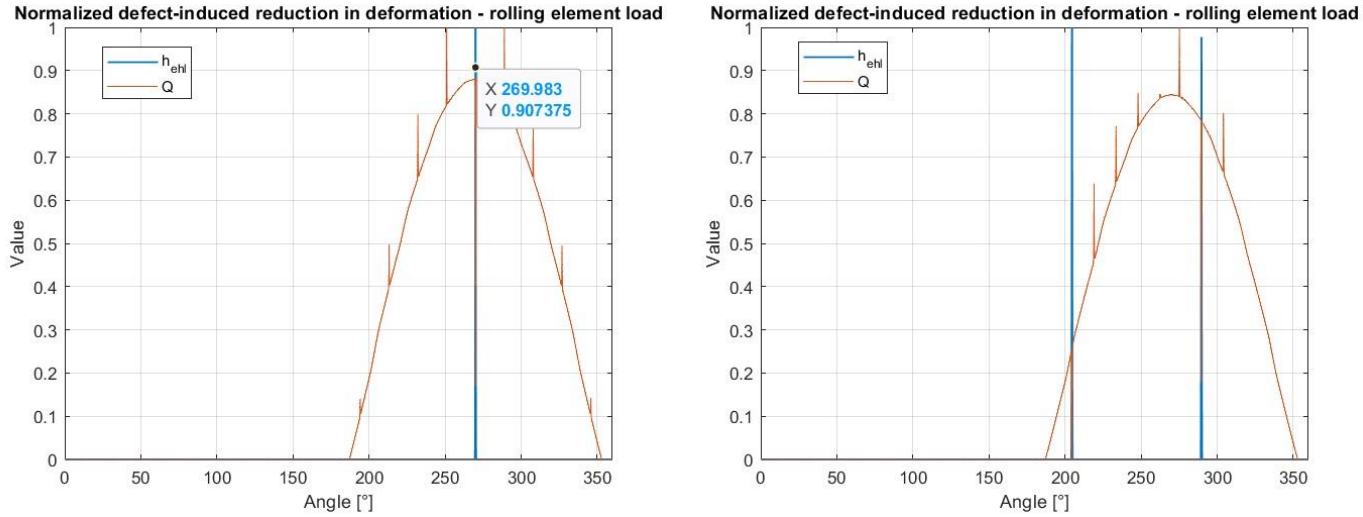

Figure 5.8: Normalized defect-induced reduction in deformation (orange) and rolling element load (orange) for a bearing with a localized defect on the outer (left) and inner (right) raceway

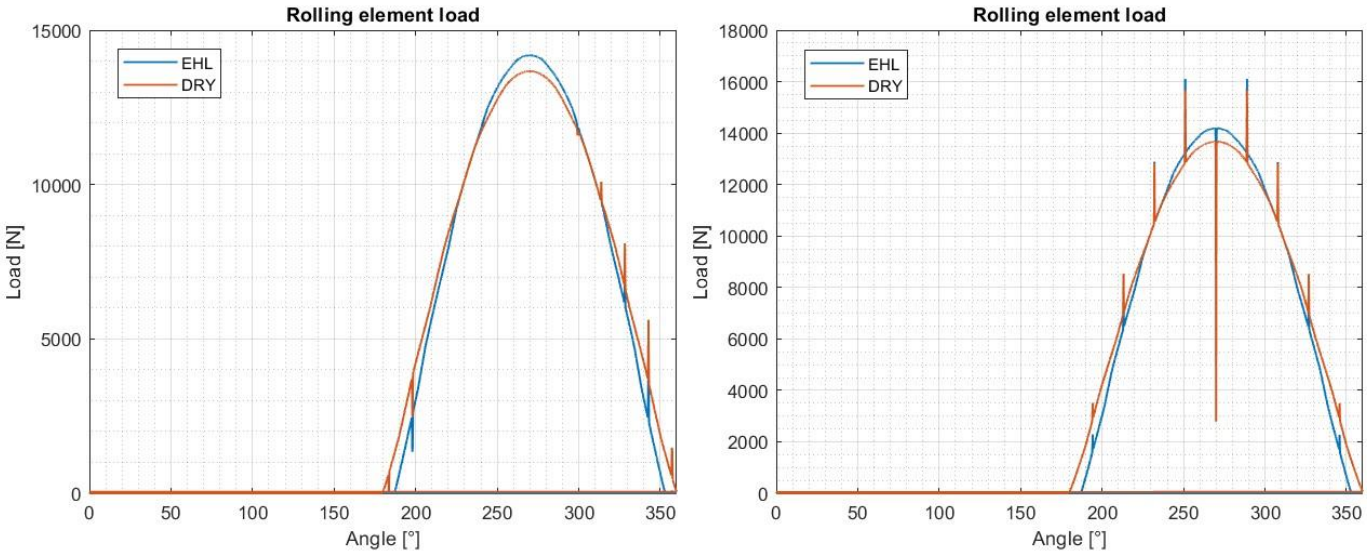

Figure 5.9: Rolling element load distribution for dry (orange) and lubricated (blue) contact in case of a localized defect on the inner (left) or outer (right) receway

6 EXPERIMENTAL TEST RIG

Numerical data obtained through simulation using the nonlinear multi-body dynamic should be compared to experimental data obtained under similar condition of rotation speed, external loads and defect type to validate the model.

Figure 6.1 shows the photo of the test rig designed by Brusa et al. [56] to conduct active monitoring of bearings.

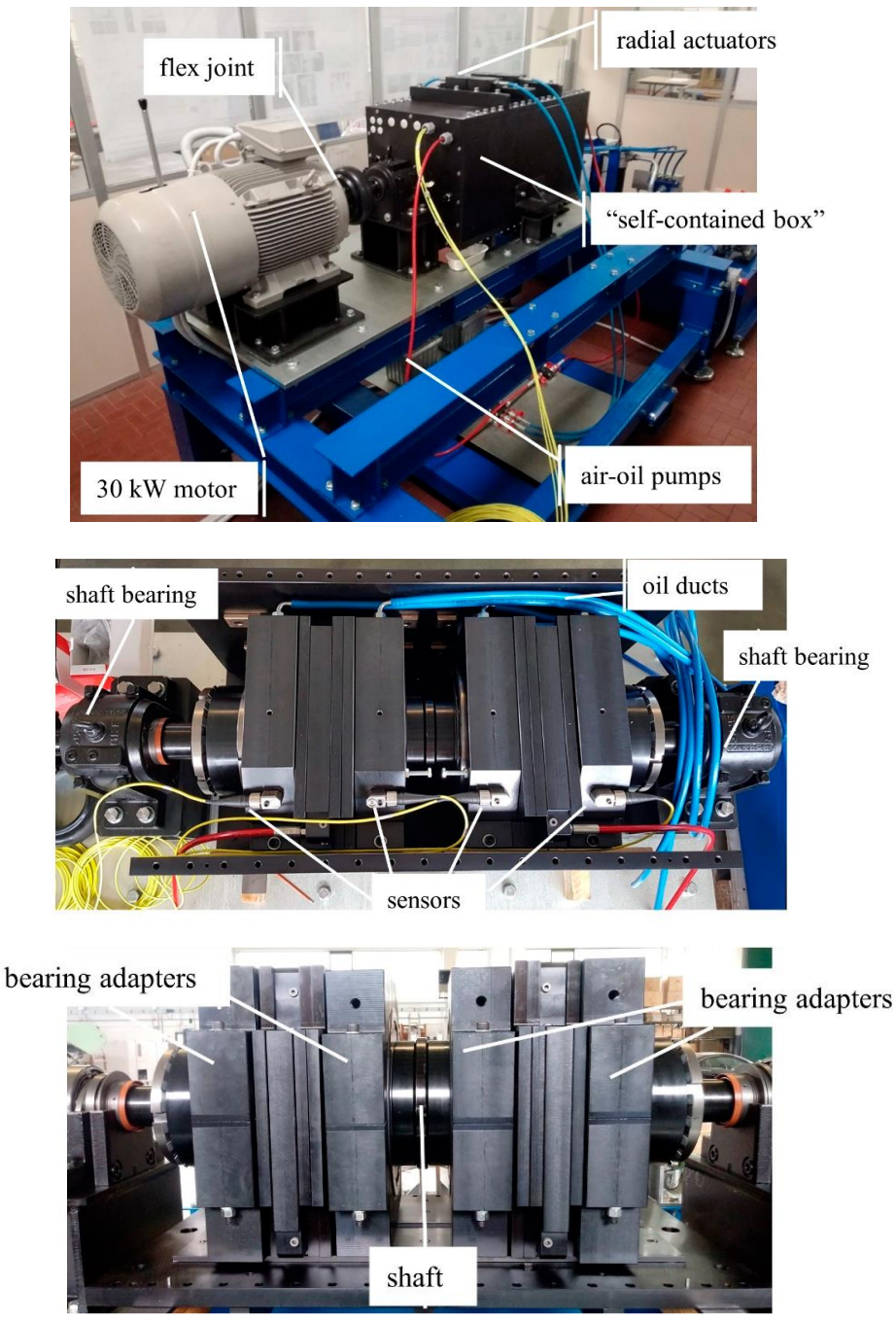

Figure 6.1: Pictures of the test rig [56]

The test rig is powered by a 30 kW three-phase motor by SIEMENS® connected to the SIEMENS® G120_CU240E_2 INVERTER with Brake Resistor, that drives the shaft up to 195 Nm of maximum torque at the rated spin speed of 1470 rpm. Speed and torque controls, managed by the inverter, can set the motor speed up to 2500 rpm. The "self-contained box" is made of two housing blocks with oil lubrication equipped with up to four different roller bearings, whose reactions are directly balanced by the box. Hydraulic actuators on the upper panel of the box and between each pair of bearings allow the application of radial and axial static loads respectively up to 200 kN.

As it is mentioned in Brusa et al. [56], the main advantages of this test rig can be resumed as follows:

- independent radial and axial loads up to 200 kN on each tested bearing
- simultaneous testing of four bearings, that allows the self-balance of applied loads with minimal transmission to the platform
- high modularity, that enables testing on different sized bearings up to 420 mm of outer diameter
- direct measure of friction torque of tested bearings

Four SKF CMSS 2200 T sensors are mounted along the radial direction on the bearing-box adapters to measure vibration amplitude and frequency and operating temperature of each bearing. Vibration signals are acquired through the DAQ system and digitally sampled at $f_s = 20480$ Hz

Figure 6.2 shows the CAD model section of the "self-containing box" that highlights the position of the accelerometer with respect to the centre of the bearing. In the reference system used in the model developed in this work:

$$\Delta y = 215.25 \text{ mm}$$

$$\Delta x = 150.77 \text{ mm}$$

The angle formed by the segment that links the centre of the bearing to the sensor and the x-axis is

$$\alpha_{sensor} = \tan^{-1}\frac{\Delta y}{\Delta x} \tag{6.1}$$

So the acceleration sensed by the sensor can be written as:

$$\ddot{s}_{sensor} = \ddot{x}_{OR}\cos\alpha_{sensor} + \ddot{y}_{OR}\sin\alpha_{sensor} \tag{6.2}$$

The 22240 CCK/W33 bearing tested on this test rig has the characteristic frequencies reported in Table 6.1

Characteristic Frequency	Value [Hz]
Ball Pass Frequency relative to Outer Race (***BPFO***)	8.176 f_r
Ball Pass Frequency relative to Inner Race (***BPFI***)	10.824 f_r
Ball Spin Frequency (***BSF***)	3.464 f_r
Fundamental Train Frequency (***FTF***)	0.43 f_r

Table 6.1: SKF 22240 CCK/W33 characteristic frequencies

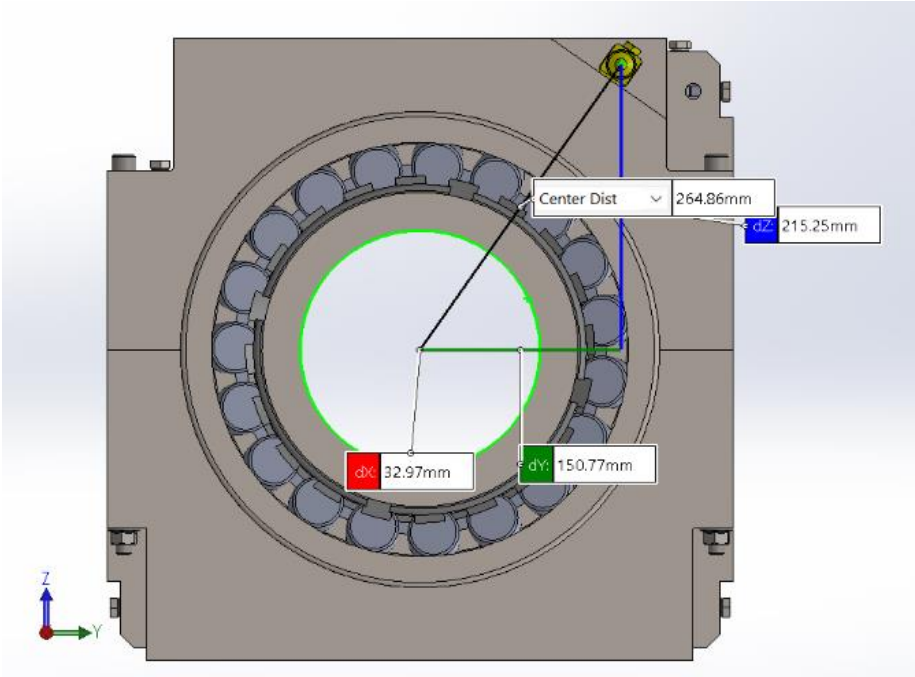

Figure 6.2: CAD model of the box showing the position of the sensor with respect to the shaft center

7 VIBRATION ANALYSIS AND FAULT DETECTION: A REVIEW

Nowadays, since manufacturing companies struggle to maintain their competitiveness on the global market, they are making great efforts to reduce costs and improve quality. Being cost saving and profitability achieved by higher industrial equipment's availability, reliability and maintainability, it is necessary to implement an effective machinery maintenance program. In fact, since it is impossible to effectively plan maintenance and production scheduling without an accurate identification of machine faults, the most important task in machinery maintenance should be fault detection and diagnostics. In addition, an accurate diagnostic activity assures fast troubleshooting and repair time that can improve machine availability significantly. [7]

The required degree of fault prevention strongly depends on the complexity of the system and its applications. For example, bearings are among the most critical components that have wide application ranges in many industries and have proven to be reliable when properly applied. As rolling element bearings have been gradually employed in more severe application with higher loads, speeds and restricted lubrication, thanks to improvements in materials, design, lubrication technology and service-life, condition monitoring and fault detection have become important to ensure safe operation of rotating machines. So, many approaches, such as thermal, oil debris and vibration analysis have been introduced in diagnostic operation. [3]

The most advanced approach to fault diagnosis is represented by Intelligent Fault Diagnosis (IFD), that is based on the application of machine learning theories, opening to an interesting way to release the contribution from the human labor, since the algorithm can automatically recognize the health status of machines. [4]

This chapter opens with a brief analysis of failure causes and a review of the main types of analysis used in fault detection with a focus on vibration analysis, that is involved in this work. Then, there is a description of the fault signatures and the characteristic frequencies of defects in bearings. In the last part of the chapter the machine learning algorithms, used in this work, and the way to train them are presented.

7.1 FAILURE CAUSES

The most important failure means, which can cause localized defects on bearings, are wear. crack, fatigue, corrosion, brinelling and lubrication starvation. On the other hand, inadequate maintenance, improper assembling and harsh operating conditions, such as insufficient lubrication or very high speed or acceleration, are responsible for the generation of distributed defects. [7] [3]

7.1.1 Wear damage

Wear is a frequent source of damage and occurs due to dirt and foreign particles entering the bearing because of inadequate sealing or contaminated lubrication. This increases the friction between metal contacts, that leads to a change in raceway profile and gives them a dull appearance due to abrasive effects. Wear damage gradually deteriorates the bearing, leading to an alteration in elements profiles and diameters (geometric errors) and increasing clearance.

Non-uniform diameters of worn rolling elements causes cage frequency vibration to be produced as the sequence of balls rotating through the load zone is period with the cage rotation frequency.

7.1.2 Fatigue damage

Fatigue damage is induced by loading: in loaded configurations, after a certain amount of time a minute fatigue crack begins below the surface and propagates towards the surface as the loading continues, until a piece of metals breaks away, leaving a pit in the contact area. If the bearing continues in its service operations, the damage spreads due to stress concentration. Fatigue effects grow faster in case of overloads, over speeding or oil starving, so the bearing may fail before the predicted lifetime. The surface damage severely disturbs the nominal motion of the rolling elements by introducing short time impacts repeated at the characteristic rolling element defect frequencies.

