All structures are subject to catastrophic failures associated with damages and its incremental nature. Therefore, the damage must be detected as early as possible, evaluating the structural integrity so that corrective actions are taken at the right time, reducing costs with downtime without losing reliability. In this work, the Multilayer Perceptron (MLP) network was used to recognize patterns from dynamic responses of healthy and damaged samples (rolling bearings and composite beams). Statistical Parameters (SP), Principal Component Analysis (PCA) and Dislocated Series (DS) were used as data treatment techniques. The results found show that the propose damage detection methodology is an efficient Structural Health Monitoring (SHM) tool.

Adviser: Dr. Ricardo De Medeiros.

Co-Adviser: Dr. Eduardo Lenz Cardoso.

Joinville, 2020

PEDRO ALMEIDA REIS| ARTIFICIAL NEURAL NETWORKS APPLIED TO STRUCTURAL DAMAGE IDENTIFICATION USING DYNAMIC RESPONSE AND SIGNAL PROCESSING

UDESC

SANTA CATARINA STATE UNIVERSITY – UDESC
COLLEGE OF TECHNOLOGICAL SCIENCE – CCT
MECHANICAL ENGINEERING GRADUATE PROGRAM – PPGEM

MASTER THESIS

ARTIFICIAL NEURAL NETWORKS APPLIED TO STRUCTURAL DAMAGE IDENTIFICATION USING DYNAMIC RESPONSE AND SIGNAL PROCESSING

PEDRO ALMEIDA REIS

JOINVILLE, 2020

Ficha catalográfica elaborada pelo programa de geração automática da

Biblioteca Setorial do CCT/UDESC,

com os dados fornecidos pelo(a) autor(a)

Reis, Pedro
ARTIFICIAL NEURAL NETWORKS APPLIED TO STRUCTURAL DAMAGE IDENTIFICATION USING DYNAMIC RESPONSE AND SIGNAL PROCESSING / Pedro Reis. -- 2020.
140 p.

Orientador: Ricardo De Medeiros
Coorientador: Eduardo Cardoso
Dissertação (mestrado) -- Universidade do Estado de Santa Catarina, Centro de Ciências Tecnológicas, Programa de Pós-Graduação em Engenharia Mecânica, Joinville, 2020.

1. Artificial Neural Networks. 2. Structural Health Monitoring . 3. Damage Detection. 4. Composite Materials . 5. Rolling Bearings. I. De Medeiros, Ricardo. II. Cardoso , Eduardo. III. Universidade do Estado de Santa Catarina, Centro de Ciências Tecnológicas, Programa de Pós-Graduação em Engenharia Mecânica. IV. Titulo.

-

Artificial Neural Networks applied to structural damage identification using dynamic response and signal processing

por

Pedro Almeida Reis

Esta dissertação foi julgada adequada para obtenção do título de

MESTRE EM ENGENHARIA MECÂNICA

Área de concentração em "Modelagem e Simulação Numérica"
e aprovada em sua forma final pelo

CURSO DE MESTRADO ACADÊMICO EM ENGENHARIA MECÂNICA
DO CENTRO DE CIÊNCIAS TECNOLÓGICAS DA
UNIVERSIDADE DO ESTADO DE SANTA CATARINA.

Banca Examinadora:

ASSINADO DIGITALMENTE
Prof. Dr. Ricardo de Medeiros
CCT/UDESC
(Orientador/Presidente)

ASSINADO DIGITALMENTE
Prof. Dr. André Tavares da Silva
CCT/UDESC

VIA VIDEOCONFERÊNCIA
Prof. Dr. Marcelo Leite Ribeiro
EESC/USP

Joinville,SC, 7 de agosto de 2020.

-

To my mother, family and friends. In memory of my grandmother Francisca.

ACKNOWLEDGMENTS

The author would like to express the following acknowledgments:

To Prof. Dr. Ricardo De Medeiros, professor, advisor, and friend for all the support in guiding the development of this work.

To Prof. Dr. Eduardo Lenz Cardoso, co-advisor, for the attention given in key points of this work.

To the Professors Prof. Dr. Pablo Andrés Muñoz Rojas and P.h.D Miguel Vaz Junior that contributed to my formation.

To Prof. Dr. Marco Túlio, MSc. Antônio Andrade, and Prof. Jardel Jesus for encouraging me.

To my laboratory partners in the weekends, Kelvin Iwasaki, Gabriel Nagafugi, and Matheus Rodrigues.

To the Santa Catarina State University (UDESC) for the high quality education provided.

To my mother for the unconditional love, affection, care and education.

To my family who supported me in this important stage of my career.

To my friends João Medeiros, Natsue Mizubuti, Gustavo Heiden, Keila Mercedes, Eduardo Menezes, Larissa Steiger, Augusto dos Santos, Luiz Gabriel, Felipe Faria, Taísa Utida, Vanessa Menezes, Luiz Fernando, José Thiesen and Luísa Völtz for having welcomed me so well.

This study was financed by the Coordenacão de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES - Finance Code 001) and PROMOP (Programa de Bolsas de Monitoria de Pós-Graduação) of the Santa Catarina State University..

The authors acknowledge funding of this research project by Brazilian agency FAPESC (Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina, grants 2017TR1747, 2017TR784 and 2019TR779)

“However difficult life may seem,
there is always something you can do,
and succeed at. It matters that you
don’t just give up”.
(Stephen Hawking)

ABSTRACT

ALMEIDA REIS, Pedro, ARTIFICIAL NEURAL NETWORKS APPLIED TO STRUCTURAL DAMAGE IDENTIFICATION USING DYNAMIC RESPONSE AND SIGNAL PROCESSING. 2020. Master Thesis (Master in Mechanical Engineering - Area: Numerical Modeling and Simulation) – Santa Catarina State University. Mechanical Engineering Graduate Program Joinville 2020.

In general, all structures can be subject to damage leading to catastrophic failures, causing loss of human life, environmental tragedies, and significant financial losses, due to incremental nature, observed from its emergence. Aiming to ensure reliability and safety, Structural Health Monitoring (SHM) has gained evidence, mainly through the use of tools such as the Vibration-Based Model (VBM) and Artificial Neural Network (ANN). VBM uses the premise that some properties are modified when damage is present, such that the dynamic response is modified, making it possible to monitor these changes to identify damage. In this way, this work focuses on the utilize of ANN to identify damage in rolling bearings and composite structures using vibration data. However, it is not feasible to directly use these high dimensional data, since it requires complex models, with a high computational cost. Therefore, it is necessary to use tools capable of processing the data without losing fundamental information for detecting and classifying damages. Thus, strategies as Statistical Parameters (SP), Dislocated-Series (DS), and Principal Component Analysis (PCA) were used to this end. Finally, it is evaluating each one from some case studies considered as damage detection in balls and inner race for rolling bearings of the data set provided by Case Western Reserve University (CWRU) and delamination damage in composite beams. The results found in this work show that the proposed damage detection methodology is an efficient SHM tool.

Key-words: Artificial Neural Networks, Structural Health Monitoring, Damage detection, Composite Materials, Rolling Bearings fault diagnosis, Vibration Based Model, Statistical Parameters, Principal Component Analysis.

RESUMO

ALMEIDA REIS, Pedro, ARTIFICIAL NEURAL NETWORKS APPLIED TO STRUCTURAL DAMAGE IDENTIFICATION USING DYNAMIC RESPONSE AND SIGNAL PROCESSING. 2020. Dissertação (Mestrado em Engenharia Mecânica - Área: Modelagem e Simulação Numérica) – Universidade do Estado de Santa Catarina. Programa de Pós-Graduação em Engenharia Mecânica Joinville 2020.