7.1.3 Corrosion damage

Corrosion damage results in rust on the bearing elements, which originates from the exposure to water, acid, acid lubrication or to external elements in case of incorrect storage and causes excessive noise during operation. Corrosion may be also caused by condensation, that originates from sudden cooling of the bearing in humid air from the working temperature and can damage its components even before the installation. Rust particles also have an abrasive effect that generates wear, while rust pits form the initiation sites for subsequent flaking and spalling.

7.1.4 Brinelling

Brinelling consists in permanent indentation created by rolling element overload. It can be seen as regularly spaced indentations distributed over the entire raceway circumference with a shape similar to the Hertzian contact. It may be caused by static loading, which leads to observable plastic deformation of raceways, vibration and shock loads and increases bearing noise and vibration, leading to a premature failure. In fact, each indentation acts as a small fatigue site producing sharp impacts with the passage of rolling elements. Continued operation leads to the development of spalling effects.

7.1.5 Lubrication starvation

Inadequate lubrication in terms of quantity and quality is responsible for the premature failure of bearings because it leads to skidding, slippage and bearing seizure phenomena. In the Hertzian contact zones contacting surface weld together and then they are torn apart due to high loads, as if there was not a sufficient lubrication. This may happen for the cage-roller, roller-race and cage-race contacts and can severe consequences because high temperatures can anneal the bearing elements, reducing hardness and fatigue life.

7.2 Fault Detection

The detection of faults in bearings, which has been recognized an important aspect for preventing catastrophic failures and planning effective maintenance, can be conducted following several approaches, such as thermal analysis, oil debris analysis and vibration analysis. [3]

7.2.1 Thermal analysis

Thermal analysis is a tool that can generates warnings about the overheating of the bearing system, but it cannot be used to identify the type and size of defects. It is performed employing thermocouples or other types of temperature monitoring devices at the inlet and outlet of the test chamber.

7.2.2 Oil debris analysis

Debris analysis is usually employed for those bearings that use oil lubrication and consists of oil sampling to find metal particles, analytical tests and data interpretation. There are many oil debris analysis techniques such as elementary spectroscopy, wear particle analysis, fine particulates analysis, molecular analysis and electrochemical chemistry, that can provide information on quantity, form and size distribution of the debris, leading to the damage type detection.

Only a small amount of oil debris extracted from oil analysis can reveal the severity of rolling contact fatigue wear and this can be a critical indication that the bearing has to be changed. Even if this type of analysis cannot differentiate between the type of damaged component because of the material similarity, this technique is often used along with the vibration analysis in bearing condition monitoring programs.

7.3 Vibration Analysis

Each rotating machine has its own vibration signature as result of the rotation of all the components of the mechanical system, such as motors, shafts, gears and bearings. Rolling element bearings play a significant role in machinery vibration, because their structural elements act as springs and add mass to the whole system, defining the vibration response to external time-varying and excitation forces, that produce time-varying loads. Those forces are accounted in the bearing design, but they can be greatly amplified by imperfections and defects in bearing components.

The detection of progressive bearing deterioration during operation through vibration measurements has been in use for a long time and has become more economically convenient in recent years. Vibration diagnostics usually consist in extracting features from a signal and associating them with healthy or faulty components.

A healthy bearing under constant load is likely to move towards steady state dynamic equilibrium because of the symmetry in rolling bearing elements, while the strength of vibration may increase when a defect occurs, since a transient force takes place each time a component contacts the defective surface, resulting in rapid accelerations. [3]

However, the success of any monitoring program largely depends on the accuracy of the measurements, so the measuring instrumentation must be properly calibrated, the sensor mounting should not limit the frequency and dynamic ranges of sensors and measurements must be always taken at the same locations. The measurements of machine vibrations can be made using a wide range of transducers, such as velocity transducers, microphones, which are inexpensive tools for the non-contact vibration measurement and piezo-electric accelerometers, that are the most used transducers, thanks to their wide frequency response and dynamic range and the possibility to use acceleration for computing velocities and displacements. [7]

The most common techniques for the vibration analysis are the time-domain and the frequency-domain techniques.

7.3.1 Time-domain techniques

Time-domain vibration signals, if properly understood, contain much information.

The most straightforward technique is the visual inspection of the time-domain waveform. Figure 7.1 shows simulated vibration signal waveforms of 0.4 s from a bearing containing an inner (above) and an outer (below) raceway defect. Repeated impacts can be observed at the time periods corresponding to the interval T_{BPFI} and T_{BPFO}. In the first plot the amplitude modulation, that depends on the entering and exit of the defect from the load zone is also visible.

For example, regarding the first plot, the time distance between two peaks is $\Delta t_{peaks} = 0.0254$ s and the theorical T_{BPFI} is:

$$T_{BPFI} = \frac{60}{10.824\, n} = 0.0254\ s \tag{7.1}$$

Vibration signals may be very complicated and impacts cannot be so clearly distinguishable if the whole mechanical system is composed of many components, making the visual recognition of defects difficult, as it can be seen in Figure 7.2. For that reason, it is convenient to extract some characteristic parameters, starting from the time-domain signal $x(t)$: [57]

- *peak level*: it is defined as the half difference between the maximum and minimum vibration level, but it is not always a good indicator, because false data caused by noise can have a strong effects on this feature

$$Peak = \frac{1}{2}(\max x(t) - \min x(t)) \tag{7.2}$$

- *mean*: it is the average vibration level

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i(t) \tag{7.3}$$

- *standard deviation*: it is the amount of dispersion of a set of values

$$STD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{7.4}$$

- *RMS value*: it is the normalized second statistical moment of the signal or its standard deviation

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}x_i^2} \tag{7.5}$$

- *crest factor*: it is the ratio of the peak value to the RMS of the signal and it is an index of its impulsive nature: the crest factor increases in presence of discrete impulses.

$$CF = \frac{Peak}{RMS} \tag{7.6}$$

- *kurtosis*: it is the normalized fourth statistical moment of the signal and provides a measure of the impulsive nature of the signal. A Gaussian distributed signal has got $Kurtosis = 3$.

$$Kurtosis = \frac{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^4}{STD^4} \tag{7.7}$$

- *shape factor*: it is the ratio of the RMS and mean of absolute values of the signal

$$SF = \frac{RMS}{\frac{1}{N}\sum_{i=1}^{N}|x_i|} \tag{7.8}$$

- *impulse factor*: it compares the height of the peak values to the mean level of the signal

$$IF = \frac{Peak}{\frac{1}{N}\sum_{i=1}^{N}|x_i|} \tag{7.9}$$

- *clearance factor*: it is the peak value divided by the squared mean value of the square roots of the absolute amplitudes; it is maximum for healthy bearings and start goes on decreasing for defective ball, defective outer and defective inner race respectively

$$CLF = \frac{Peak}{\left(\frac{1}{N}\sum_{i=1}^{N}\sqrt{|x_i|}\right)^2} \tag{7.10}$$

- *skewness*: it is the measure of the asymmetry of a signal distribution

$$Skew = \frac{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^3}{STD^3} \tag{7.11}$$

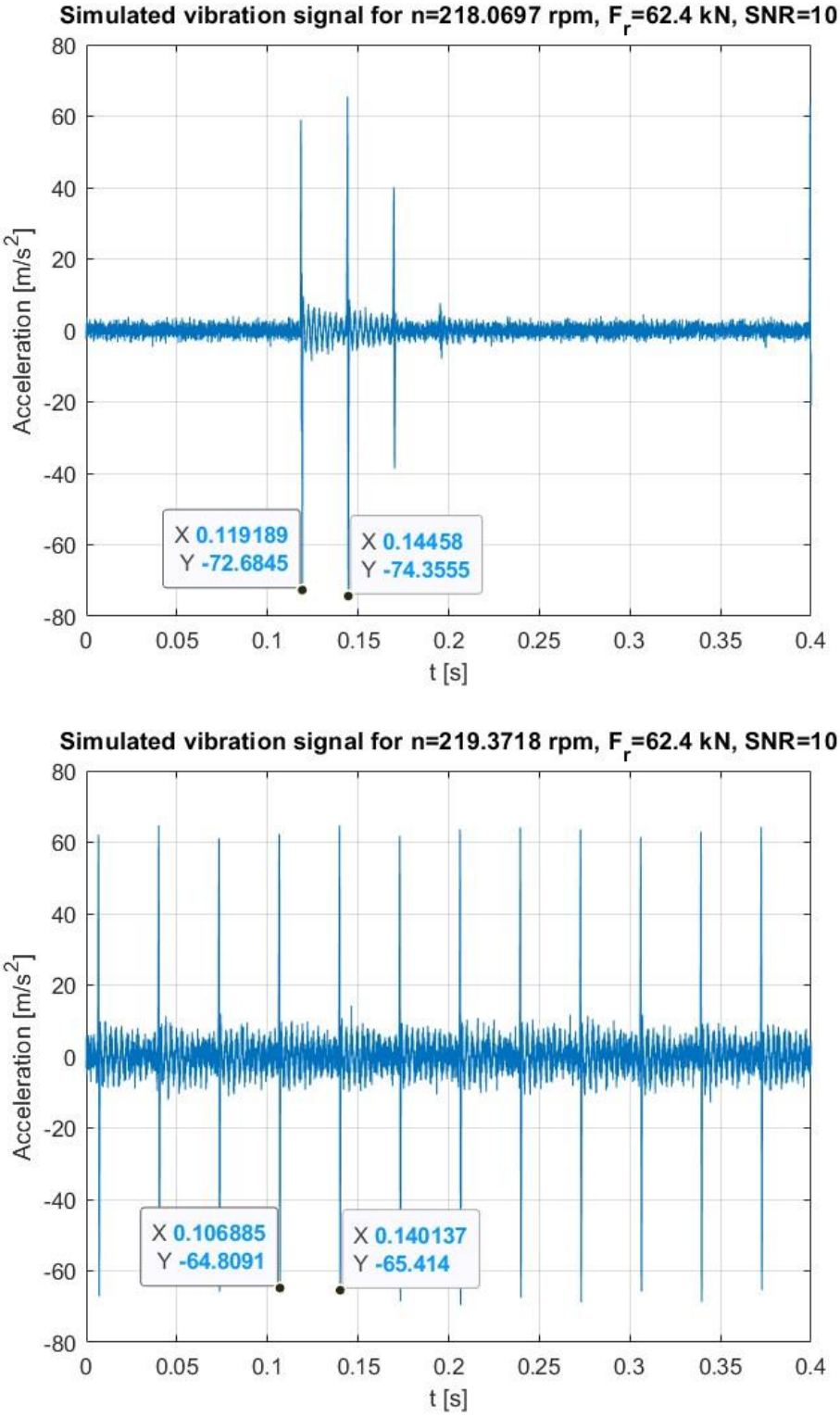

Figure 7.1: Simulated time-domain vibration plots of a defective bearing on the inner race (top) and outer race (bottom)

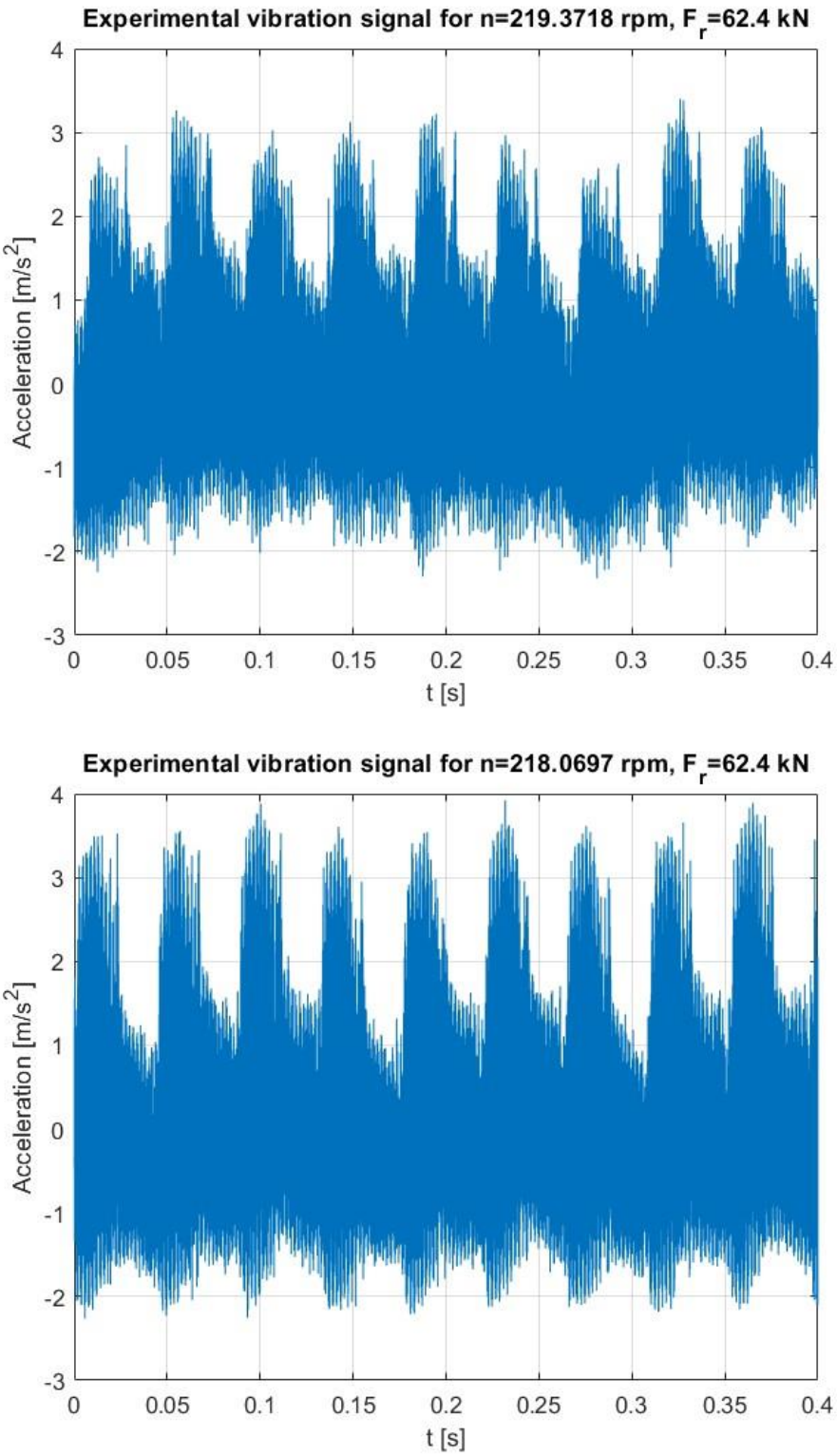

Figure 7.2: Experimental time-domain vibration plots of a defective bearing on the inner race (top) and outer race (bottom)

7.3.2 Frequency-domain techniques

Frequency analysis is a method used to extract the frequency content from the time-domain signal and it consists in the measure of vibration over many discrete neighboring narrow frequency bands.

Discrete Fast Fourier analysis of the time waveform is the most popular method to derive the frequency domain signal, but also the envelope analysis is gaining popularity. Power spectrum can be used to locate the defects by relating the defect characteristic frequencies to the mayor frequency components: a baseline spectrum is taken when bearing is in good conditions and the comparison with subsequent data is used to locate those frequencies in which significant increase in magnitude have occurred.

The conversion from a time-domain vibration signal to a frequency-domain one is computed through the Fourier Transform:

$$x(f) = \int_{-\infty}^{+\infty} x(t) e^{-j2\pi ft}\, dt \tag{7.12}$$

The operation on digitized data can be performed through the Discrete Fourier Transform:

$$x(k) = \int_{n=0}^{N-1} x(n) e^{-j2\pi \frac{kn}{N}}\, dt \tag{7.13}$$

The vibration level at each frequency represents the vibration over a narrow frequency band centered at the designed frequency with a bandwidth determined by the conversion parameters employed. [58]

Being s_k the spectrum for $k = 1,2,\dots,K$, where K is the number of spectrum lines, and f_k is the frequency of the kth spectrum line, the main frequency-domain features are:

- *mean frequency*: it is the average spectral vibration level

$$\overline{x_f} = \frac{1}{K} \sum_{k=1}^{K} s_k \tag{7.14}$$

- *frequency center*:

$$x_{fc} = \frac{\sum_{k=1}^{K} f_k s_k}{\sum_{k=1}^{K} s_k} \tag{7.15}$$

- *root mean square frequency*:

$$RMSF = \sqrt{\frac{\sum_{k=1}^{K} f_k^2 s_k}{\sum_{k=1}^{K} s_k}} \tag{7.16}$$

- *standard deviation frequency*:

$$STDF = \sqrt{\frac{\sum_{k=1}^{K} \left(f_k - x_{fc}\right)^2 s_k}{\sum_{k=1}^{K} s_k}} \tag{7.17}$$

Envelope analysis is another very common method in defect diagnosis and it was developed in early 1970s by Mechanical Technology Inc. The fundamental concept involved in the envelope analysis is that each time a localized defect makes contact under load with another bearing component, an impulse with a very short duration compared to that of the interval between pulses vibration is generated. So, its energy is distributed across a wide range of frequencies and many resonances of the system are excited. The frequencies of occurrence of impulses are referred to the bearing characteristic frequencies.