Em geral, todas as estruturas podem estar sujeitas a danos que levam a falhas catastróficas, causando perda de vidas humanas, tragédias ambientais e perdas financeiras significativas, devido à natureza incremental, observadas desde o seu surgimento. Com o objetivo de garantir confiabilidade e segurança, o Monitoramento de Integridade Estrutural (SHM) ganhou evidência, principalmente por meio do uso de ferramentas como o Modelo Baseado em Vibração (VBM) e Rede Neural Artificial (ANN). O VBM usa a premissa de que algumas propriedades são modificadas quando o dano está presente, de modo que a resposta dinâmica é modificada, possibilitando o monitoramento dessas alterações para identificar danos. Dessa forma, o presente trabalho enfoca o uso de RNA para identificar danos em rolamentos e estruturas compostas usando dados de vibração. No entanto, não é possível usar diretamente esses dados que possuem alta dimensão, pois requer modelos complexos, com alto custo computacional. Portanto, é necessário usar ferramentas capazes de processar os dados sem perder informações fundamentais para detectar e classificar os danos. Assim, estratégias como Parâmetros Estatísticos (SP), Séries Deslocadas (DS) e Análise de Componentes Principais (PCA) foram utilizadas para esse fim. Avaliando cada estudo de caso considerado como detecção de danos nas esferas e pista interna de rolamentos do conjunto de dados fornecido pela Case Western Reserve University (CWRU) e danos por delaminação em vigas compostas. Os resultados encontrados neste trabalho mostram que a metodologia de detecção de dano proposta é uma ferramenta eficiente para o SHM.

Palavras-chave: Redes Neurais Artificiais, Monitoramento de Integridade das Estruturas, Detecção de dano, Materiais Compósitos, Diagnóstico de falha em rolamentos, Modelo baseado em vibrações, Parâmetros Estatísticos, Análise de Componente Principal.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Developments in the last century provided several advances in mechanical systems, becoming transport, natural resources extraction, and industrial production faster, and more effective to supply a population with almost 7.7 billion inhabitants in the world, according to the UN (2019). However, techniques used in the development of these systems have limitations in terms of physical and mathematical modeling, leading to uncertainties in its respective applications. So, it can lead to unexpected failures that pose the risk humans to the environment as well as economic losses. Thus, strategies aimed at inspection of the structures such as Non-Destructive damage detection methods are required. However, they need prior knowledge about the existence and location of damage, in addition to another limiting factor, which is accessibility. Therefore, as a way to cope with critical systems, Structural Health Monitoring (SHM) has been widely studied by scientists and engineers as a strategy to carry out continuous monitoring during the use of structure.

In 2015 and 2019, two similar tragedies occurred in Brazil: the rupture of tailing dams used by mining companies in Mariana and Brumadinho cities, as can be seen in Fig. 1.1(a) and Fig. 1.1(b), respectively. The total death toll is around 200, with 108 missings. Besides, there was a higher environmental impact with the deposition of mud in rivers, decimating fish, and other living beings in the surroundings. After the first occurrence, the *Agência Nacional de Mineração* (ANM) took action to increase the number of inspections for 220 dams located in the same region to check its integrity by the current rules (ANM, 2019). Nevertheless, the tragedy of equal nature ended up occurring in Brumadinho, resulting in a considerable number of deaths and environmental impact. The aforementioned suggests the demand for more qualified techniques to detect damages in structural effectively. Once the defect becomes critical, depending on the structure mechanical properties, the deterioration process can happen in an accelerated manner, decreasing the structural capacity to resist the design requirements.

Figure 1.1 – Disasters caused by the rupture of the Mariana and Brumadinho dams.

(a) Mariana town devastated after dam rupture.

(b) Trail of destruction left by the rupture of the misty dam.

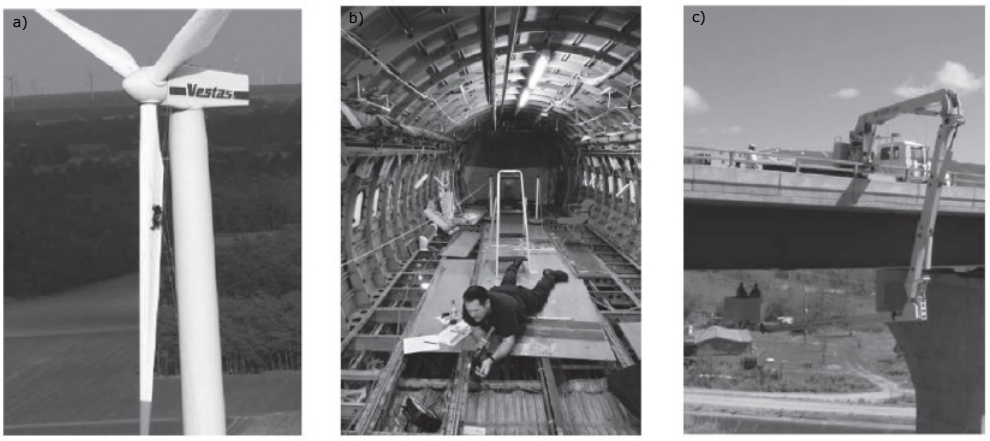
1.1 Background and motivation

Structures are subject to damage in several ways like improper use, natural degradation due the use time, errors in the design phase, as well as in manufacturing and assembly, and factors related to the environment such as earthquakes, for example (PIAZZAROLI, 2019). The damage is, therefore, intrinsic to the structures, associated with changes in their physical properties, which can increase over time, reducing performance and increasing the chances of sudden failures (MEDEIROS et al., 2018). Thus, it becomes essential to develop techniques to allow the identification of defects at an early stage, evaluating the life of the structure, so that necessary repairs can be carried out. One of the most used methods in the early stages of a visual inspection. However, it is limited to some types of structures and damages, motivating the development of more advanced techniques.

Non-destructive Techniques (NDT) as eddy current, radiography, and acoustic emission gained attention and are commonly used. However, these techniques require prior knowledge about the existence and location of the damage, and a trained professional to correctly handle the equipment, impacting downtime and rising maintenance costs. To find an efficient solution, researchers have concentrated efforts to develop a methodology able to detect and localize the damage as early as possible, through continuum monitoring, evaluating its severity, and predicting the residual lifetime, called Structural Health Monitoring (SHM) (FERDINAND, 2014). Fundamentally, the structure is observed along the time, using periodically spaced dynamic response measurements, identifying damage-sensitive features, and statistical analysis of these features to determine the current state of system health (FARRAR; WORDEN, 2012). Hence, SHM can replace the traditional maintenance strategy, minimizing the numbers of inspections based on conventional *Non-Destructive Techniques* (NDT), as Fig. 1.2 shows, reducing costs and operational time required. In the case of composite structures, for example, SHM can be even part of the structure, using piezoelectric sensors inserted between layers (MEDEIROS et al., 2015).

Composite materials were widely used in the industry, much faster than anticipated in

Figure 1.2 – An overview of potential fields for SHM applications: a) Inspection of wind turbines b) Inspection of aeronautical fuselages c) Bridge Monitoring.