This impulsive excitation is normally repetitive in bearings, so the energy is concentrated in a narrow band, which makes easier the detection of faults.

The envelope spectrum is obtained through the High Frequency Resonance Technique (HFRT), which involves three steps:

- band-passing or high passing a measured signal with a chosen resonant frequency
- demodulating the filtered signal with a full-wave rectifier
- low-pass filter application to cancel high-frequency components and retain the low-frequency information associated with bearing defect

From the application of those filters centered at structural resonances, defect-induced vibration can be separated from the background noise; the demodulation operation extracts the defect-induced vibration signal and the use of frequency domain clearly identifies the repetitive occurrence of defects.

7.3.3 Fault signatures and characteristic frequencies

When a single component of a bearing is defective because of one of the causes mentioned in Section 7.1, it easy to identify the fault signature, since each time a rolling element rolls over the defect, an impulsive force occurs causing the whole bearing system to vibrate. Thus, the bearing starts vibrating at its natural frequency and the response rapidly decays due to internal damping. The fundamental frequency of the response waveforms is the rate at which the rolling elements roll over the defect. For that reason, each type of defect is characterized by its own frequency that depends on the components on which the defect occurs, the bearing geometry and the rings rotational speeds.

Fundamental characteristic frequencies generated in rolling bearings cover a wide range of frequencies and can interact giving a complex signal, that is further complicated by the presence of other sources of mechanical, structural or electromechanical vibration of the system.

The characteristic frequencies are the following ones:

- FTFI: fundamental train frequency relative to inner raceway
- FTFO: fundamental train frequency relative to outer raceway
- FTF: fundamental train frequency
- BPFO: ball pass frequency on an outer race defect
- BPFI: ball pass frequency on an inner race defect
- BSF: ball spin frequency

and they are defined as follows:

$$FTFI = \frac{1}{2} f_{IR}(1-\gamma) \tag{7.18}$$

$$FTFO = \frac{1}{2} f_{OR}(1+\gamma) \tag{7.19}$$

$$FTF = \frac{1}{2}(FTFI + FTFO) = \frac{1}{2}\left[f_{IR}(1-\gamma) + f_{OR}(1+\gamma)\right] \tag{7.20}$$

$$BPFO = \frac{Z}{2}(f_{IR} - f_{OR})\,(1-\gamma) \tag{7.21}$$

$$BPFI = \frac{Z}{2}\,(f_{IR} - f_{OR})\,(1+\gamma) \tag{7.22}$$

$$BSF = \frac{d_m}{2\,D_r}\,(f_{IR} - f_{OR})\,(1-\gamma) \tag{7.23}$$

where:

$$f_{IR} = \frac{\omega_{IR}}{2\pi} \tag{7.24}$$

$$f_{OR} = \frac{\omega_{OR}}{2\pi} \tag{7.25}$$

In case of stationary outer ring and rotating inner ring:

$$FTF = \frac{1}{2} f_{IR}(1-\gamma) \tag{7.26}$$

$$BPFO = \frac{Z}{2} f_{IR}(1-\gamma) \tag{7.27}$$

$$BPFI = \frac{Z}{2} f_{IR}\,(1+\gamma) \tag{7.28}$$

$$BSF = \frac{d_m}{2\,D_r} f_{IR}(1-\gamma) \tag{7.29}$$

The frequency equations assume that there is no sliding in the rolling element movement, but in practice its motion is a combination of rolling and sliding. Thus, the real characteristic defect frequencies may differ slightly from those predicted, but this is also very dependent on the type of bearing, operation conditions and fits.

It is very important when comparing vibration signature that data are obtained at identical shaft rotational speeds, since, as it can be seen in above mentioned formulas, vibration frequencies are dependent on velocity, so both frequencies and amplitudes may differ a lot for different working conditions. [1]

Imperfections on the raceway surface and rolling element interact to produce other discrete frequencies and sidebands that are summarized in

Surface defect		Frequency
Component	**Imperfection**	
Inner raceway	Eccentricity Waviness Discrete defect	f_{IR} $n\,Z\,FTFI \pm f_{IR}$ $n\,Z\,FTFI \pm f_{IR}$
Outer raceway	Eccentricity Discrete defect	$n\,Z\,FTFO$ $n\,Z\,FTFO \pm f_{IR}$ $n\,Z\,FTFO \pm FTFO$
Rolling element	Diameter variation Waviness Discrete defect	$Z\,FTFO$ $2n\,BSF + FTFO$ $2n\,BSF + FTFO$

Table 7.1

Table 7.1: Frequencies related to surface imperfections [1]

7.3.4 Raceway defects

A discrete defect on the inner raceway will generate a series of high energy pulses at a rate equal to the ball pass frequency relative to the inner raceway. The amplitude of pulses varies, because, since the inner ring rotates, the defect enters and exits the load zone, so the signal is modulated in amplitude at the inner ring rotational frequency. In the frequency domain this gives rise to a discrete peak at the $BPFI$ and a pair of sidebands equally spaced by f_{IR} at the two side of the main frequency. A discrete fault on the outer race will generate a series of high energy pulses at a rate equal to the ball pass frequency relative to the outer raceway $BPFO$ and, since the outer ring is fixed, only that frequency and its multiple are stressed.

Figure 7.3 and Figure 7.4 show the frequency-domain (obtained through the Fast Fourier Transform method) and squared envelope plots of a simulation of 60 s with a radial load $F_r = 62.4$ kN, an inner ring rotational velocity $n_{IR} = 120.6405$ rpm and with the presence of a defect on the inner (former plot) or on the outer raceway (the latter).

Vertical lines highlight the multiple of the characteristic frequencies $BPFI$ and $BPFO$ respectively. In case of the presence of a defect on the inner raceway, sidebands spaced by a length equal to the inner race rotational frequency f_{IR}.

In this case

$$f_{IR} = \frac{n_{IR}}{60} = 2.01 \text{ Hz} \tag{7.30}$$

and the distance between the highlighted central and the first sideband frequency is

$$\Delta f = 2.02\ Hz \tag{7.31}$$

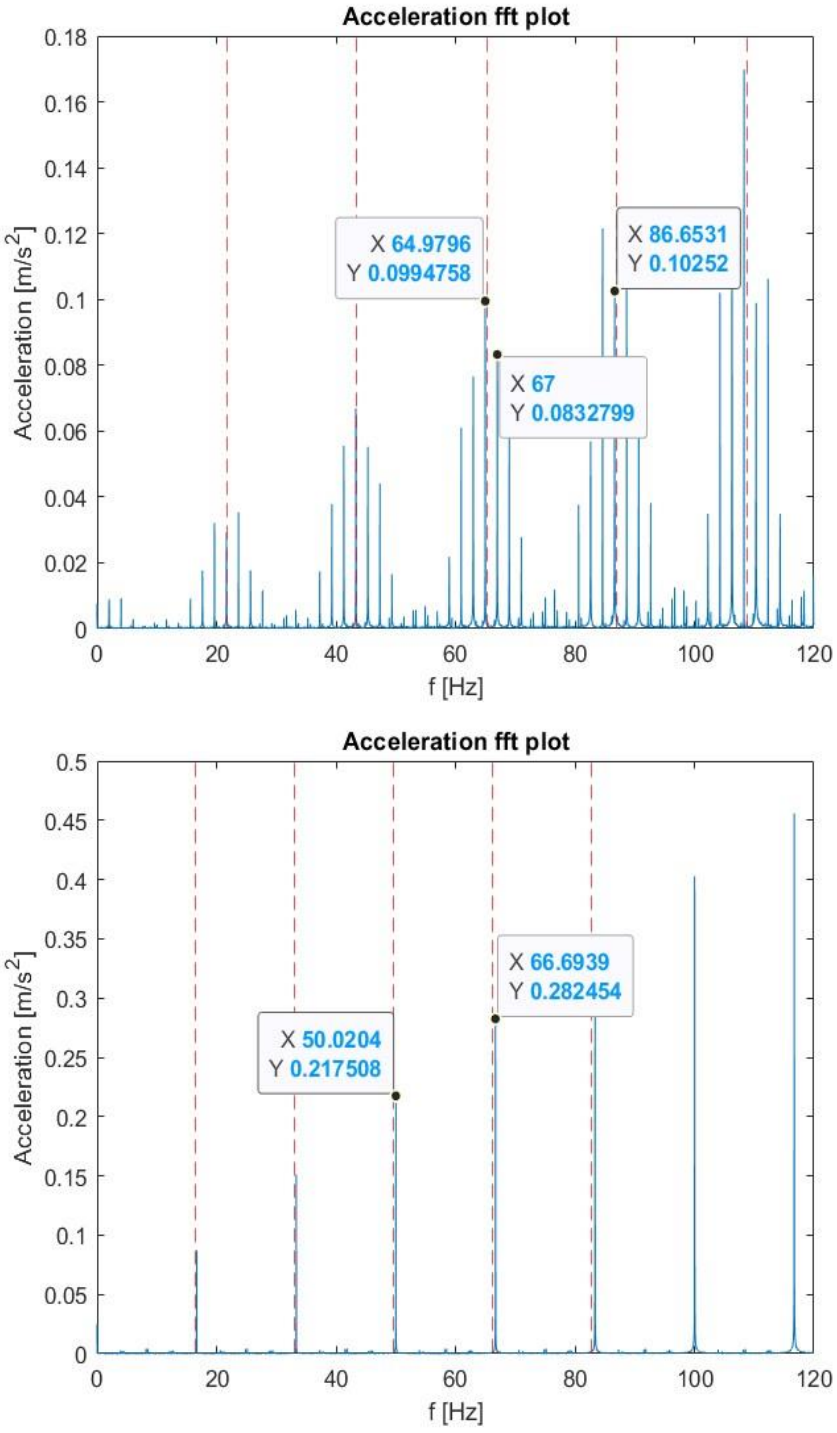

Figure 7.3: Frequency-domain vibration plots of defective bearing on the inner race (above) and outer race (below)

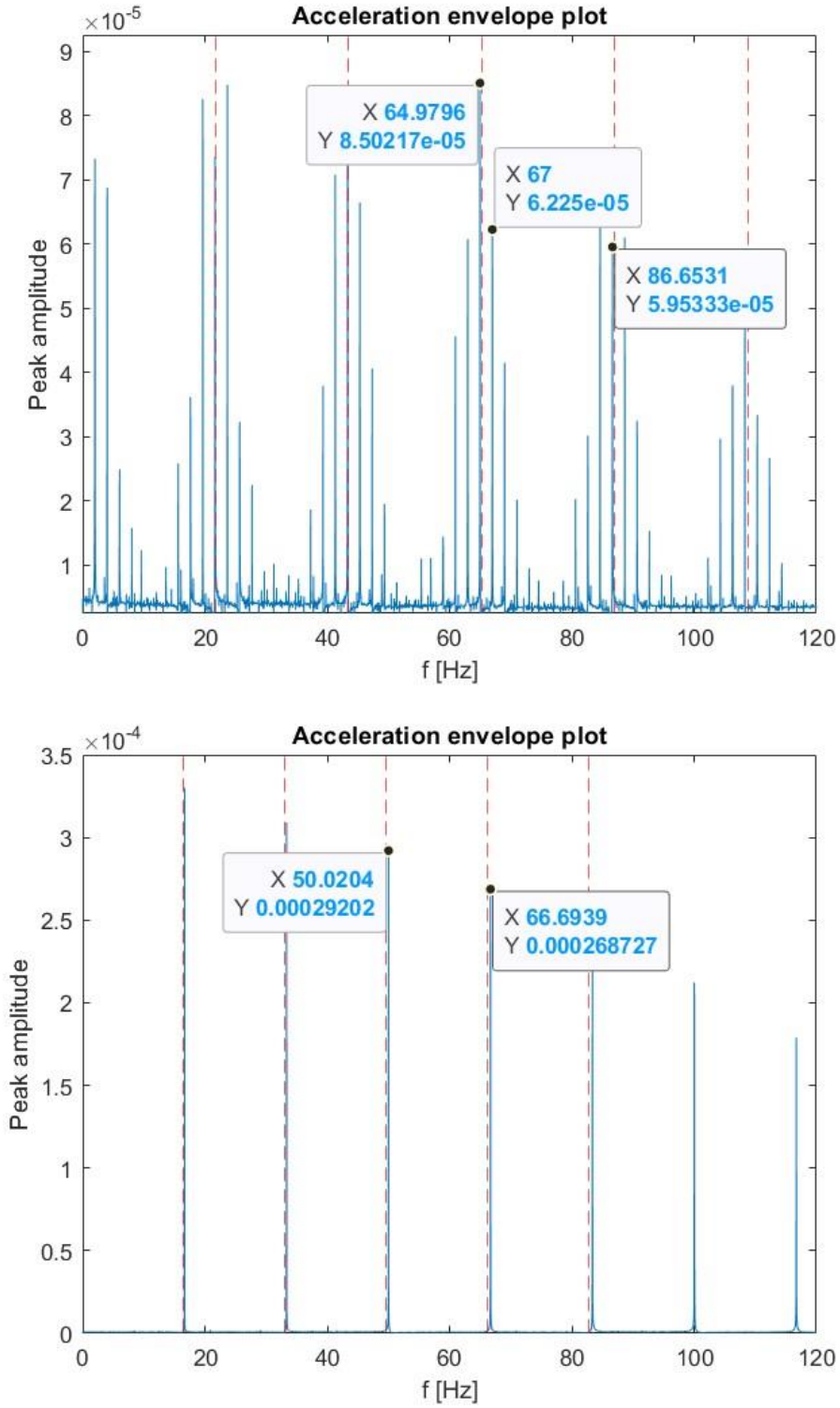

Figure 7.4: Envelope plots of defective bearing on the inner race (above) and outer race (below)

7.4 INTELLIGENT FAULT DIAGNOSIS (IFD)

Intelligent Fault Diagnosis (IFD) refers to the application of machine learning theories to machine fault diagnosis. In the past, traditional machine learning algorithms began to reduce the human intervention in machine fault monitoring: form the debut in the 1950s to the 2010s Artificial Neural Networks (ANN), Support Vector Machine (SVM), k-Nearest Neighbor (kNN) have been the most used algorithms. The health status of a machine could be predicted extracting sensitive features from vibration data. From the 2010s deep learning theories reformed IFD: traditional algorithms could recognise the health status of machines, but feature extraction still depended on human labor and only small databases could be handled; instead, deep learning approaches, which include deep belief networks (DBP) and convolutional neural networks (CNN), can automatically learn fault features from collected data. Thus, these models are expected to directly connect the raw monitoring data to the corresponding health status of machines, further reducing the contribution of human labor. [4]

In this section there is a review of SVM and kNN algorithms that have been used in this work to perform fault diagnosis.

7.4.1 Support Vector Machine (SVM)

Support Vector Machine is an AI algorithm that can handle very large feature spaces, because the dimension of the classified vectors used for the training phase does not have an influence on the performance as it happens for conventional classifiers. Thus, SVM is a good approach for fault diagnosis, since vibration features may not have to be limited. SVM has also good generalization properties compared to conventional classifiers, because in the training phase the structural misclassification risk is minimized, whereas traditional algorithms minimise the empirical risk.

The Structural risk minimization (SRM) is an inductive principle of use in machine learning introduced by Vladimir Vapnik and Alexey Chervonenkis [59] and is implemented in the training phase by the minimization of

$$E_{train} + \beta\, H(W) \tag{7.32}$$

where E_{train} is the training error, e.g. the mean square between the value of the learned model and the label of the data vector, and $H(W)$ is a regularization function that controls the trade off between minimizing the training error and minimizing the expected gap between the training and test errors. The cost function of the algorithm can be written as:

$$J(w) = \frac{1}{2n}\sum_{i=1}^{n}(h_w(x_i) - y_i)^2 + \frac{\lambda}{2}\sum_{j=1}^{d} w_j^2 \tag{7.33}$$

On the other hand, the Empirical Risk Minimization (ERM) is a principle that is used to give theoretical bounds to machine learning algorithms performance. The core idea behind ERM is that we cannot know how well an algorithm can work in practice using new data, because their distribution is unknown, but its performance can be evaluated on a known set of training data.

Thus, the empirical risk is computed by averaging the loss function of the training set:

$$R_{emp}(h) = \frac{1}{n}\sum_{i=1}^{n} L(h(x_i), y_i) \tag{7.34}$$

The learning algorithm chooses the hypothesis $\hat{h}$ that minimizes it:

$$\hat{h} = \arg_{h \in H} \min R_{emp}(h) \tag{7.35}$$

Given a set of data input

$$\mathbf{x_i} \quad i = 1, 2, \ldots, N$$

where N is the number of samples, which have labels indicating their class, in the linear case (Figure 7.5) SVM theory supposes that there exists a hyperplane with function

$$\mathbf{w}^{\mathrm{T}}\,\mathbf{x} + b = 0 \tag{7.36}$$

which implies $y_i(\mathbf{w}^T\,\mathbf{x} + b = 0) \geq 1, \quad i = 1, \ldots, N$

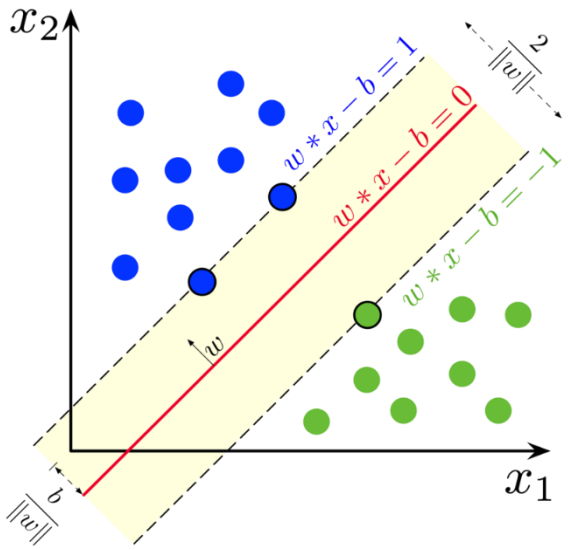

Figure 7.5: Plot of maximum-margin hyperplane and margins for a linear SVM with two classes

Taking into account the possible presence of noise in data with the slack variable ζ_i and the error penalty constant β, the optimal hyperplane that separates data can be obtained as a solution to the following optimization problem:

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\|\boldsymbol{w}\|^2 + \beta\sum_{i=1}^{N}\zeta_i \\ &\text{subject to} \quad \begin{cases} y_i(\mathbf{w}^{\mathbf{T}}\,\mathbf{x_i} + b = 0) \geq 1 - \zeta_i, & i = 1, \ldots, N \\ \zeta_i \geq 0, & i = 1, \ldots, N \end{cases} \end{aligned} \tag{7.37}$$

Replacing the weight vector $\mathbf{w}$ with

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x_i} \tag{7.38}$$

the optimization problem can be written in the form

$$f(x) = \text{sign}\left(\sum_{i,j=1}^{N} \alpha_i y_i (\mathbf{x_i}\, \mathbf{x_j}) + b \right) \tag{7.39}$$

SVM can also be used in nonlinear classification problems applying kernel functions, that map the n-dimensional input vector $\mathbf{x}$ into a l-dimensional feature space to make a linear classification feasible.

Using the non-linear vector function

$$\mathbf{\Phi}(\mathbf{x}) = \big(\phi_1(\mathbf{x}), \dots, \phi_l(\mathbf{x})\big) \tag{7.40}$$

the decision function can be written in the form:

$$f(x) = \text{sign}\left(\sum_{i,j=1}^{N} \alpha_i y_i (\mathbf{\Phi^T}(\mathbf{x_i})\, \mathbf{\Phi}\big(\mathbf{x_j}\big)) + b \right) \tag{7.41}$$

The kernel function returns the dot product of the feature space mapping from the original data points:

$$K(\mathbf{x_i}, \mathbf{x_i}) = \Big(\mathbf{\Phi^T}(\mathbf{x_i})\, \mathbf{\Phi}\big(\mathbf{x_j}\big)\Big) \tag{7.42}$$

There are many types of kernel functions, such as linear, polynomial and Gaussian RBF (radial basis function). The selection of the most suitable kernel is crucial, because it defines the feature space in which the training set data will be classified. [60] [61]

7.4.2 k-Nearest Neighbors (kNN)

k-Nearest Neighbors algorithm is a non-parametric supervised learning methos developed by Fix and Hodges [62] in which the input is made of the k closest training samples of a dataset and the output for a classification problem is a class membership: the object is assigned to the class most common among its k nearest neighbors.

Given a set of observations

$$D = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\} \tag{7.43}$$

where $\mathbf{x}_i$ is the feature vector and y_i is the corresponding class label and $(\mathbf{x}_i, y_i)$ is assumed to be i.i.d. from some unknown distribution P of $(\mathbf{x}, y)$ on $R^d \times \{\omega_1, \dots, \omega_M\}$, the goal is the design of a function $\phi_n : R^d \to \{\omega_1, \dots, \omega_M\}$ that maps a feature vector $\mathbf{x}$ into its desired class from $\{\omega_1, \dots, \omega_M\}$.

The performance of the classifier is evaluated through the probability error

$$L(\phi_n) = P\{(\mathbf{x}, y): \phi_n(\mathbf{x}) \neq y\} \tag{7.44}$$

If the underlying distribution is known, then the optimal decision rule that minimizes the probability of error is the bayes decision rule:

$$\phi^*(\mathbf{x}) = \arg\max P(y|\,\mathbf{x}) \tag{7.45}$$

It can be shown that at any given point $\mathbf{x}$ the probability that its nearest neighbor $\mathbf{x}'$ belongs to class ω_i converges to the corresponding a posteriori probability $P(\omega_i|\mathbf{x}')$ as the number of reference observations goes to infinity and it was shown by Cover and Hart [63] that under some continuity conditions on the underlying distribution the asymptotic probability of error L_{NN} for a multi-class kNN classifier is bounded by

$$L^* \leq L_{NN} \leq L^* \left(2 - \frac{M}{M-1} L^*\right) \tag{7.46}$$

where L^* is the optimal Bayes probability of error and M is the number of classes. Thus, the nearest-neighbor rule is asymptotically optimal when the classes are not overlapping. [64]

Practically, in the classification phase, after the constant k has been defined by the user, an unlabelled vector is classified by assigning the label which is most frequent among the k training samples that are nearest to that query point. The distance metric is usually represented by the Euclidean distance, Hamming distance or correlation coefficients. In the example shown in Figure 7.6, the test sample can be classified as a blue square or a red triangle: if $k = 3$ is selected, it is assigned to red triangles because the nearest points are represented by two triangles and one square, whereas if $k = 5$ it is assigned to blue squares.

A drawback of this algorithm occurs when the class distribution is skewed: a class tends to dominate the prediction of new samples, because its examples tend to be common among the k nearest neighbors due to their large number. This issue can be avoided introducing a weight proportional to the inverse of the distance from the point that has to be labelled to the examples.

The value of k deeply affects the classification performance [65]: larger values reduce the effects of noise but make less distinct boundaries between classes. Accuracy is also affected by the presence of noisy or irrelevant features, or if the feature scales are not consistent with their importance. For that reason, feature selection and scaling algorithms play an important role for kNN and the most used methods are principal component analysis (PCA), linear discriminant analysis (LDA) and canonical correlation analysis (CCA) [66]. Finally, data-reduction is one of the biggest problems when working with huge datasets, because only some of data point are usually needed for accurate classification and those are called prototypes: after the recognition of the outliers, the dataset is divided into the prototypes that are used for classification decisions and the absorbed point that can be correctly classified using prototypes and can be removed from the training set.

An algorithm for dataset reduction is the Hart algorithm [67], that is applied to 1NN and selects the prototypes U such that the classification algorithm with U has the same accuracy as if it uses the whole set of data.

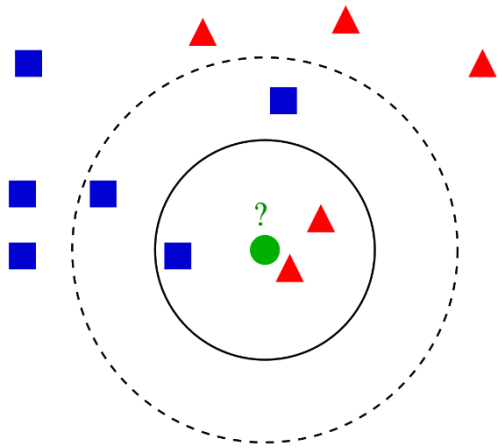

Figure 7.6: Example of kNN classification

7.4.3 Features selection and Shapley values

As discussed in Section 7.4.3, the use of a large set of features extracted from vibration signals can reduce the machine learning algorithm accuracy due to information redundancy, increasing at the same time complexity and computational effort. Therefore, the extraction of useful information from the high-dimensional feature set becomes crucial. For that reason, many dimensionality reduction techniques, also called data fusion techniques, have been developed and the principal ones are the filter-based, wrapper-based and embedded methods.

The filter-type feature selection algorithm measures the feature importance focusing on the characteristics of each of them, such as variance and relevance to the response; the filter feature selection is uncorrelated to the training algorithm, since it is a part of the data pre-processing step.

The wrapper-type feature selection algorithm takes place in the training phase using a subset of features and then adds or removes one of them using a selection criterion, based on the model performance that results from adding or removing a selected feature. The training phase is repeated until the stopping criterion is satisfied.

The embedded-type feature selection algorithm learns feature importance as part of the model learning process: during the training phase the importance of each feature in the trained model is computed, so that the most relevant ones can be chosen.

These are some of the mostly used dimensionality reduction techniques:

- *Principal Component Analysis* (PCA) is a very popular traditional linear dimensionality reduction technique that selects the features that represents the maximum variance in input data

- *Linear Discriminant Analysis* (LDA) determines the projection direction maximizing the distance between the data clusters used for class discrimination
- *Locally Linear Embedding* (LLE) [68] is an unsupervised learning algorithm that tries to find nonlinear structures in high-dimensional data exploiting the local symmetry of linear reconstructions
- *ISOMAP* extracts significant features from observed data by retaining local neighborhood information. [69]

In this work the Shapley values were used to highlight the most relevant features in the defect type classification of the spherical roller bearing under investigation. Shapley values [70] is a method invented to assign pay-outs to players depending on their contribution towards the total playout but can be effectively used to measure the global importance of a feature.

Using the game theory, Shapley values can be explained as follows: given a cooperative game with M players aiming at maximizing a payoff and letting $\mathrm{S} \subseteq \mathrm{M} = \{1, \dots, M\}$ be a subset of $|\mathrm{S}|$ players, the contribution function $v(\mathrm{S})$, which maps a subset of players to the real number, called the contribution of coalition S, describes the total expected sum of payoffs the member of S can obtain by cooperation. Thus, Shapley values are a method to distribute the total gains to players and the amount that player j earns is [71]

$$\phi_j(v) = \phi_j = \sum_{\mathrm{S} \subseteq \mathrm{M} \setminus \{\mathrm{j}\}} \frac{|\mathrm{S}|!\,(M - |\mathrm{S}| - 1)!}{M!} (v(\mathrm{S} \cup \{\mathrm{j}\}) - \mathrm{v}(\mathrm{S})), \qquad \mathrm{j} = 1, \dots, \mathrm{M} \tag{7.47}$$

that is a weighted mean over contribution functions differences for all subsets S of players not containing player $\{j\}$. The algorithm also accounts for the non-distributed gain $\phi_0 = v(\emptyset)$.

The main properties of Shapley values can be summarized as follows: [72]

- *Efficiency*: the total gain is distributed

$$\sum_{j=0}^{M} \phi_j = v(\mathrm{M}) \tag{7.48}$$

- *Symmetry*: if i and j are two players that contribute equally for all possible coalitions, then their Shapley value is the same

$$v(\mathrm{S} \cup \{\mathrm{i}\}) = v(\mathrm{S} \cup \{\mathrm{j}\}) \rightarrow \phi_i = \phi_j \tag{7.49}$$

- *Dummy player*: if a player does not contribute to any coalitions, then his Shapley value is zero

$$\phi_j = 0 \tag{7.50}$$

- *Monotonicity*: if two games v and v' a player always makes a greater contribution to v than to v' for all S, then the gain for v will be greater than that for v':

$$v(\mathrm{S} \cup \{\mathrm{i}\}) - v(\mathrm{S}) \geq v'^{(\mathrm{S} \cup \{\mathrm{i}\})} - v'^{(\mathrm{S})} \quad \forall \mathrm{S} \qquad \rightarrow \qquad \phi_i(v) \geq \phi_i(v')$$

$$L^* \leq L_{NN} \leq L^* \left(2 - \frac{M}{M-1} L^*\right) \tag{7.51}$$

- *Linearity*: if two coalition games described by the gain functions v and w are combined, the distributed gain is the sum of the two gains. This is valid also for a multiplication by a constant value:

$$\phi_j(v + w) = \phi_j(v) + \phi_j(w) \tag{7.52}$$

$$\phi_j(\alpha\, v) = \alpha\, \phi_j(v) \tag{7.53}$$

Given a training set of N samples $D = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$ used to train a predictive machine learning model $f(\mathbf{x})$ whose outcome should be as close as possible to label y, Shapley values can be used to explain the prediction $y^* = f(\mathbf{x}^*)$ for a specific feature vector $\mathbf{x} = \mathbf{x}^*$.

Adopting the game theory, single predictions take the place of payout and features become players. Thus, the prediction can be written as

$$f(\mathbf{x}^*) = \phi_0 + \sum_{j=1}^{M} \phi_j^* \tag{7.54}$$

where $\phi_0 = \mathrm{E}[f(\mathbf{x})]$ and ϕ_j^* is the ϕ_j for $\mathbf{x} = \mathbf{x}^*$. Therefore, Shapley values explain the difference between the prediction $y^* = f(\mathbf{x}^*)$ and the global average prediction.

For the computation of the Shapley values for a prediction explanation, the contribution $v(\mathrm{S})$ for a certain subset S has to be defined. This function should take values similar to $f(\mathbf{x}^*)$ when only the value of the subset S of these features is known. As cited in Lundberg [73] the expected output of the predictive model, conditional to the feature values $\mathbf{x}_\mathrm{S} = \mathbf{x}_\mathrm{S}^*$ of this subset, can be used:

$$v(\mathrm{S}) = \mathrm{E}[f(\mathbf{x}) | \mathbf{x}_\mathrm{S} = \mathbf{x}_\mathrm{S}^*] \tag{7.55}$$

Thus, the conditional expectation summarizes the whole probability distribution and, if it is considered as the prediction for $f(\mathbf{x}^*)$, it is also the minimizer of the squared error loss function.

In the case of the linear regression model, which can be written in the form

$$f(\mathbf{x}) = \beta_0 + \sum_{j=1}^{M} \beta_j x_j \tag{7.56}$$

and where all the features are independent, the Shapley values are written as:

$$\phi_0 = \beta_0 + \sum_{j=1}^{M} \beta_j E[x_j] \tag{7.57}$$

$$\phi_j = \beta_j (x_j^* - E[x_j]) \tag{7.58}$$

In case of dependent features and nonlinear predictive models, no explicit formula exists. If the feature vector has got M elements the number of possible sets involved in computation are 2^M, so the exact solution to the problem becomes computationally infeasible for large values of M. [71]

There are two main algorithms for the computation of the Shapley values and the difference among them is in the definition of the value function. Both of them define the value function such that the sum of the Shapley values of a query point over all features corresponds to the total deviation of the prediction for the query point from average:

$$\sum_{j=1}^{M} \phi_j(v_x) = f(\mathbf{x}) - \mathrm{E}[f(\mathbf{x})] \tag{7.59}$$

- *KernelSHAP* [73]
  the KernelSHAP algorithm defines the value function of features in S at the query point **x** as the expected prediction with respect to the interventional distribution D, which is the joint distribution of feature in $\mathrm{S^C}$

  $$v_x(\mathrm{S}) = \mathrm{E_D}\left[f\left(x_\mathrm{S}, X_{\mathrm{S^C}}\right)\right] \tag{7.60}$$

  where x_S is the query point value for features in S and $X_{\mathrm{S^C}}$ are the features in $\mathrm{S^C}$.
  If we assume that features are not highly correlated, the interventional distribution can be computed through

  $$v_x(\mathrm{S}) = \mathrm{E_D}\left[f\left(x_\mathrm{S}, X_{\mathrm{S^C}}\right)\right] \approx \frac{1}{N}\sum_{j=1}^{N} f\left(x_\mathrm{S}, \left(X_{\mathrm{S^C}}\right)_j\right) \tag{7.61}$$

  where N is the number of observations and $\left(X_{\mathrm{S^C}}\right)_j$ contains the values of the features in $\mathrm{S^C}$ for the j-th observation

- *Extension to KernelSHAP* [71]
  the extension to KernelSHAP algorithm defines the value function of the feature in S at the query point **x** using the conditional distribution of $X_{\mathrm{S^C}}$, given that $\mathrm{X_S}$ is the query point values:

  $$v_x(\mathrm{S}) = \mathrm{E}_{X_{\mathrm{S^C}}|\mathrm{X_S}}\left[f\left(x_\mathrm{S}, X_{\mathrm{S^C}}\right)\right] \tag{7.62}$$

  To evaluate the value function $v_x(\mathrm{S})$ at point x, the algorithm uses the nearest neighbors of that query point; thus, it does not require the assumption of feature independence. On the other hand, this algorithm is computationally more expensive, does not support categorical predictors and cannot handle missing values in the feature vectors X.