Source: Gomes et al. (2018)

the 1970s, especially in the field of experimental and military aviation. According to Zanatta (2012), researches and engineers considered unlikely that any further progress will be made in the area of materials for aeronautical structure. Thus, composite materials bring together the main characteristics requested by the sector; low weight, high strength, corrosion resistance, and the possibility of manufacturing elements with complicated shapes. However, these materials can exhibit unconventional and complex damages types, like transverse crack and delamination, usually not visible to human eyes. Therefore, it is essential to detect the damage as early as possible to prevent sudden failures with catastrophic consequences. Considering fault-tolerance on structural elements of aircraft practiced nowadays, NDTs are applied periodically as a more conservative alternative to give the reliability required, resulting in high cost for airlines (KATUNIN et al., 2015). In 2017, it was reported that $ 70 billion was spent by airlines for maintenance, repair, and overhaul, with the expectation that this figure will reach $ 115 billion over the next ten years (MICHAELS, 2018). In this scenario, the SHM is a promising technique to perform continuous monitoring in these structures, identifying the damage, evaluating its severity, as well as residual strength, and determining the lifetime of the structure. It minimizes the number of inspections, reducing the maintenance cost, without compromising the reliability.

According to Güemes et al. (2020), SHM consists of three-element keys:

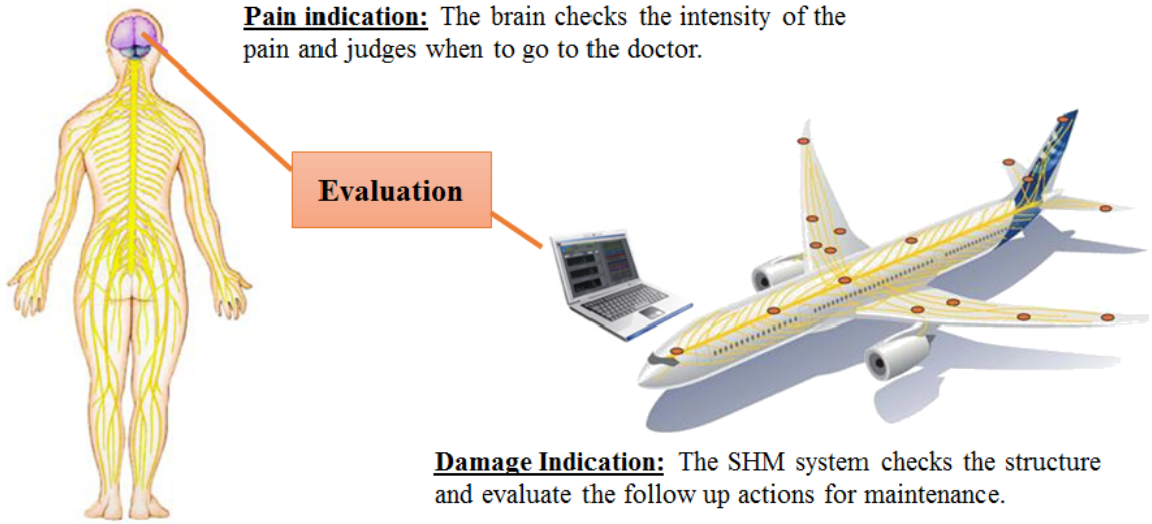
1. **A network of a sensor attached to the structure**: Through the use of a sensor network in the structures, it is possible to carry out an automated inspection essential for continuous data acquisition.

2. **On-boarding data handling and computing facilities**: It is necessary because high computational efforts are required to store the amount of data extracted through the sen-

sors. SHM was feasible when large-capacity of computers were available (in the mid-1980s).

3. **Algorithms**: Necessary to compare the initial state of the structure (healthy), evaluating changes associated with damage, i.e, an appropriate analogy with the human nervous system can be made, according to Fig. 1.3, where changes in the performance appear as disease symptoms in humans.

Figure 1.3 – The analogy between the operation of the human nervous system and a structure SHM (photo from OWI application lab http://www.owi-lab.be)

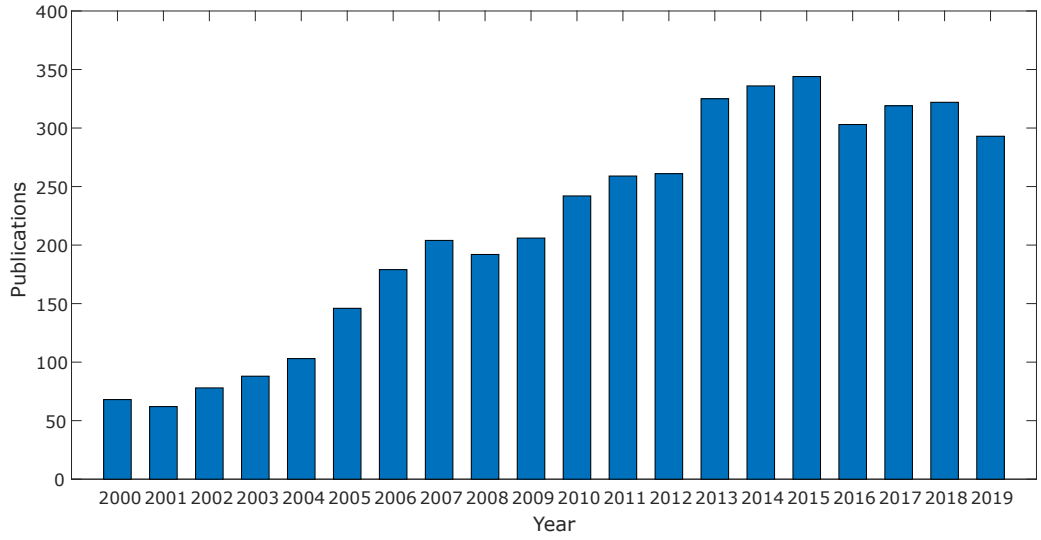


Since the 1970's several studies had been developed to evaluate the dynamic sensitives related to damage, aiming to obtain a more efficient strategy for detection, considering that properties as damping, stiffness, and mass are modified by damage (HUMAR et al., 2006). Recently, machine learning has proven to be a powerful technique for pattern recognition problems. Besides, it is a useful tool in new SHM applications, acting as a bridge connecting a large amount of data from the structure/machinery and systems devices able to make failure predictions (ZHAO et al., 2019). Cawley and Adams (1979) were the pioneers using variations in natural frequency as indicative of damage using a Finite Element (FE) computational model. In the sequence, several advances were implemented, such as the use of Modal Assurance Criterion (MAC) by Allemang and Brown (1982), correlating intact and damaged modes. As an advance, the Damage Location Assurance Criterion (DLAC) was proposed, locating the damage by frequency changes for specific modes (MESSINA et al., 1998). However, according to studies carried out by Doebling et al. (1996), only considerable damage is capable of causing sensitive variations in natural frequencies, making these methods not feasible for more conservative applications. Another factor that hinders the application of these methods is the presence of noises or other external factors. Alternatively, as a solution to these problems, the application of statistical pattern recognition has gained notable attention in the past decade. One of the

main advantages of this type of methodology is that it only requires data from the damaged and undamaged structure to perform the machine learning process (CATBAS; AKTAN, 2002).

The robustness of damage detection is extremely sensitive to the treatment of collected data, which can remove information. Therefore, several studies are being carried out for its consolidation, as illustrated by Fig. 1.4 that exposes growth in the publication of works in the damage detection area using pattern recognition (TANG et al., 2019). Recently, the use of techniques based on Artificial Intelligence (AI) such as *machine learning* and *deep learning* has redefined the state of the art of several kinds of researches, through object identification, image segmentation modeling, and prediction of non-linear systems (ZHAO et al., 2019).

Figure 1.4 – Number of publications per year from 2000 to 2019.