The main difference between the two approaches is in the way the feature contribution to the final prediction is evaluated: interventional methods (the latter) attribute an influence on x_j only if the value function, computed considering that element, is significantly different if it was computed without it; on the contrary, the conditional approach may attribute influence to feature with no interventional effect, only because of their presence in the model. Thus, this kind of approach requires further modelling of how features are correlated. [74]

8 AI ALGORITHM SETUP AND EVALUATION

This chapter presents the path followed to setup the machine learning algorithm aimed at recognizing the presence of defects on the bearing raceways. The main goal of this section is recognizing whether numerical vibration paths can be used to train an AI algorithm capable to perform fault detection on experimental data obtained from an operating machine.

The first step is represented by the creation of an experimental vibration database, using the test rig presented in Chapter 6, and a numeric one, relying on the nonlinear multi-body dynamic model presented in Chapter 4, under the same external radial load, shaft rotation speed, mean temperature and defect conditions.

In the second phase vibration paths are partitioned in shorter segments to get a bigger dataset to train the AI algorithm. Then the main time and frequency domain features are extracted and the database of features linked to each category of fault (non-defective, inner race fault, outer race fault) is created.

The third step is represented by the setup of the AI algorithm: training and validation data are selected randomly from the numerical database while test data are selected among the experimental data. Then SVM and kNN machine learning algorithms are trained and their performance evaluated.

The last part of this chapter is focused on the selection of the most important features applying the Shapley values method.

8.1 DATABASE GENERATION

Firstly, two database of vibration paths, the former with experimental data obtained from the bearing test rig and the latter with numerical data computed through the nonlinear multi-body dynamic model of the bearing in the same conditions of external load, rotation speed, mean temperature and defect conditions. Therefore, numerical data were supposed to have the same characteristics of the experimental ones.

Experimental data were obtained imposing different working conditions that are resumed in Table 8.1. As it can be seen from the table three load cases, ten nominal rotational speeds and three defect types were used, so the whole dataset contains ninety vibration paths with a duration of 60 s.

Since the nominal rotational speed refers to the three-phase motor synchronous velocity, the real shaft rotation speed is lower due to the presence of friction in the whole mechanical. The difference between the two rotation speed is accounted through the slip parameter. Thus, a tachometer in the test rig is necessary to register the true shaft rotation speed.

Simulated data were obtained using the operating conditions reported in Table 8.1, but imposing for each case the shaft rotational speed measured through the tachometer and the mean operating temperature of the related experimental case. The duration of numerical simulations was set equal to that of experimental paths.

	CASE 1	CASE 2	CASE 3
Radial load [kN]	0	62.4	124.8
Nominal speed [rpm]	127, 227, 353, 457, 523, 607, 727, 877, 937, 997		
Defect type	non defective (0), inner race defect (1), outer race defect (2)		

Table 8.1: Test rig operating conditions

8.2 FEATURES EXTRACTION

The feature extraction activity allows the creation of a database of labelled samples that can be used to train, validate and test machine learning algorithms.

To get more samples, each experimental and numerical vibration path was divided into shorter segments: the developed Matlab algorithm allowed the creation of 200 ms, 400 ms or 800 ms segments from original data.

As it can be seen from Figure 8.1, experimental data are corrupted by noise, coming from the other mechanical and electrical components linked to the system. To simulate this phenomenon, white Gaussian noise was added to each numerical data segment. Three different cases of signal-to-noise ratio (SNR) were considered: 0, 10, −3 and −10 dB (Figure 8.2).

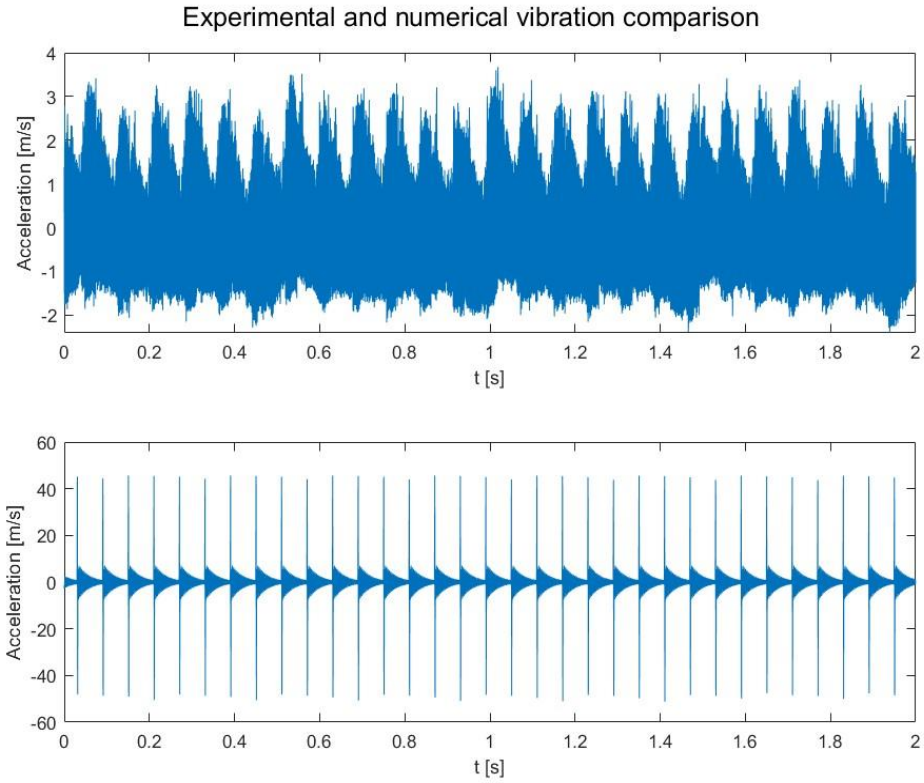

Figure 8.1: Comparison of experimental (top) and numerical (bottom) vibration paths for operating conditions of 127 rpm *and 62.4* kN

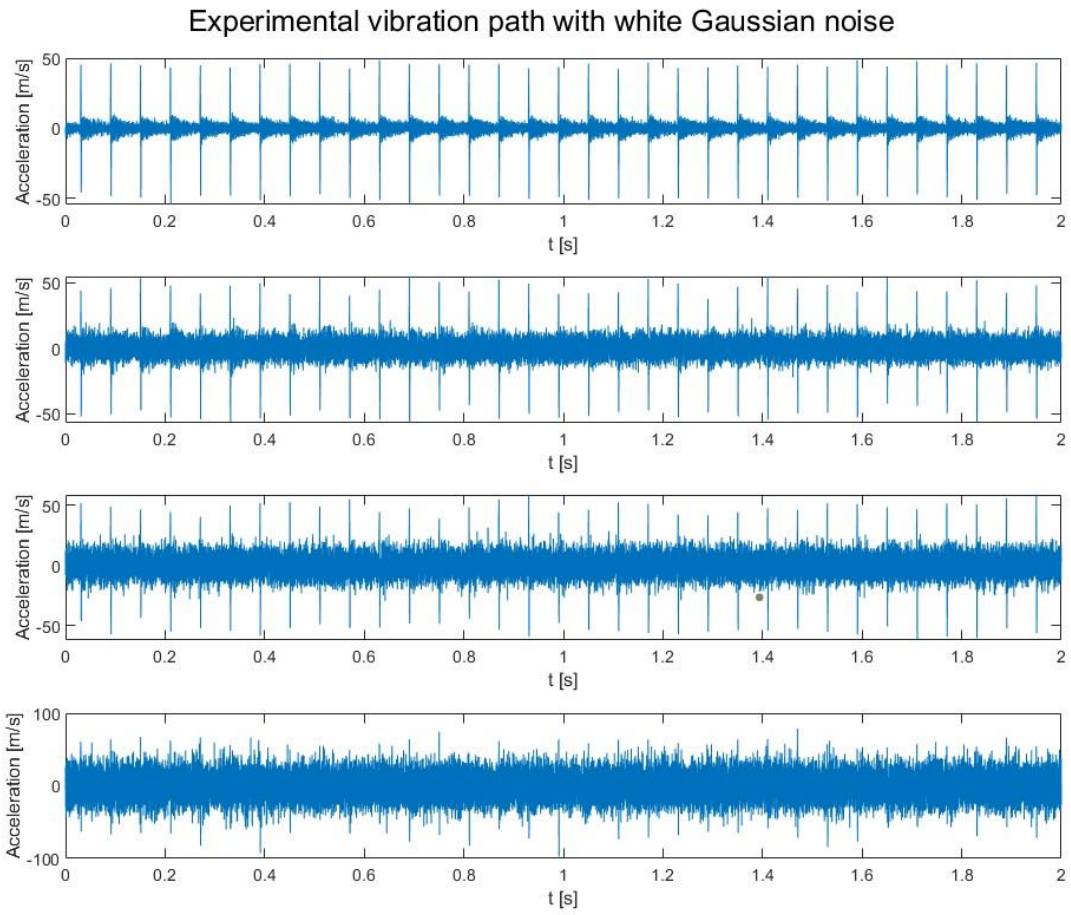

Figure 8.2: Experimental vibration path added with noise with SNR of 10 dB *(first), 0* dB *(second), -3* dB *(third) and -10* dB *(fourth plot) for operating conditions of 127* rpm *and 62.4* kN *load*

At this point vibration data segments were ready to feature extraction. In this work twenty-three features, taken from Lei et al. [75], were selected: the first eleven parameters belong to the time-domain, while the other to the frequency domain. The extraction of the frequency-domain features was performed after the computation of the fast Fourier Transform algorithm on the time-domain signals. All the selected features for this work are listed in Table 8.2.

After feature extraction activity a sample with twenty-three features and a label depending on the defect type was generated from each segment: label 0 refers to non-defective bearing conditions, while 1 and 2 are related to a localized damage on the inner and on the outer raceway respectively.

Time-domain features		Frequency-domain features	
$\boldsymbol{p_1 = \frac{1}{N}\sum_{i=1}^{N} x(i)}$	Mean	$p_{12} = \frac{1}{K}\sum_{k=1}^{K} s(k)$	Mean frequency
$\boldsymbol{p_2 = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x(i) - p_1)^2}}$	STD	$p_{13} = \frac{1}{K-1}\sum_{k=1}^{K}(s(k) - p_{12})^2$	
$\boldsymbol{p_3 = \left(\frac{1}{N}\sum_{i=1}^{N}\sqrt{\lvert x(i)\rvert}\right)^2}$	SMRA	$p_{14} = \sum_{k=1}^{K}\frac{(s(k) - p_{12})^3}{K\left(\sqrt{p_{13}}\right)^3}$	
$\boldsymbol{p_4 = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x(i))^2}}$	RMS	$p_{15} = \sum_{k=1}^{K}\frac{(s(k) - p_{12})^4}{K\, p_{13}^4}$	
$\boldsymbol{p_5 = \max\lvert x(i)\rvert}$	Peak	$p_{16} = \frac{\sum_{k=1}^{K} f_k\, s(k)}{\sum_{k=1}^{K} s(k)}$	Frequency center
$\boldsymbol{p_6 = \sum_{i=1}^{N}\frac{(x(i) - p_1)^3}{(N-1)\, p_2^3}}$	Skewness	$p_{17} = \sqrt{\frac{\sum_{k=1}^{K}(f_k - p_{16})^2\, s(k)}{K}}$	STDF
$\boldsymbol{p_7 = \sum_{i=1}^{N}\frac{(x(i) - p_1)^4}{(N-1)\, p_2^4}}$	Kurtosis	$p_{18} = \sqrt{\frac{\sum_{k=1}^{K} f_k^2\, s(k)}{\sum_{k=1}^{K} s(k)}}$	RMSF
$\boldsymbol{p_8 = \frac{p_5}{p_4}}$	CF	$p_{19} = \sqrt{\frac{\sum_{k=1}^{K} f_k^4\, s(k)}{\sum_{k=1}^{K} f_k^2\, s(k)}}$	
$\boldsymbol{p_9 = \frac{p_5}{p_3}}$	CLF	$p_{20} = \frac{\sum_{k=1}^{K} f_k^2\, s(k)}{\sqrt{\sum_{k=1}^{K} s(k)\sum_{k=1}^{K} f_k^4\, s(k)}}$	
$\boldsymbol{p_{10} = \frac{p_4}{\frac{1}{N}\sum_{i=1}^{N}\lvert x(i)\rvert}}$	SF	$p_{21} = \frac{p_{17}}{p_{16}}$	
$\boldsymbol{p_{11} = \frac{p_5}{\frac{1}{N}\sum_{i=1}^{N}\lvert x(i)\rvert}}$	IF	$p_{22} = \sum_{k=1}^{K}\frac{(f_k - p_{16})^3\, s(k)}{K\, p_{17}^3}$	
		$p_{23} = \sum_{k=1}^{K}\frac{(f_k - p_{16})^4\, s(k)}{K\, p_{17}^4}$	

Table 8.2: Table of features parameters

8.3 AI ALGORITHM DEVELOPMENT

From the data preparation procedure presented in the section above, nine databases were generated, three from experimental data, which differ from the segmentation time interval (200, 400, 800 ms), and other six from numerical data, that have different duration (200, 400, 800 ms) and SNR (0 and 10 dB) of the added white Gaussian noise. Due to the huge number or samples in each database, it was decided to use only a part of them for the machine learning algorithm elaboration.

8.3.1 Training and testing

For the training and testing phases of the AI algorithm $N = 2000$ labelled samples were selected randomly from the original numerical database and split in a ratio $80 - 20$ % for the two phases respectively.

The evaluation of the best performing algorithm was conducted using the MATLAB Classification Learner App. The machine learning algorithms used in this work and their principal characteristics are resumed in Table 8.3 and in Table 8.4.

Algorithm	Number of neighbors	Distance metric	Distance weight
Fine kNN	1	Euclidean	equal
Medium kNN	10	Euclidean	equal
Coarse kNN	100	Euclidean	equal
Cosine kNN	10	cosine	equal
Cubic kNN	10	Minkowski	equal
Weighted kNN	10	Euclidean	squared inverse

Table 8.3: kNN algorithms characteristics

Algorithm	Kernel function	Kernel scale
Linear SVM	linear	/
Quadratic SVM	quadratic	/
Cubic SVM	cubic	/
Fine Gaussian SVM	Gaussian	1.2
Medium Gaussian SVM	Gaussian	4.8
Coarse Gaussian SVM	Gaussian	19

Table 8.4: SVM algorithms characteristics

Firstly, the above mentioned algorithms were trained and tested using numerical data: 1600 random labelled samples were used for training and 400 for testing. All the algorithms were evaluated using samples obtained from 200, 400 and 800 ms numerical vibration path segments corrupted by white Gaussian noise with SNR equal to 10, 0, −3 and −10 dB. Their accuracy on testing data is reported in Figure 8.3. In computation 5-fold cross validation was adopted and all the $D = 23$ features were employed.

From the bar graphs it can be seen that the overall accuracy of all the AI algorithms is very high, as expected, and tends to become higher if longer vibration path segments are used. Only in the last case, in which white Gaussian noise SNR $= 10$ dB, the algorithms accuracy is lower, because of the stronger corruption of vibration data due to added noise.

These results led to the consideration that AI algorithm can effectively detect the presence and the type of defect in rolling element bearings, so supervised learning algorithms like SVM and kNN can be effectively used in IFD.

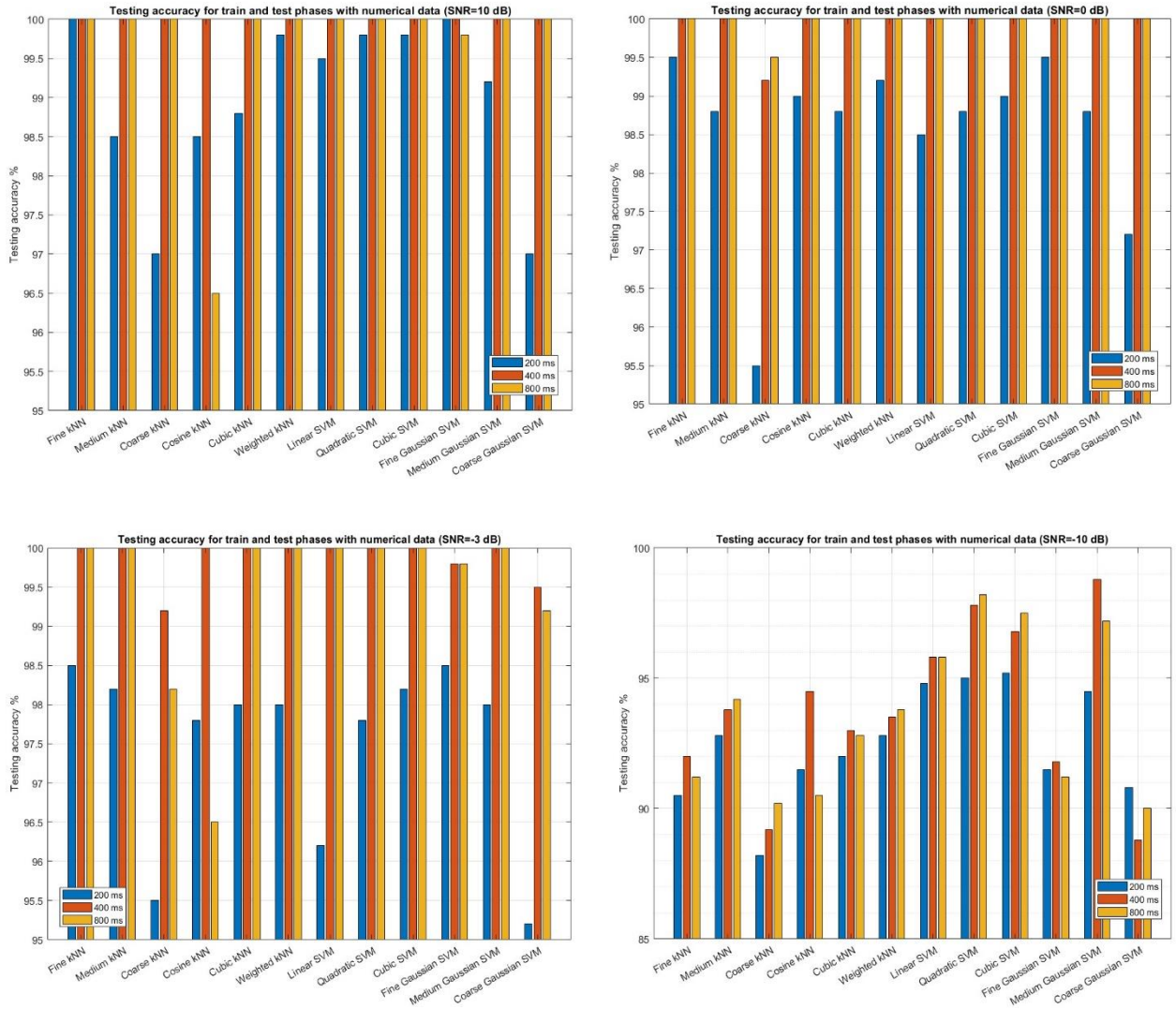

Figure 8.3: Testing accuracy for training and testing phases with numerical data and different white Gaussian noise SNR: 10 dB *(top left), 0* dB *(top right), -3* dB *(bottom left) and -10* dB *(bottom right)*

Then all the machine learning algorithms were trained and tested using only samples coming from experimental vibration paths. Also in this case 1600 samples were used for the training phase and 400 for the testing one. The testing accuracy of each AI algorithm, computed with samples obtained from different time-duration segments (200, 400 and 800 ms), is reported in Figure 8.4. It is evident that all kNN algorithms work quite well, but coarse SVM and kNN and linear SVM have a lower performance. The result about linear SVM machine was expected because the high nonlinearity of bearing model is reflected in the nonlinearity of features and so of the boundaries between classes.

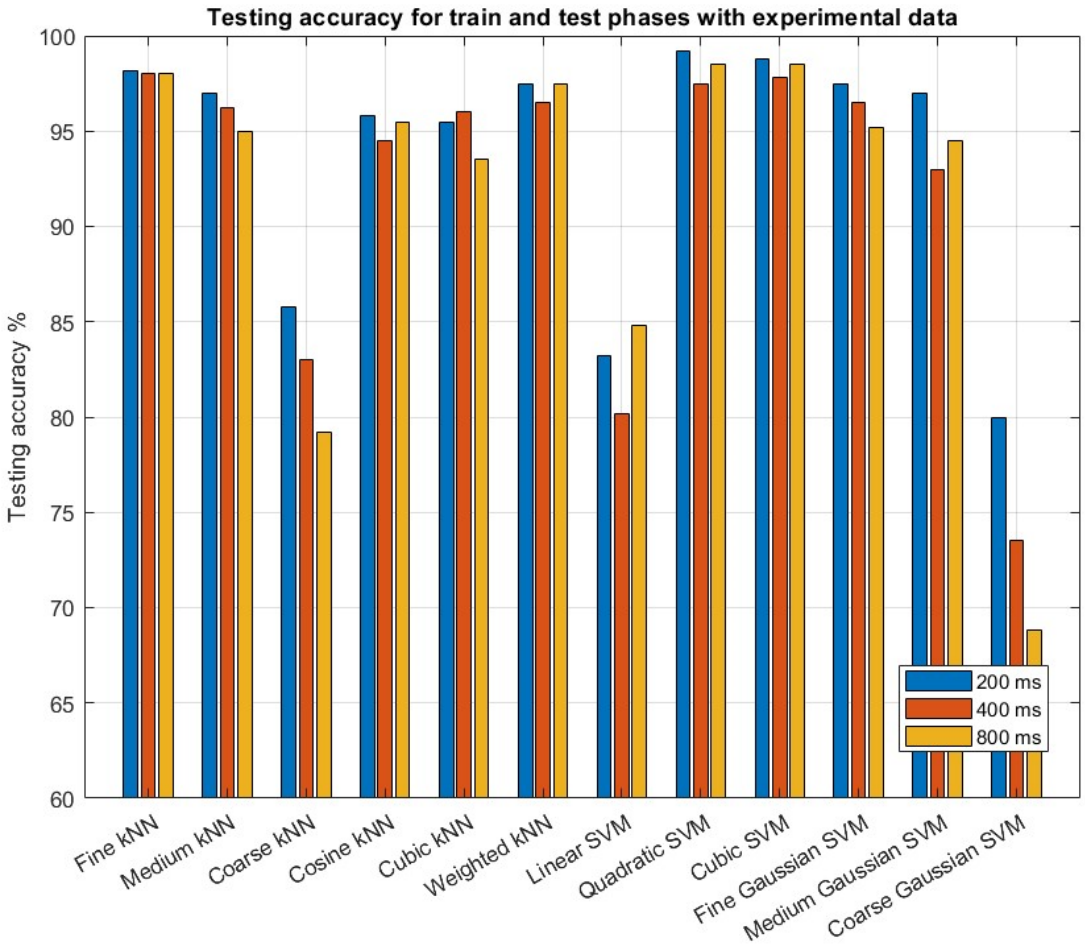

Figure 8.4: Testing accuracy for training and testing phases with experimental data

8.4 FEATURE SELECTION WITH SHAPLEY VALUES

As described in Section 7.4.3, the Shapley values represent the contribution of each feature to the outcome of the classification problem handled by the machine learning algorithms.

Shapley values can be computed through the Matlab command `shapley`, that evaluates them for a specific query point and for all the features. The Shapley value of a feature for a query point explains the deviation of the prediction for the query point from the average prediction due to the feature itself. For each query point, the sum of the Shapley values of each feature corresponds to the total deviation of the prediction from the average.

In the computation of the Shapley values, the extension to KernelSHAP method was used, because it does not require the feature independence assumption. In fact, not all the selected features in this analysis are independent, because some of them are linked through mathematical relationships (see Table 8.2).

Shapley values were computed for quadratic SVM and fine kNN algorithms, because, as it can be seen from Figure 8.4, if they are trained and tested on experimental data, they are the most accurate ones for the two types of algorithms analysed in this work. The boxplots in Figure 8.5 and Figure 8.6 show the Shapely values of the three defect classes for the two algorithms trained on numerical (on the right side) and experimental data (on the left side of the figure).

Boxplot were obtained training the algorithms with $N_{train} = 1600$ samples, as it was done for the calculation of the previous section, evaluating the Shapley values for $N_{Shapley} = 100$ samples, recording their values only for correctly classified points and finally averaging the recorded Shapley values for each class (0, 1 and 2) and for each of the twenty-three features. In each box, the central mark indicates the median value, while the bottom and top edges of the box indicate the 25th and 75th percentiles respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '+' marker symbol. Names and mathematical definitions of features are resumed in Table 8.2.

From the analysis of the boxplots related to the quadratic SVM algorithm trained with numerical data it can be seen that features $p_2 - p_5$ (STD, SMRA, RMS, peak) and the last four ones ($p_{20} - p_{23}$) have almost the same median value and little variance across the three defect classes, so they can be considered as non-relevant features; p_1 (mean) has got the highest median value and very little variance for class 1, so it is a relevant feature for defect-free condition; p_6, which is the measure of signal skewness, seems to be determinant in the recognition of an inner race defect, since it is positive only for class 1; the other features that contribute to this class are $p_7 - p_{11}$ (Kurtosis, CF, CLF, SF, IF), which all belong to time domain; p_{14} and p_{15} are the most relevant feature in outer race defect recognition, but they have a quite big variance, so they could be ineffective in some cases; they are also positive for class 2 and negative for class 0; p_{19} has got a strange behaviour, because the median value is positive only for class 1, but it has a huge variance, so it cannot be considered as a reliable feature.

In case the SVM algorithm has been trained with experimental samples, it can be seen from the graphs that p_1, which is the mean value, does not have any influence (its value is almost zero for all the classes), in fact as it can be seen from Figure 9.1, experimental time-domain vibration does not change a lot if the type of defect is different and therefore, the mean value is almost the same; STD, SMRA, RMS and peak value ($p_2 - p_5$) discriminate classes 0 and 1 from class 2, p_6 (skewness) has got a great impact on all the classes, but especially on outer defect class and it has got a great variability; frequency-domain features ($p_{12} - p_{23}$) have got a median value near to zero for class 0, so they seem to be crucial in the recognition of the type of defect.

In the boxplots regarding the fine kNN algorithm trained with numerical data, it can be seen that p_1, as in the case of quadratic SVM, has got a great impact on class 0, while $p_2 - p_5$ do not give a significant contribution, because they have almost the same value for all the classes; also in this case p_6 is a relevant feature for inner defect recognition; the other time-domain features ($p_7 - p_{11}$) have almost the same median value in the three cases and have a great variability for the two defect classes, so they cannot be considered as reliable; frequency-domain features do not have great influence, because they all have big variance and nearly the same median value across the classes;

only p_{19}, which helps to recognize an inner race defect, and p_{13} for the recognition of an outer race defect are significant features.

In case of the training phase with experimental data, the mean vibration level, described by feature p_1, does not play an important role, as it happens for SVM algorithm; $p_2 - p_5$ discriminate class 2 from the other ones; p_6 has got a very high variance in all the three cases and but the median Shapley value is barely the same; the other time-domain features does not seem to have a high discriminating capability; also for kNN algorithm, frequency-domain features are relevant in the recognition of the type of defect, so between class 1 and 2.

From the analysis on the Shapley values made in this section it is evident once more the difference highlighted between numerical and experimental data: if the machine learning algorithm is changed, the most important features are almost the same, but they have a quite different impact on the classification, if the algorithm has been trained with numerical or experimental samples. This can be seen in the discriminating capability of the mean value of the signal, which has got a great impact in the former case, while it is irrelevant in the latter

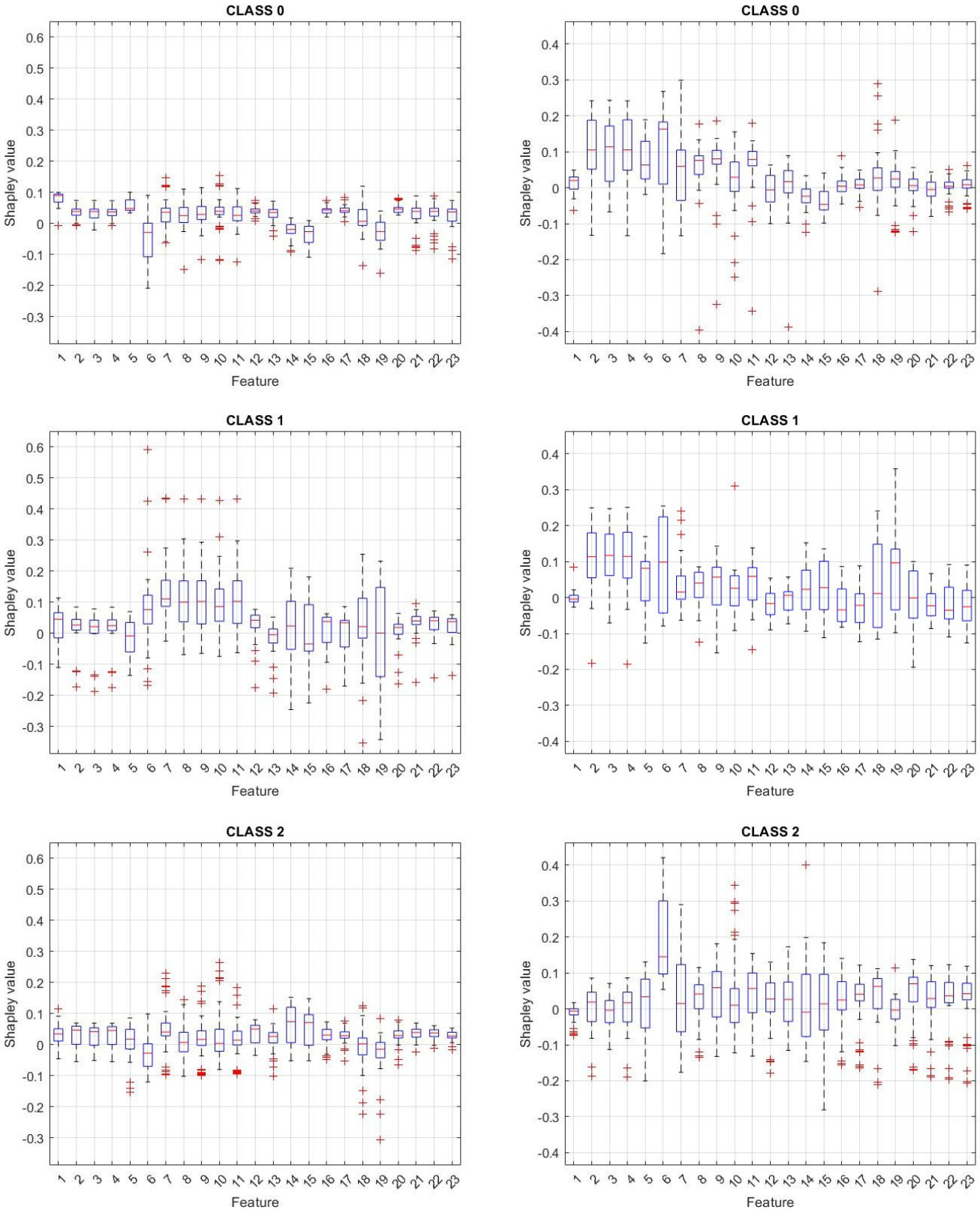

Figure 8.5: Shapley values of quadratic SVM algorithm trained on numerical data (left) and on experimental data (right) related to the three different defect classes 0 (top), 1 (middle), 2 (bottom)