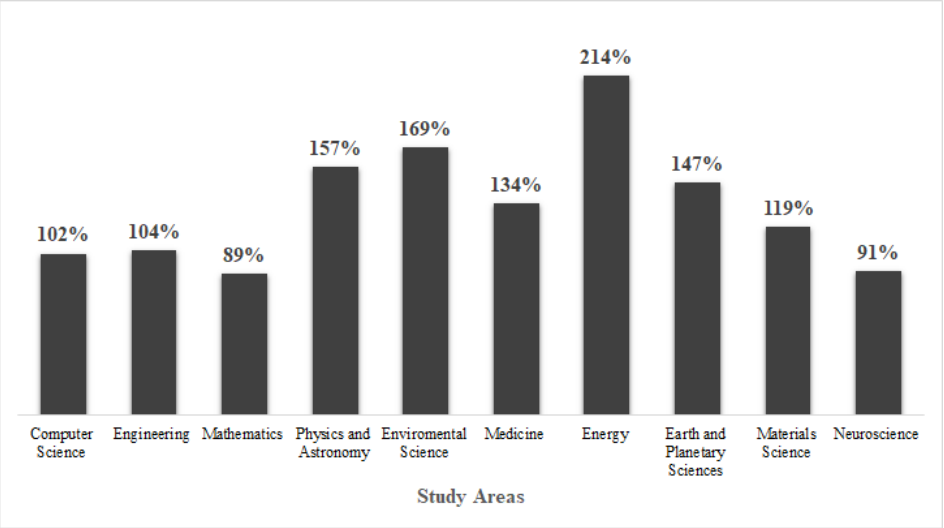
Auhtor's production. Source: Web of Science

With machine learning popularization nowadays, one may think it is a technology recently developed. However, studies in this area started in the 1940s, introduced by McCulloch and Pitts (1943), presenting a logical calculation of the neural networks that unified the studies of neurophysiology and logic. This study assumes that neurons are connected one by one through synaptic connections, following the *all or nothing law*, and have specific adjustable values. Thus, a sufficient number of these units and adjusted synaptic connections set up, any computable function could be represented. Later, Hebb (1949) made a higher contribution in the area, presenting for the first time explicit formulation of a physiological learning rule for synaptic modification. He introduced the famous *learning postulated*, stating that the efficiency of a variable synapse between two neurons is increased by the repeated activation of one neuron caused by others. Another significant contribution was implemented by Rosenblatt (1958), the original methodology of *supervised learning*, though of Rosenblat's Perceptron. Widrow and Hoff (1960) introduced the least square mean algorithm and used it to formulate Adalaine and Madalaine. Amari (1967) used a stochastic gradient method for adaptive pattern classifi-

cation. After the development of the Hopfield and Kohonen networks in 1982, Ackley et al. (1985) developed a stochastic machine known as the *Boltzmann machine*, the first successful realization of a multilayered neural network. Another important step was taken by Rumelhart et al. (1986) through the development of the *back-propagation algorithm*. Broomhead and Lowe (1988) developed a radial basis function. In the 90's, a powerful computational technique known as *Support Vector Machine* (SVM) was developed, gaining prominence in the pattern recognition problems (BOSER et al., 1992; VAPNIK; CORTES, 1995; VAPNIK, 1999). Hinton et al. (2006) coined the term *deep learning* to explain the new algorithms that allow for more in-depth learning, with a higher power of abstraction; so that the machine can distinguish objects, texts, and videos (MARR, 2016). The deep learning advances are more related to the called Convolutional Neural Networks (CNNs) widely used, providing excellent accuracy in object-detection as GoogLeNet proposed by Szegedy et al. (2015). Therefore, based on this high performance, Artificial Neural Networks have gain attention in several fields of science, as (*c.f* Fig. 1.5). From a survey carried out in the SCOPUS database, counting the works published from 2010 to 2020, the growth in the use of artificial neural networks doubled in most of the main research lines studied in the world, with emphasis on the Energy area that presented a growth of 214% in the period assumed. Figure. 1.6 shows a timeline, summarizing important advances in ANN development.

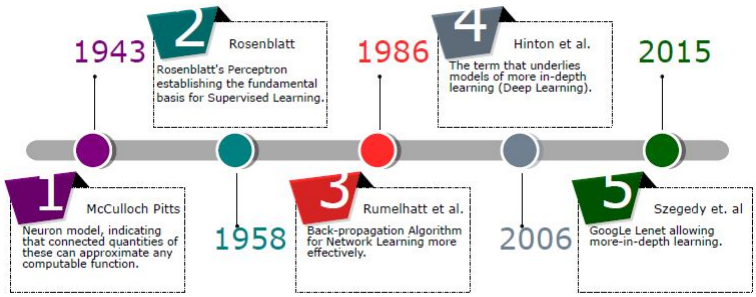
According to Montalvao et al. (2006), Artificial Neural Networks gained space in the detection of damage in structures of composite materials, due to the complexity of correctly modeling all failure mechanisms in a composite. The networks proved to be useful, since they can represent any dynamic system as a black box, via non-linear mapping.

Figure 1.5 – Increased use of neural networks in different areas of study.



Author's produciton. Source: SCOPUS

Figure 1.6 – A short timeline of Artificial Neural Network development.



Auhtor's production.

1.2 State of the Art

Considering the power of Neural Networks to recognize patterns automatically from a data set, through an input/output model. The use of modal parameters to detect damage such as natural frequency, modal shapes and transfer function has proved to be a suitable approach. Thus, Rhim and Lee (1995) verified the feasibility of the use of them as an alternative to detect delamination damage with different sizes in a cantilever beam. In the first step, a computational model was developed, generating some samples which had Frequency Response Functions (FRF) evaluated. After that, an auto-regressive model with exogenous input (ARX) was used. Polynomial coefficients served as input, where good results were found, guiding others' works. Okafor and Chandrashekhara (1996) performed modal testing for intact and damaged composite beams through a built-in piezoelectric sensor, comparing experimental data with analytical issues, where promising results were found. Natural frequencies up to the fourth mode serve as the input of the Multilayer Perceptron (MLP) network to predict the size of damage in delaminated beams. However, only damages with magnitudes as large as $15.24cm$, presented allowable prediction errors. A significant remark was that only the frequency of the fourth mode had shown considerable sensibility for the damage.

Later, another important work was published by Chakraborty (2005), where the first ten natural frequencies were used as input to train the ANN, detecting size, shape, and damage location, through samples generated via the Finite Element (FE) model. In the results, frequencies corresponding to the first two modes decreases. However, it was observed that when the damage was moved away from the center, the other frequencies increase. It was possible to verify the size of damage affects higher modes, and the shape modification causes minor changes. As an alternative, Jiong Tang (2006) had studied a widely used statistical method known as Principal Component Analysis (PCA). The method to extract features from FRFs signals and trains neural networks, towards determining a more robust and efficient damage detection technique for aluminum beams samples at free-free boundary conditions.

During years, many works negligence modeling errors in the Finite Element (FE) model

used to generate vibration data caused by uncertainties present in real models (inaccuracy of physical parameters, non-ideal boundary conditions, finite element discretization, and non-linear properties), which had shown discrepant results among them. Thus, Bakhary et al. (2007) introduced an approach called Statistical Neural Networks (SNN), where uncertainties are applied as noisy in the frequencies and modal shapes. The maximum and minimum intervals used to train four artificial neural networks, and all outputs related to the Stiffness Reduction Factor (SF) are evaluated. In further, the mean and variance of them are determined by evaluating the Probability Damage Existence (PDE) through normal distribution. The probabilistic approach demand four networks, which make the application more computationally expensive. Padil et al. (2017) improved this technique using a non-probabilistic approach with the interval bonds of natural frequencies and modal shapes. He determines a failure surface and evaluates the Possibility Damage Existence (PoDE), and it is not necessary to calculate the means and standard deviation to determine normal distribution. Recently, Padil et al. (2020) presented the best results in this technique though PCA as feature extraction in FRFs.

In this scenario, Wavelet Transform coefficients have success in damage detection applications. Yam et al. (2003) studied and developed an automatic and non-destructive damage detection method for the PVC sandwich plate through theoretical, numerical, and experimental investigation. It was observed that the wavelet energy increases, providing the use of coefficients as a feature extraction technique, setting up a reliable strategy. Moreover, in 2-D analysis plotting, Wang et al. (2017) used images generated from wavelet coefficients detecting damage successfully in a gear-box via Deep Convolutional Neural Networks (DCNNs). Tang et al. (2019) proposed the image processing technique to generate images from time and frequency plots together were used to detect damage with Convolutional layers in the Multilayer Perceptron Model (MLP).

Therefore, several studies have been developed in the last years in the direction to improve feasibility for different damages types and structures, under diverse engineering applications. Table 1.1 shows recent works publishers in this study area.

As it is possible seen, artificial neural networks with convolution layers have gained much attention in SHM applications due to improvements in damage detection and classification. However, more computational efforts are required.

Table 1.1 – Recent works and respective contributions in the state of the art.

Author	Main contribution
Abdeljaber et al. (2017)	The called 1D-CNN is used as a pattern recognition model to detect damage through a frame vibration response.
Liu et al. (2017)	Dislocated Time Series Convolutional Neural Networks (DTS-CNN) used to damage detection in rotating machines.
Khoshnoudian et al. (2017)	Competitive learning using a 2-D PCA approach with competitive learning, detecting damages of different sizes in a truss computational structure.
Piazzaroli (2019)	Use statistical parameters as feature extraction and data reduction to detect damage in civil engineering structures.
Hu et al. (2020)	Locate and identify damage extent on a Finite Element frame of Xi'an Bell tower using the total Wavelet energy rate.
Teng et al. (2020)	Convolutional Neural Networks used to damage detection in a computational model of steel frame structure used by the dynamic response and Modal Strain.
Guo et al. (2020)	To use Convolutional Neural Networks (CNN) for damage detection in beams.
Verstraete et al. (2020)	Semi-supervised and Unsupervised learning approach to detect a fault in rolling bearings through PCA and k-means clustering and GANS-CNN.
Kumar et al. (2020)	Different damage sizes in rolling bearings were detected using Deep Convolutional Neural Networks (DCNN) in image processing by 2-D.

1.3 Objectives and Scope

As the state of art shows, methods based on *deep learning* have been increasingly used to detect damage to diverse structures. Thus, the present work aims to study methodologies that testing different damage detection approaches to analyze bearings and beams of composite materials. The data is extracted from a vibration-based model, where several tools for evaluating characteristics will be used, such as Statistical Parameters (SP), Principal Component Analysis (PCA), and Series-Dislocated (SD), taking into account the particular characteristics of each one. Therefore, the objectives can be split by the items below:

- Proposes a damage detection methodology using the most appropriate feature extraction tool to obtain the more effectively Neural Networks.
- Perform a preliminary study using a data set consolidated in scientific society to validate the methodology, detecting damage on rolling bearings.
- To use data from time-domain and frequency-domain to detect damage on composite

beams by different approaches.

- Identify uncertainties in the manufactured samples and insert this into a computational model, to generalize a network with only computational model data, so, it can extend to detect damage to the experimental beams.

1.4 Thesis Outline

This work is divided into the following chapters:

- **Chapter 1: Introduction -** It provides an overview of the work, showing the motivation of the study and its objectives.
- **Chapter 2: Structural Health Monitoring -** Present the fundamental concepts of the Structural Health Monitoring (SHM) in a general form, showing each step. The theoretical basis for the Vibration-Based Model (VBM). Failure mechanism in the composite beams, and damages on Rolling Bearings.
- **Chapter 3: Composite Materials -** An introduction to important concepts approached in the work like composite materials, classification, types of damage, damage tolerance and general aspects.
- **Chapter 4: Rolling Bearings -** Section provided to define Rolling Bearings and how vibration responses are used in damage monitoring.
- **Chapter 5: Pattern Recognition -** Present a review of Artificial Neural Networks, serving as a basis for studies to be carried out later.
- **Chapter 6: Materials and method -** It presents the data, as well as the strategies established to perform damage detection in bearings and composite beams.
- **Chapter 7: Results and Discussion -**Present and discuss results obtained to the relevant cases studied.
- **Chapter 8: Conclusions -** Describe the potentialities and limitations of the methodology and the recommendation of future works.
- **Bibliography:** Where the references used in the work are presented.
- **Appendix:** It assists in presenting the results and formulations necessary for a better understanding of the text.

Chapter 2

Structural Health Monitoring

2.1 General Concepts

Structural Health Monitoring is defined as a set of activities; acquisition, validation, and analysis of data generated by a system that allows assessing the residual strength of structures during its work, providing essential information to detect damages, prevent failures and, therefore, increasing the reliability (KESSLER et al., 2002). It requires that a data acquisition system be assembled to collect information, evaluating continuously the structural health, based on the performance changes (XIAO, 2012). Thus, the SHM is established as a new maintenance philosophy, in which a system can monitor a structure without having to access it, detecting the damage automatically, allowing a significant reduction in the resources needed for the task.

SHM has been gained much attention in the last years about the Non-Destructive Technique (NDT). Commonly, used in systems where extreme safety is mandatory because it can detect the damage at the initial phase, following its evolution, estimating the life of the structure. Avoiding a higher number of stops inspections, such as airplane wing replacing after some flight hours (FARRAR; WORDEN, 2012). The SHM process was introduced into the pattern recognition field, such that the dynamic response of the structures has specific features that can be evaluated and classified into healthy and damaged by the fundamental steps:

1. **Operational evaluation**: In this initial step, it is necessary to answer some meaningful questions about the system to be monitored such that damage types that the system is subjected to, and the most critical of it, environmental conditions to investigate limitations about monitoring and data acquisition (FARRAR; WORDEN, 2006).

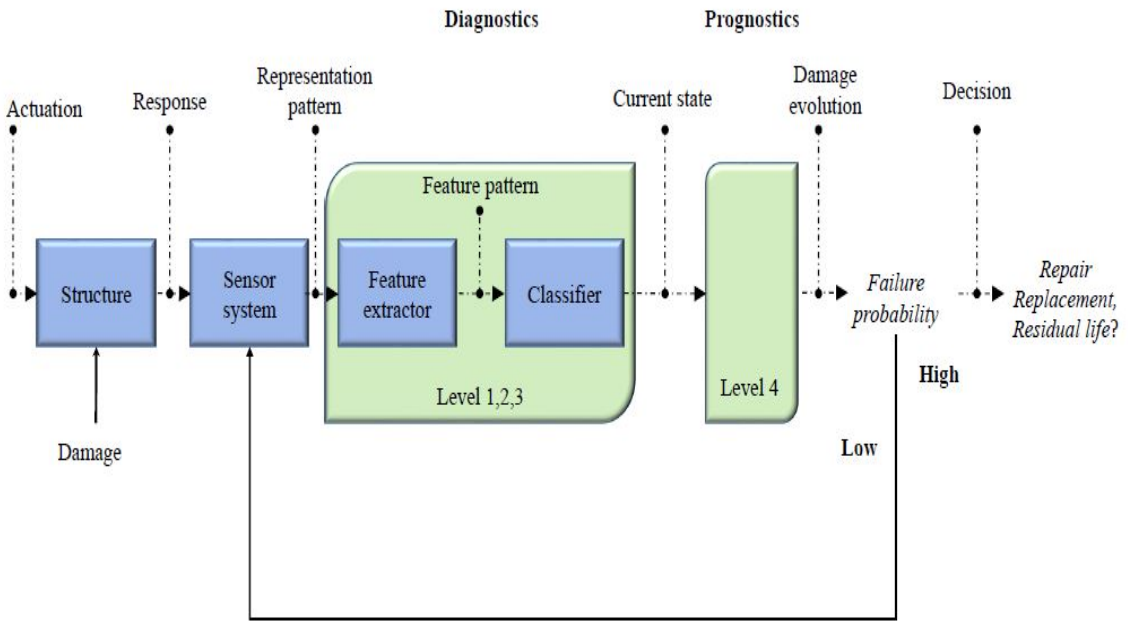
2. **Acquisition data and treatment**: It is associated with the data acquisition about the type of sensors to be used, quantity, positions, specifications about bandwidth choice, and hardware. After extracted, the data should be treated, cleaning up the undesirable pieces of the signals and noises associated, which can make difficult the damage classification after (MONTALVAO et al., 2006).