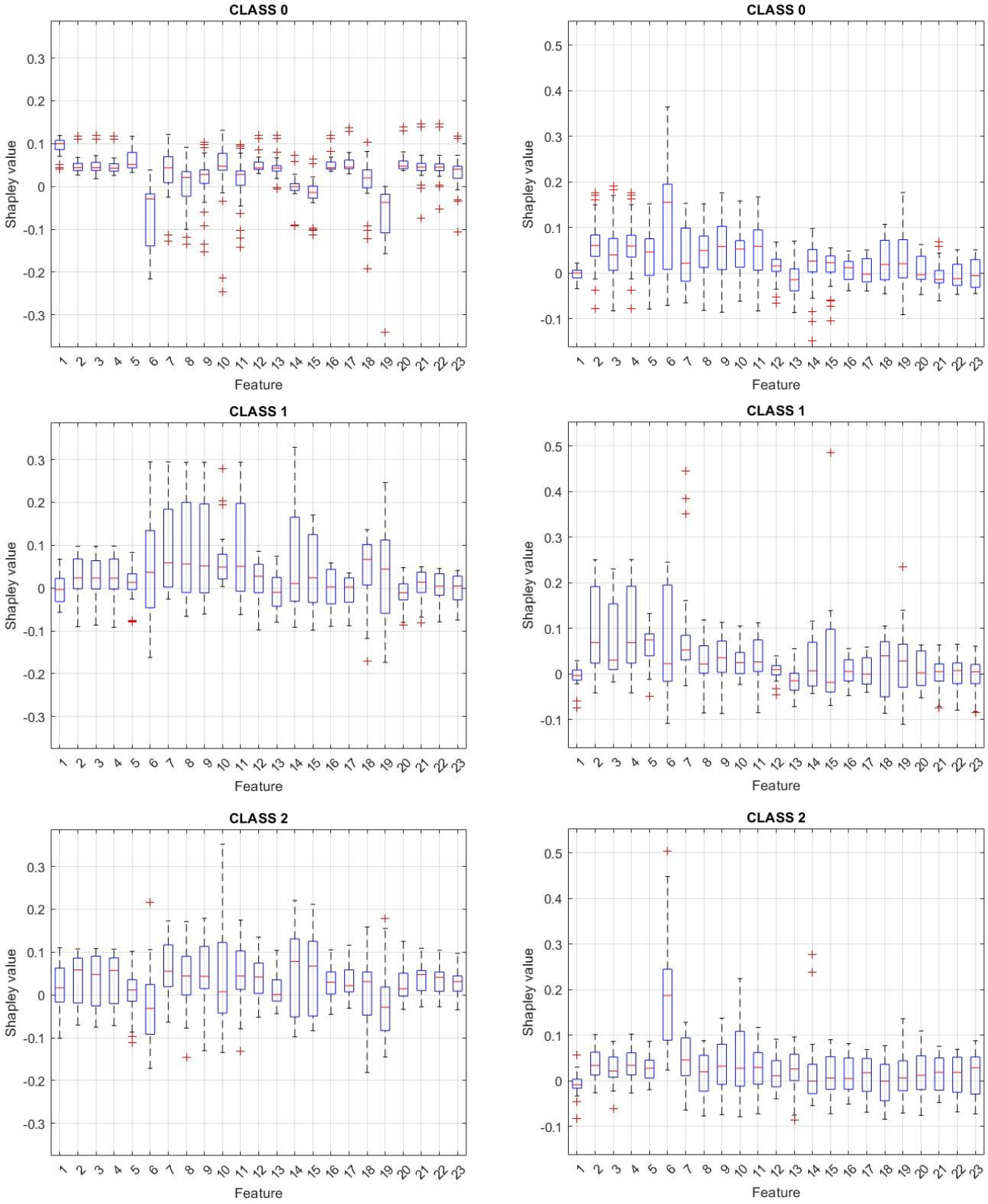

Figure 8.6: Shapley values of fine kNN algorithm trained on numerical data (left) and on experimental data (right) related to the three different defect classes 0 (top), 1 (middle), 2 (bottom)

9 CONCLUSIONS AND FUTURE WORK

9.1 SUMMARY OF THE OBTAINED RESULTS

In this work a nonlinear multi-body dynamic model of a spherical roller bearing was presented. It can simulate the bearing behavior under average radial load and shaft velocity conditions and the effects produced by localized defects on raceways. The effects of the operating temperature and lubricant film thickness and damping in the contacts between rolling elements and raceways are also added in the model through empirical formulas involving parameters related to geometric and loading conditions, lubricant properties, material elasticity, roughness and hardness.

A Simulink model was developed to carry out the simulations under a wide range of shaft rotational velocity, radial load and defect conditions to replicate vibration data obtained from a test rig, developed by Brusa et al. [56], which is located at Politecnico di Torino.

From the simulation results presented in Chapter 5, it could be seen that the model simulates the load distribution among rolling elements, the lubricant film thickness and damping coefficient in accordance with theory. Numerical vibration data correctly simulate the presence of defects on raceways as they present impulsive response when a defect is in the load zone and a rolling element falls in it. By performing the FFT on time signals, it was seen that the presence of a defect leads to the excitation of the characteristic frequencies of the bearing related to that kind of defect, as it was expected.

Numerical and experimental vibration paths were divided into shorter time segment lasting $200, 400$ or 800 ms and twenty-three features among those of time and frequency domains were extracted from each of them. Therefore, from each segment a sample, containing the related features, with a label depending on the type of defect, was obtained and a wide database was created. Before the feature extraction activity, numerical data were added with a white Gaussian noise with different levels of SNR to simulate the noise present in experimental vibration paths, that comes from the other mechanical and electrical components of the test rig.

A number of randomly selected samples was extracted to train many machine learning algorithms belonging SVM kNN and their accuracy was tested on a new set of data, taken randomly from the unused samples of the same dataset. It could be seen that testing accuracy was close to 100 % and had best values for longer time segments. It was also noticed that a higher level of noise worsened the performance of the AI algorithms, because defect-induced peaks are hidden. The same operations were done using experimental data and the best testing accuracy was achieved by fine kNN, quadratic and cubic SVM. Therefore, it was successfully checked that AI algorithms can be used in the diagnostics of bearing health conditions.

Finally, the Shapley values, which represent the contribution of each feature to the classification of a query point, were computed for 100 query points using quadratic SVM and fine kNN algorithms to highlight the most relevant feature in IFD. The Shapley values were computed for the algorithms trained with both numerical and experimental data and results were compared: it could be seen that the most relevant features are nearly the same for the two kinds of AI algorithms, but they differ due to the types of data used for the training phase.

9.2 LIMITATIONS AND DIRECTIONS FOR FUTURE WORK

The developed nonlinear multi-body dynamic model was designed for spherical roller bearings, but it can be easily modified to simulate the behavior of any kind of bearing. A more comprehensive and accurate simulation could be obtained with the introduction of two more degrees of freedom (one for each ring) in the z-direction, that allows the simulation of external axial loads and preloads.

In the first formulation of the model presented in this work the high frequency resonator was not added, but it is one of the topics for future developments as well as the inclusion of the two rows of rolling elements. Further analysis on the effects of rolling elements slippage and the way to accurately simulate it should also be developed.

As it can be seen from Figure 9.1, which shows time, frequency and envelope spectrum signals for a spherical roller bearing under 62.4 kN load, with a rotational speed of 127 rpm and in different defect conditions, the numerically simulated bearing and the real one have a quite different behavior. From the time signal it is evident that the biggest discrepancy is the vibration amplitude, but it is a common problem related to multi-body dynamic models due to the initial assumptions and simplifications, as it mentioned in Liu et al. [33].

The numerical vibration signal is dominated by the defect-induced peaks, while the experimental one seems to be only slightly influenced, since the biggest contribution to acceleration comes from the other parts of the mechanical system. This can also be seen from experimental frequency and envelope spectra because the highlighted characteristic frequencies are not excited significantly. On the contrary, in the plots coming from simulations, characteristic frequencies and their multiples are the only excited frequencies.

Discrepancies between numerical and experimental data can also be seen in extracted features; in fact, once algorithms have been trained with numerical data, they can be tested on experimental data in a number equal to testing data samples $N_{test} = 400$. The evaluation of testing accuracy was performed using samples coming from experimental vibration paths lasting 400 ms and it is reported in Figure 9.2. It is evident that overall testing accuracy on experimental data is quite unsatisfactory, since it is below 50 %, which is the coin toss case. Moreover, if the SNR level of the white Gaussian noise is higher, algorithms tend to work better on testing data and “coarser” algorithms have better performances (coarse kNN, coarse SVM). This is a proof of the corruption of experimental vibration paths by a random noise.

The plot in Figure 9.3 shows the Area Under Curve (AUC) for the selected machine learning algorithms tested on experimental data. AUC corresponds to the integral of the Receiver Operating Characteristics (ROC) curve, which relates the true positive rate (TPR), or sensitivity, to the false positive rate (FPR), or specificity. AUC values are in the range $[0,1]$ and larger values corresponds to better classifier performances, since it indicates the ability of the algorithm to distinguish between classes: a perfect classifier has $AUC = 1$, because it always assigns correctly the positive class observation to the positive class and does the same for the negative ones, while a random classifier has got $AUC = 0.5$, since false positive rate and true positive rate have always the same value.

From the graph it can be seen that the overall AUC value is not much higher than 0.5, so the performance of evaluated algorithms is not good. Moreover, for kNN algorithms AUC becomes higher if numerical training data are more corrupted by noise, while it is the opposite in case of SVM, which reaches the highest values of AUC in case of linear and quadratic kernels.

The most common problem found in algorithms is their incapability of detecting faults, as they almost classify all samples as non-defective. Better results were obtained with medium and coarse Gaussian SVM, especially if the noise added to numerical data is comparable or little higher than the original signal ($SNR = 0, -3$ dB), because they can detect the presence of the three defect classes, even if their accuracy is not high.

The source of problems seems to be again the big difference in vibration behavior between the modelled and real bearing. Thus, machine learning algorithms have a very low accuracy if they are trained with numerical data and tested on experimental ones, because extracted features have different distributions.

All those problems could be related to the difference between real and simulated defects; in fact, the bearing tested on the rig has got a circular defect with a diameter of 2 mm, while the modelled one is 2 mm wide a has a length equal to the roller contact length. This surely leads to greater accelerations. Therefore, the defect model has to be changed, even if the one introduced in this work, coming from Wen et al. [52], simulates in a more realistic way than many other bearing models the penetration of the rolling element into the defect.

Lastly, a more comprehensive and accurate simulation of the vibration, that can replicate with better precision the experimental data, can be obtained introducing in the dynamic model all the mechanical components of the test rig: in this case the number of DoF may increase considerably and so the numerical effort. In this case the excited frequencies, which can be seen in experimental frequency spectrum, could be simulated.

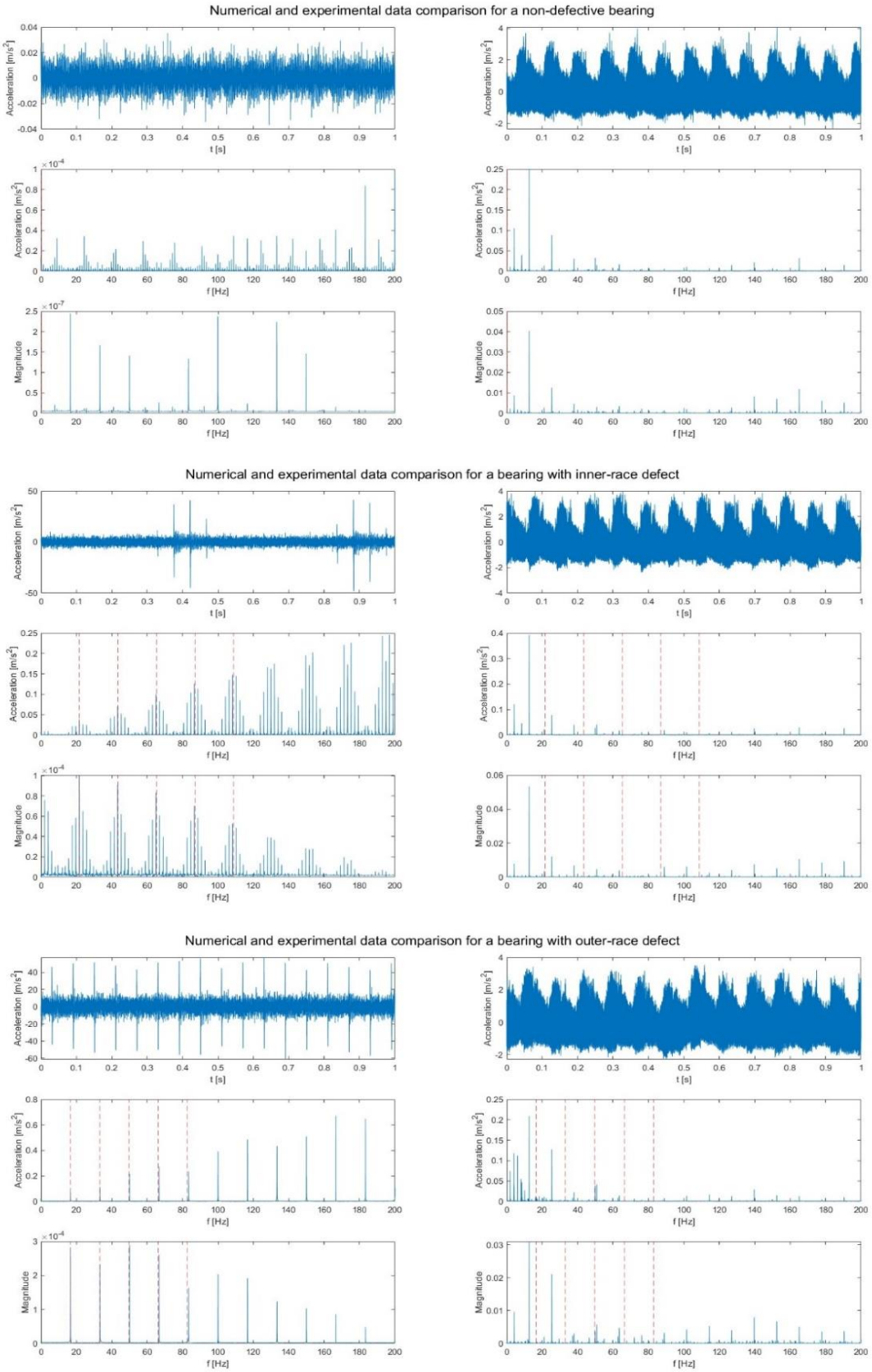

Figure 9.1: Comparison of numerical (left) and experimental (right) time-domain, frequency-domain and envelope signal for non-defective (top), inner race-defective (middle) and outer race-defective (bottom) bearing

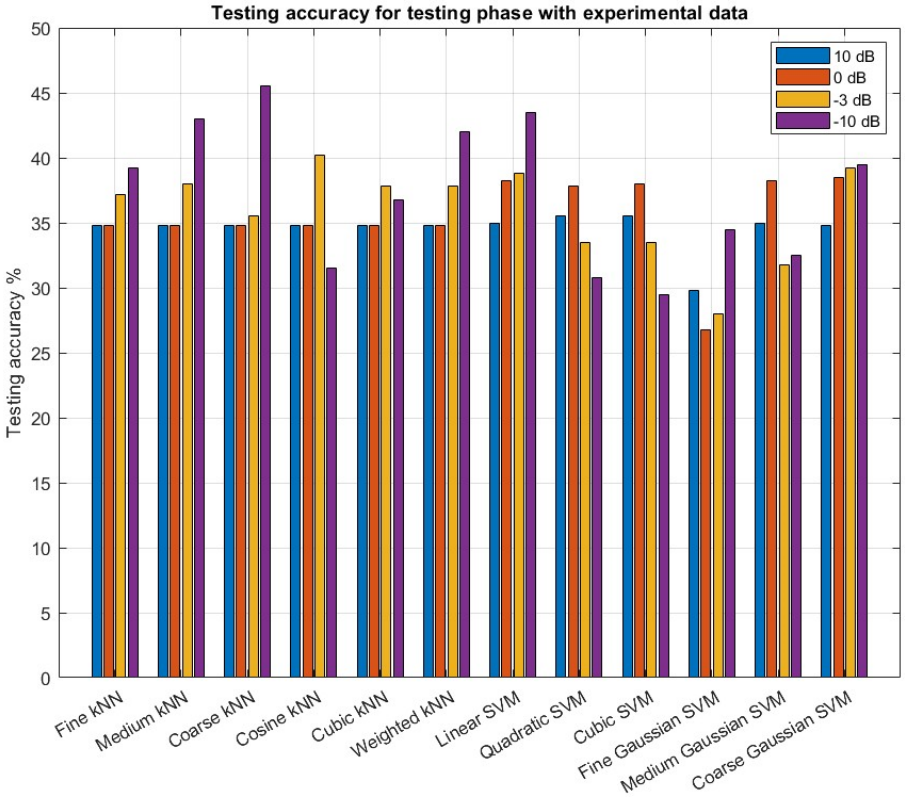

Figure 9.2: Testing accuracy for AI algorithms trained with numerical data affected by white Gaussian noise with different SNR and tested on experimental data