3. **Features extraction**: This step in the structural health monitoring process receives the most attention in the literature. Feature extraction is the process of identifying damage sensitive parameters from measured data. These damage features are defined in the time, frequency, or modal domain. Information reduction and condensation are also of concern for a large quantity of data, particularly if comparisons of many measurements of the structure are required (OOIJEVAAR, 2014).

4. **Classification**: Finally, through the extraction of features performed in the previous step, an algorithm is implemented to classify samples of the signals collected for different states of the structure, based on the damage sensitivity presented by the features, determining its integrity (FARRAR et al., 2001).

Figure 2.1 shows a simplified process, explaining the SHM, where the two main phases are *diagnostic the current state* and *failure prognostic*. Once that the damage has previously been identified, the lifetime is determined based on residual strength.

Figure 2.1 – The multidisciplinary Structural Health Monitoring (SHM) process.



Adapted from Ooijevaar (2014).

As a fundamental part of SHM, according to Doebling et al. (1996), damage identification is divided into hierarchical four levels:

- **Level 1**: Answer a binary question about the existence of the damage.
- **Level 2**: Identify the location where the damage was found in the geometry of the structure.

- **Level 3**: Quantify the severity of the damage.
- **Level 4**: Predict the life of the structure while it is subjected to those conditions.

2.2 Vibration Based Model (VBM)

In an efficient application of the SHM, it is necessary to identify the damage earliest as possible. So, that corrective actions can be made, minimizing risks associates with disasters, avoiding financial and economic consequences. Methods traditionally used for damage detection involve visual inspection and instrumental assessment. However, it depends on the structure to be accessible everywhere, which can be impractical in some cases when it is complex. Thus, vibration-based models satisfy the accessibility requirements in most of the structures, enabling continuous monitoring. Also, presents good sensitivity to damage because natural frequencies, modal shapes, and damping are directly affected by changes in stiffness and mass that occur due to the presence of damage (HUMAR et al., 2006). Hence, changes resulted from the damage will causes changes in the modal parameters (natural frequencies, mode shapes, and modal damping ratio). Usually, the damage will decrease the mass and stiffness of the structure and increase the damping ratio locally. Among them, mass is less sensitive to the damage, while the damping is most sensitive (ZOU et al., 2000). Thus, the model based on vibrations response, not only serves as a fundamental instrument to evaluate the performance of the structures but also make diagnoses regarding the existence of the damage and its severity concerning an initial or healthy state.

Vibrations can be understood as the oscillatory movement of an object around the equilibrium position, which can be; periodic, non-periodic, and transient. Two elements are fundamental for vibration occur; inertia and stiffness. Therefore, according to Rao (1995), when the system is shifted out from equilibrium, the inertia generates kinetic energy T that is absorbed by the stiffness as potential elastic energy V, establishing a time-dependent energy exchange

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{x}_k}\right) - \frac{\partial T}{\partial x_k} + \frac{\partial V}{\partial x_k} = F_k \quad k = 1, 2, ..., n, \tag{2.2.1}$$

such that x_k are system coordinates and F_k the generalized non-conservative forces. Additionally, if damping is present, an additional term of energy dissipation is added to the energy equilibrium

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{x}_k}\right) - \frac{\partial T}{\partial x_k} + \frac{\partial V}{\partial x_k} = -\frac{\partial \Omega}{\partial \dot{x}_k} + F_k, \tag{2.2.2}$$

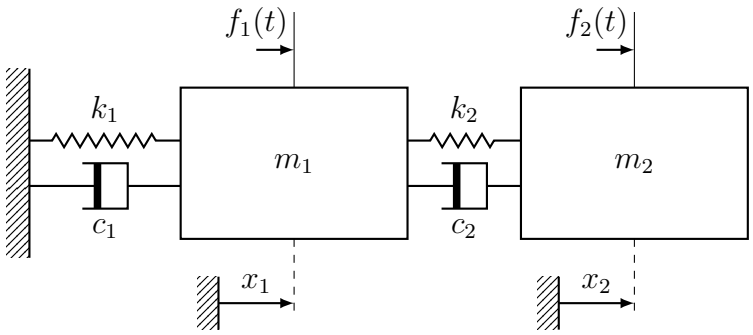
where the left side terms are the rate of variation of the kinetic and potential energies inside the control volume, respectively. On the other hand, right side terms represent, respectively, the rate of variation for the dissipation energy and the external force variation ratio.

2.2.1 Vibration systems with multiple DOFs

The engineering systems are commonly complex to modeling as a continuum one because it requires the solution of a partial differential equations system, which can be quite difficult. For many partial differential equations, analytical solutions do not exist. On the other hand, when the system is analyzed in terms of multiple degrees of freedom, the only solution of an ordinary differential equation is required, which is relatively simple. Hence, it is conventional to simplify the continuum system analyses in terms of multiple degrees of freedom. However, if more degrees of freedom are assumed in the system, the solution becomes more complex, and more mathematical resources are needed. So, the use of numerical methods being convenient in some cases (RAO, 1995).

Degrees of freedom are the number of the time-dependent coordinates used to describe the motion. An example is showed in Fig. 2.2, where x_1 and x_2 are displacements of motion for inertia m_1 and m_2, respectively, defining 2 DOFs. Identifying the number of degrees of freedom is an essential task to analyze the system because it determines the number of ordinary differential equations and, consequently, the number of natural frequencies and mode shapes (KELLY, 2012).

Figure 2.2 – Example of a vibration system with 2 Degrees of Freedom (DOF).



Author's production

After performing a kinematic analysis of the system, defining its respective degrees of freedom, the general equation can be found both by the equilibrium of forces and moments from Newton's laws and by Lagrange's equations due to the energies present in the system. However, whichever method is used, the general equation is

$$[M]\{\ddot{x}(t)\} + [C]\{\dot{x}(t)\} + [K]\{x(t)\} = \{\, f(t)\}, \tag{2.2.3}$$

where $[M]_{N\times N}$ represents the inertia of system, $[C]_{N\times N}$ damping, $[K]_{N\times N}$ equivalent stiffness and $\{f\}_{N\times 1}$ the excitation force (BALACHANDRAN; MAGRAB, 2011), corresponding to the system of N partial differential equations.