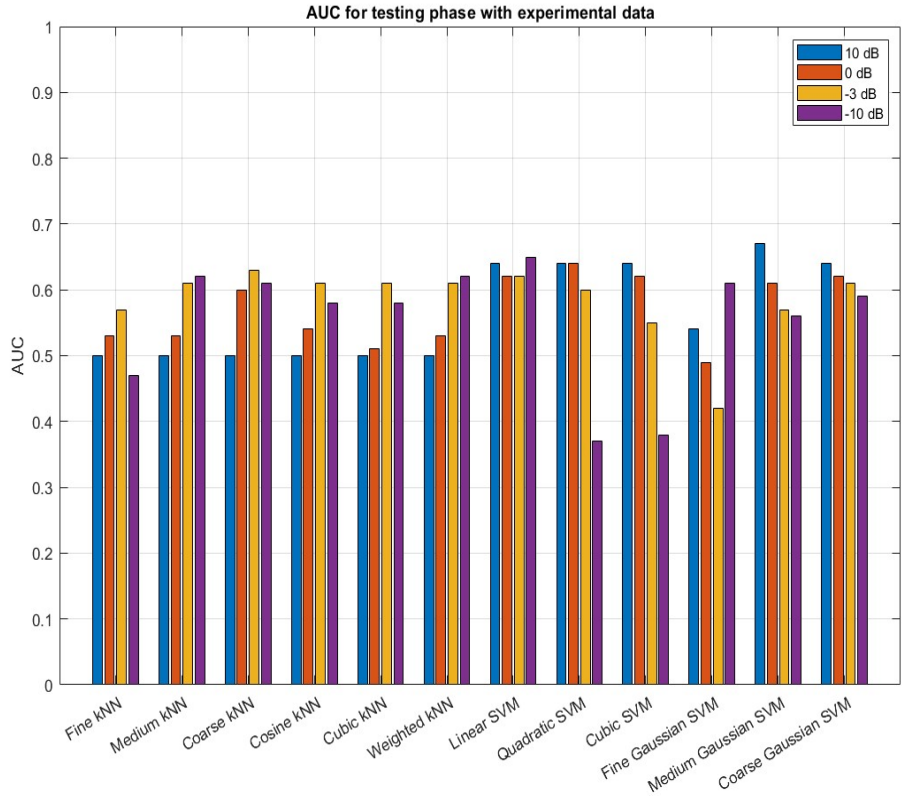

Figure 9.3: AUC of each AI algorithm trained with numerical data corrupted by white Gaussian noise having different levels of SNR and tested on experimental data

REFERENCES

[1] S. J. Lacey, «An overview of bearing vibration analysis,» *Maintenance & asset management,* vol. 23, n. 6, pp. 32-42, 2008.

[2] S. Singh, C. Q. Howard and C. H. Hansen, «An extensive review of vibration modelling of rolling element bearings with localised and extended defects,» *Journal of Sound and Vibration,* vol. 357, pp. 300-330, 2015.

[3] A. Rezaei, «Fault Detection and Diagnosis on the rolling element bearing,» Charleton University, 2007.

[4] Y. Lei et al., «Applications of machine learning to machine fault diagnosis: A review an a roadmap,» *Mechanical Systems and Signal Processing,* n. 138, 2020.

[5] P. K. Gupta, Advanced dynamics of rolling elements, Springer Science & Business Media, 2012.

[6] T. A. Harris and M. N. Kotzalas, Essential Concepts of Bearing Technology (5th ed.), CRC Press, 2006.

[7] P. Chen, «Bearing condition monitoring and fault diagnosis,» University of Calgary, 2001.

[8] B. Ghalamchi, «Dynamic Analysis Model of Spherical Roller Bearings with Defects,» 2014.

[9] SKF, Rolling Bearings, 2021.

[10] P. Dietl, Damping and stiffness characteristics of rolling element bearings, 1997.

[11] B. J. Brewe and D. E. Hamrock, «Simplified solution for elliptical-contact deformation between two elastic solids,» *Journal of tribology,* n. 99,4, pp. 485-487, 1977.

[12] A. Palmgren and G. Lundberg, Dynamic capacity of rolling bearings, 1947.

[13] H. Sjovall, «The load distribution within ball and roller bearings under given external radial and axial,» *Teknisk Tidskrift,* vol. 9, 1933.

[14] T. A. Harris, Rolling bearing analysis, Wiley-Interscience-Publication, 1991.

[15] B. Hamrock, Fundamentals of fluid film lubrication, McGraw Hill, 1994.

[16] H. Spikes, «Basics of EHL for practical application,» *Lubrication science,* vol. 27, n. 1, pp. 45-67, 2015.

[17] A. Grubin, «Fundamentals of the hydrodynamic theory of lubrication of heavily loaded cylindrical surfaces,» *Investigation of the Contact Machine Componets,* vol. 2, 1949.

[18] N. A. H. Tsuha and K. L. Cavalca, «Stiffness and damping of elastohydrodynamic line contact applied to cylindrical roller bearing dynamic model,» *Journal of Sound and Vibration,* vol. 481, 2020.

[19] N. A. H. Tsuha and K. L. Calcalca, «Finite line contact stiffness under elastohydrodynamic lubrication considering linear and nonlinear force models,» *Tribology International,* vol. 146, 2020.

[20] C. H. Venner, Multilevel solution of the EHL line and point contact problems, 1991.

[21] W. Yunlong et al., «Effects of raceway surface roughness in an angular contact ball bearing,» *Mechanism and Machine Theory,* vol. 121, pp. 198-212, 2018.

[22] N. Patir and H. S. Cheng, «Average Flow Model for Determining Effects of Three-Dimensional Roughness on Partial Hydrodynamic Lubrication,» *ASME J. Lubr. Technol.,* pp. 12-17, 1978.

[23] M. Masjedi and M. M. Khonsari, «Film Thickness and Asperity Load Formulas for Line-Contact Elastohydrodynamic Lubrication With Provision for Surface Roughness,» *Journal of Tribology,* vol. 134, n. 1, 2012.

[24] M. Masjedi and M. M. Khonsari, «Mixed elastohydrodynamic lubrication line-contact formulas with different surface patterns,» *Journal of Engineering Tribology,* vol. 228, n. 8, pp. 849-859, 2014.

[25] E. Kramer, Dynamics of rotors and foundations, Springer Science & Business Media, 1993.

[26] R. Zeillinger, «Report on preliminary investigations on damping mechanisms in rolling bearing joints,» 1992.

[27] S. H. Crandall, «The role of damping in vibration theory,» *Journal of sound and vibration,* vol. 11, n. 1, 1970.

[28] B. J. Lazan, Damping of materials and members in structural mechanics, Pergamon Press, 1968.

[29] K. H. Hunt and C. F. R. E., «Coeffcient of restitution interpreted as damping in vibroimpact,» *ASME, Journal of Applied Mechanics,* pp. 440-445, 1975.

[30] P. Krempf and J. Sabot, «Identification of the damping in a Hertzian contact from experimental non-linear response curve,» in *Proceedings of the IUTAM Symposium on Identification of Mechanical Systems*, University of Wuppertal, Germany, 1993.

[31] G. Van Nijen, On the overrolling of local imperfections in rolling bearings, 1997.

[32] C. Zhou and Z. Xiao, «Stiffness and damping models for the oil film in line contact elastohydrodynamic lubrication and applications in the gear drive,» *Applied Mathematical Modelling,* vol. 61, pp. 634-649, 2018.

[33] J. Liu and Y. Shao, «Overview of dynamic modelling and analysis of rolling element bearings with localized and distributed faults,» *Nonlinear Dynamics,* vol. 93, n. 4, pp. 1765-1798, 2018.

[34] N. S. Feng, E. J. Hahn and R. B. Randall, «Using transient analysis software to simulate vibration signals due to rolling element bearing defect,» *Applied Mechanics: Progress and Applications,* pp. 689-694, 2002.

[35] S. Fukata et al., «On the radial vibration of ball bearings: computer simulation,» *Bulletin of JSME 28.239,* vol. 28, n. 239, pp. 899-904, 1985.

[36] G. Lungberg and A. Palmgren, «Dynamic capacity of rolling bearing,» *Acta Polytechnic Mechanical Engineering Series,* vol. 1, n. 3, 1947.

[37] P. K. Gupta, «Transient ball motion and skid in ball bearings,» *Journal of Lubrication technology,* vol. April, 1975.

[38] M. Tiwari, K. Gupta and O. Prakash, «Effect of radial internal clearance of a ball bearing on the dynamics of a balanced horizontal rotor,» *Journal of Sound and Vibration,* vol. 238, n. 5, pp. 723-756, 2000.

[39] M. Tiwari, K. Gupta and P. O., «Dynamic response of an unbalanced rotor supported on ball bearings,» *Journal of Sound and,* vol. 238, n. 5, pp. 757-779, 2000.

[40] Y. H. Wijnat, J. A. Wensing and G. C. Van Nijet, «The influence of lubrication on the dynamic behaviour of ball bearings,» *Journal of Sound and Vibration,* vol. 222, n. 4, pp. 576-596, 1999.

[41] J. Sopanen and A. Mikkola, «Dynamic model of a deep-groove ball bearing including localized and distributed defects. Part 1: Theory,» *Journal of Multi-body Dynamics,* vol. 217, n. 3, pp. 201-211, 2003.

[42] J. Sopanen and A. Mikkola, «Dynamic model of a deep-groove ball bearing including localized and distributed defects. Part 2: Implementation and results,» *Proceedings of the Institution of Mechanical Engineers, Part K: Journal of Multi-body Dynamics,* vol. 217, n. 3, pp. 213-223, 2003.

[43] N. Sawalhi and R. B. Randall, «Simulating gear and bearing interactions in the presence of faults: Part I. The combined gear bearing dynamic model and the simulation of localised bearing faults,» *Mechanical Systems and Signal Processing,* vol. 22, n. 8, pp. 1924-1951, 2008.

[44] N. Sawalhi and R. B. Randall, «Simulating gear and bearing interactions in the presence of faults: Part II: Simulation of the vibrations produced by extended bearing faults,» *Mechanical Systems and Signal Processing,* vol. 22, n. 8, pp. 1952-1966, 2008.

[45] D. Petersen et al, «Analysis of bearing stiffness variations, contact forces and vibrations in radially loaded double row rolling element bearings with raceway defects,» *Mechanical Systems and Signal Processing,* vol. 50, pp. 139-160, 2015.

[46] D. Petersen, C. Howard and Z. Prime, «Varying stiffness and load distributions in defective ball bearings: Analytical formulation and application to defect size estimation,» *Journal of Sound and Vibration,* vol. 337, pp. 284-300, 2015.

[47] A. M. Ahmadi, D. Petersen and C. Howard, «A nonlinear dynamic vibration model of defective bearings–The importance of modelling the finite size of rolling elements,» *Mechanical Systems and Signal Processing,* vol. 52, pp. 309-326, 2015.

[48] D. S. Shah and V. N. Patel, «A review of dynamic modeling and fault identifications methods for rolling element bearing,» *Procedia Technology,* vol. 14, pp. 447-456, 2014.

[49] P. Yan et al., «5-DOF dynamic modeling of rolling bearing with local defect considering comprehensive stiffness under isothermal elastohydrodynamic lubrication,» *Shock and Vibration,* 2020.

[50] C. F. Yan, «Dynamics modeling on local defect of deep groove ball bearing under point contact elasto-hydrodynamic lubrication condition,» *Journal of Vibration and Shock,* vol. 35, n. 14, pp. 61-70, 2016.

[51] C. F. Yan, «Dynamic modeling for local defect displacement excitation in rolling bearing systems under elasto-hydrodynamic lubrication and slip,» *Journal of Vibration and Shock,* vol. 37, n. 5, pp. 56-64, 2018.

[52] C. Wen et al., «Dynamic behaviors of angular contact ball bearing with a localized surface defect considering the influence of cage and oil lubrication,» *Mechanism and Machine Theory,* vol. 162, 2021.

[53] A. Liew, N. Feng and E. J. Hahn, «Transient rotordynamic modeling of rolling element bearing systems,» *Gas Turbines Power,* vol. 124, n. 4, pp. 984-991, 2002.

[54] S. Sassi, B. Badri and M. Thomas, «A Numerical Model to Predict Damaged Bearing Vibrations,» *Journal of Vibration and Control,* vol. 13, n. 11, pp. 1603-1628, 2007.

[55] ASME Research Commitee on Lubrication, Pressure-Viscosity Report-Vol. II, ASME, 1953.

[56] E. Brusa et al., «Design of an Innovative Test Rig for Industrial Bearing Monitoring with Self-Balancing Layout,» *Machines,* vol. 10, n. 1, 2022.

[57] Y. Lei et al., «A multidimensional hybrid intelligent method for gear fault diagnosis,» *Expert Systems with Applications,* vol. 37, n. 2, pp. 1419-1430, 2010.

[58] A. Oppenheim, Discrete-time signal processing, Pearson Education India, 1999.

[59] V. Vapnik, The nature of statistical learning theory, Springer science & business media, 1999.

[60] A. Widodo and B. S. Yang, «Support vector machine in machine condition monitoring and fault diagnosis,» *Mechanical systems and signal processing,* vol. 21, n. 6, pp. 2560-2574, 2007.

[61] A. Widodo et al., «Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine,» *Expert systems with applications,* vol. 36, n. 3, pp. 7252-7261, 2009.

[62] E. Fix and J. L. Hodges, «Discriminatory analysis. Nonparametric discrimination: Consistency properties,» *nternational Statistical Review/Revue Internationale de Statistique,* vol. 57, n. 3, pp. 238-247, 1989.

[63] T. Cover and P. Hart, «Nearest neighbor pattern classification,» *IEEE transactions on information theory,* vol. 13, n. 1, pp. 21-27, 1967.

[64] J. Wang, P. Neskovic and L. N. Cooper, «Neighborhood size selection in the k-nearest-neighbor rule using statistical confidence,» *Pattern Recognition,* vol. 39, n. 3, pp. 417-423, 2006.

[65] S. Landau et al., Cluster analysis, John Wiley & Sons, 2011.

[66] B. Shaw and T. Jebara, «Structure preserving embedding,» in *Proceedings of the 26th Annual International Conference on Machine Learning*, 2009.

[67] P. Hart, «The condensed nearest neighbor rule,» *IEEE transactions on information theory,* vol. 14, n. 3, pp. 515-516, 1968.

[68] L. K. Saul and S. T. Roweis, «An introduction to locally linear embedding,» 2000.

[69] S. Buchaiah and P. Shakya, «Bearing fault diagnosis and prognosis using data fusion based feature extraction and feature selection,» *Measurement,* vol. 188, 2022.

[70] L. Shapley, «A value for n-person games,» *Contributions to the theory of games,* vol. 2, 1953.

[71] K. Aas, M. Jullum and A. Løland, «Explaining individual predictions when features are dependent: More accurate approximations to Shapley values,» *Artificial Intelligence,* vol. 298, 2021.

[72] I. Covert, S. M. Lundberg and S. I. Lee, «Understanding global feature contributions with additive importance measures,» *Advances in Neural Information Processing Systems,* vol. 33, pp. 17212-17223, 2020.

[73] S. M. Lundber and S. I. Lee, «A unified approach to interpreting model predictions,» *Advances in neural information processing systems,* vol. 30, 2017.

[74] I. E. Kumar et al., «Problems with Shapley-value-based explanations as feature importance measures,» in *International Conference on Machine Learning*, 2020.

[75] Y. Lei et al., «Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs,» *Mechanical systems and signal processing,* vol. 21, n. 5, pp. 2280-2294, 2007.

[76] Z. Xie et al., «Investigations on transitions of lubrication states for water lubricated bearing. Part I: Determination of friction coefficients and film thickness ratios,» *Industrial Lubrication and Tribology,* 2016.