2.2.1.1 Undamped systems: natural frequencies and mode shapes

Modal shapes and natural frequencies represent, respectively, displacements between the different degrees of freedom and the specific frequency in which each one occurs. According to Inman and Singh (1994), these parameters are closely linked to mass and stiffness, such that external forces and damping terms can be disregarded from Eq. (2.2.3), resulting in

$$[M]\{\ddot{x}(t)\} + [K]\{x(t)\} = \{0\}, \tag{2.2.4}$$

with general solution

$$\{x(t)\} = \{\phi\}e^{\lambda t}. \tag{2.2.5}$$

Differentiating Eq. (2.2.5), and substituting in Eq. (2.2.4) results in

$$([K] + \lambda^2[M])\{\phi\}e^{\lambda t} = \{0\}, \tag{2.2.6}$$

such that a non-trivial solution can be found by solving an eigenproblem. The eigenvalues are the roots of the characteristic polynomial

$$det([K] + \lambda[M]) = 0, \tag{2.2.7}$$

where $\lambda = -(\omega i)^2$, and ω is the natural frequency. Therefore, for a system with multiple degrees of freedom, each natural frequency corresponds to an eigenvalue λ, defined by

$$([K] - {\omega_j}^2[M])\{\phi_i\} = \{0\}. \tag{2.2.8}$$

2.2.1.2 Damped systems

Depending on the coordinates used to describe the motion, both the inertia and stiffness matrices can present the coupling of inertial and elastic forces, respectively. It is worth mentioning that this is not a basic feature of the system, once that $[M]$ and $[K]$ are constant matrices and non-diagonal, solving the system of second-order ordinary differential equations from Eq. (2.2.3), with constant coefficients may not be a simple task. Thus, one of the ways to solve it is through a transformation of coordinates

$$\{x(t)\} = [\Phi]\{\eta(t)\}, \tag{2.2.9}$$

so that any displacement $x_j(t)(1, 2, ..., n)$ is a linear combination of $\eta_j(t)(1, 2, ..., n)$. The matrix $[\Phi]$ is constant, non-singular and represents a linear transformation. Hence the relationship can be extended to the velocity $\{\dot{x}(t)\}$ as

$$\{\dot{x}(t)\} = [\Phi]\{\dot{\eta}(t)\}, \tag{2.2.10}$$

and acceleration

$$\{\ddot{x}(t)\} = [\Phi]\{\ddot{\eta}(t)\}. \tag{2.2.11}$$

Substituting these terms in the Eq. (2.2.3) results in

$$[M][\Phi]\{\ddot{\eta}\} + [C][\Phi]\{\dot{\eta}\} + [K][\Phi]\{\eta\} = \{f\}. \tag{2.2.12}$$

Premultiplying both sides by $[\Phi]^T$

$$[\Phi]^T[M][\Phi]\{\ddot{\eta}\} + [\Phi]^T[C][\Phi]\{\dot{\eta}\} + [K][\Phi]\{\eta\} = [\Phi]^T\{f\}, \tag{2.2.13}$$

where

$$[M_D] = [\Phi]^T[M][\Phi], \quad [C_D] = [\Phi]^T[C][\Phi], \quad [K_D] = [\Phi]^T[K][\Phi], \quad [Q] = [\Phi]^T\{f(t)\}. \tag{2.2.14}$$

Therefore, the Eq. (2.2.3) can be rewritten

$$[M_D]\{\ddot{\eta}\} + [C_D]\{\dot{\eta}\} + [K_D]\{\eta\} = [\Phi]^T\{f\}. \tag{2.2.15}$$

At this point is necessary observe that the transformation has as objective to produce diagonal matrices $[M_D]$, $[C_D]$ and $[K_D]$ simultaneously, since the system consists of independent equations of motions in generalized coordinates. Hence, if such transforming matrix $[\Phi]$ can be found, then the system of differential equations is decomposed in N independent equations

$$\hat{M}_{jj}\ddot{\eta}_j(t) + \hat{C}_{jj}\dot{\eta}_j(t) + \hat{K}_{jj}\eta_j(t) = Q_j(t) \qquad j = 1, 2..., n. \tag{2.2.16}$$

The transformation matrix exists and is called *modal matrix*, where all *modal vectors* are stored. Therefore, the coordinates $\{\eta(t)\}$ are called *modal coordinates*. Once the orthogonality between *modal vectors* and matrices $[K]$ and $[M]$ is shown in apendix A. Thus, an operation commonly used in the context of linear algebra knows as *diagonalization* can be performed to decouple the system as wished. Hence, each degree of freedom can be analyzed and solved efficiently in terms of the independent differential equations. However, this operation is only available for damping matrix in the cases of proportional damping which is a linear combination between mass and elasticity (MEIROVITCH, 2010).

2.2.2 Modal Analysis

In the past decades, motivated by the demand for lighter, flexible, and yet strength structures, modal analysis has become a good alternative for determining, improving, and optimizing

dynamic characteristics of engineering structures. The fundamental idea is to formulate a mathematical model, considering the stiffness, mass, and other properties as adjustable parameters, aiming to make the model response as close as possible to the experiment. Nowadays, computational modeling based on the Finite Element Method (FEM) has shown as a good alternative to perform the modal analysis once complex structures can be modeled and evaluated accurately. Therefore, natural frequencies, mode shapes, and Frequency Response Functions (RFs) are widely used as a measurement in these analyses because it has sensitive to the parameters previously mentioned (FU; HE, 2001).

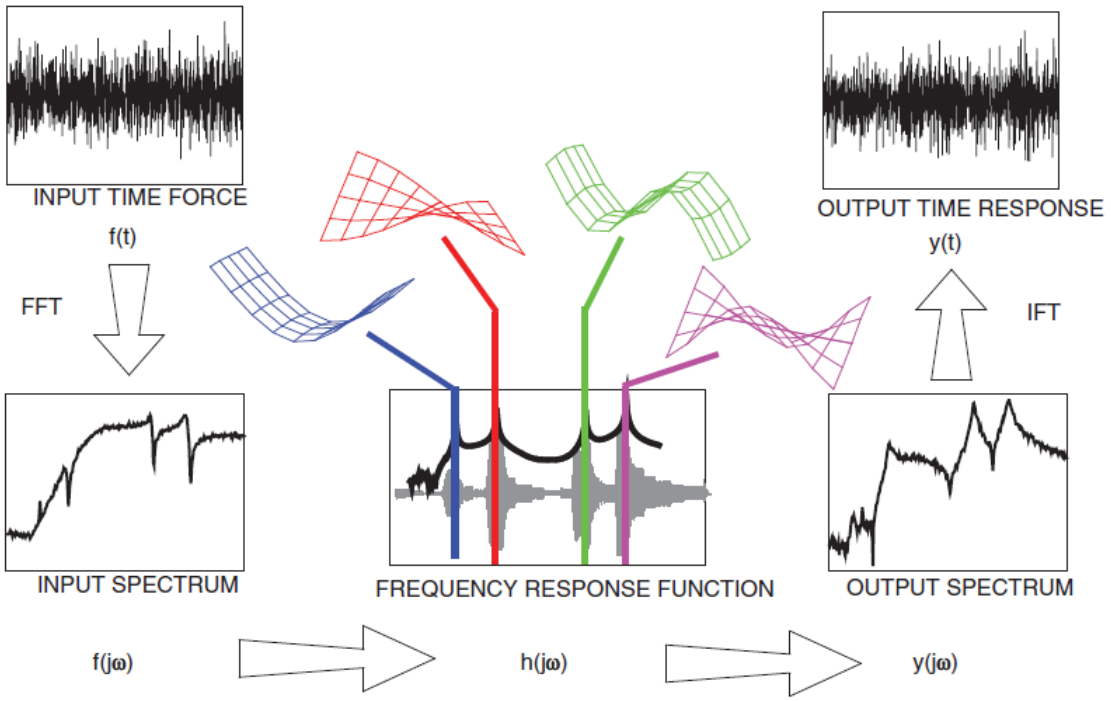
Understanding the dynamics of structures is essential to assess their performance under loads that change over time, damping capacity, friction, and fatigue endurance. However, besides a good mathematical model, it is fundamental to collect the data from structures accurately. To this end, modal testing is the study of a structure, machine, or component subjected to vibration with known excitation, often outside its normal service environment. So, the information is collected in greater detail and precision. This type of test requires a good data acquisition system and subsequent analysis. In summary, the experimental model analysis involves three constituent phases: test preparation, frequency response measurements, and model parameter identification. Test preparation involves the selection of a structural geometry model, which could lead to inaccurate measurement. During the test, a set of Frequency Response Functions (FRFs) data is measured and stored, which is then analyzed to identify modal parameters of the structure (EWINS, 1984).

As an example, Fig. 2.3 shows a simple plate under a random input excitation, and, consequently, the output time response follows the same characteristic. From the time-domain, therefore, it is not easy to observe how each mode shape and natural frequency affect the dynamic response analysis. However, using a Fast Fourier Transform (FFT), transforming the input signal to the frequency-domain, there is a much clearer picture of the input force excitation. In the frequency-domain, the modes of the system (the natural frequencies, damping, and mode shapes) act just like band-pass filters. Each mode ”knows” exactly how to amplify and attenuate the input excitation on a frequency basis, and each mode has a separate effect on the input. However, all the responses from each filter (each mode) are added together to determine the overall response. In this output response spectrum, all modes are not equally excited, because the input force spectrum does not have equal energy at all frequencies (AVITABILE, 2017).

2.2.2.1 Frequency Response Function (FRF)

The Frequency Response Function (FRF) is a relevant measurement in the Modal Analysis because it corresponds to the complete dynamic data of a system, evaluating the relationship among the response for a system and respective excitation located in the structure (KHOSHNOUDIAN et al., 2017). This approach is based on reciprocity between the excitation point (j) and the response (i), analyzed by the matrix $H_{ij}(\omega)$, which has symmetric properties (FU; HE,

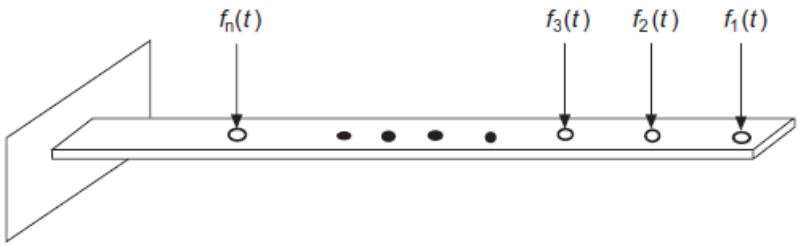
Figure 2.3 – Signal flow diagram showing modal filtering of input resulting in an output.



Source: (AVITABILE, 2017)

2001).

Figure 2.4 – Cantilever beam excited by external forces.



Source: (FU; HE, 2001)

Taking a cantilever beam from Fig. 2.4, excited in n different points by a harmonic force defined by

$$\{f(t)\} = \{F(\omega)\}e^{i\omega t}, \tag{2.2.17}$$

as the system will harmonically vibrate, the displacement corresponds

$$\{x(t)\} = \{X(\omega)\}e^{i\omega t}. \tag{2.2.18}$$

Differentiating and substituting in Eq. (2.2.3), resulting

$$-\omega^2[M]\{X(\omega)\}\cancel{e^{\omega it}} + \omega i[C]\{X(\omega)\}\cancel{e^{\omega it}} + [K]\{X(\omega)\}\cancel{e^{\omega it}} = \{F(\omega)\}\cancel{e^{\omega it}}. \tag{2.2.19}$$

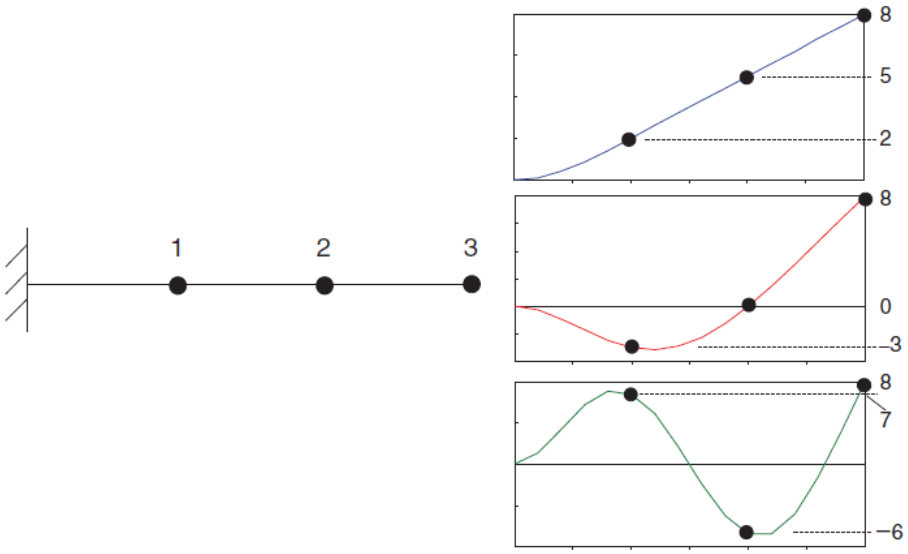
Considering small displacements and linear case, the terms $e^{\omega it}$ are canceled of the equations. Grouping, the relation between input forces $\{F(\omega)\}$ and system response $\{X(\omega)\}$

$$\frac{X(\omega)}{F(\omega)} = \left(-\omega^2[M] + i\omega[C] + [K]\right)^{-1}, \tag{2.2.20}$$

Thus, the frequency transfer function matrix $[H(\omega)]$ relates the input and output in the form $H_{out,in}$

$$[H(\omega)] = (-\omega^2[M] + i\omega[C] + [K])^{-1}. \tag{2.2.21}$$

Figure 2.5 – Model for input-output frequency response function matrix.



Source: (AVITABILE, 2017)

Another important point observed is that matrix $[H(\omega)]$ is complex, allowing FRFs to be evaluated according to real, imaginary, phase, and magnitude, each one with its corresponding characteristics and properties. Figure 2.5 presents an example where three degrees of freedom are used to model a beam and obtain FRFs measurements. Therefore, there is a mode shape for each degree of freedom, resulting in a matrix $[H]_{3\times3}$ (RAO, 1995). Figure 2.6(a) shows only the portion corresponding to the real part, where a peak is accompanied by a substantial decrease for resonance frequencies. As an example, Fig. 2.6(b) consists only of the computed imaginary part, showing that in the diagonal elements of the matrix there are not opposite peaks,

indicating no phase changes for these subsequently modes (2.6(c)). Figure. 2.6(d) shows the anti-resonance effect when the response is collected at an exciting point. Otherwise, unlike the other positions, changes in the phase are observed, as well as opposite peaks for imaginary and minimum between peaks for magnitude.

The phase is evaluated by

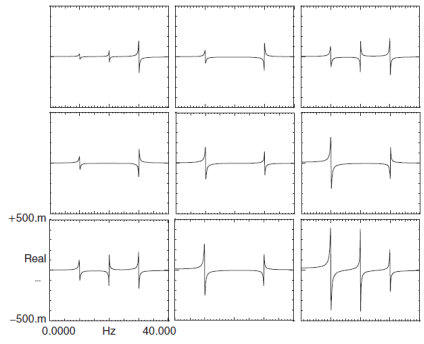
$$[\phi(\omega)] = Tg^{-1}\left(\frac{Img([H(\omega)])}{Real([H(\omega)])}\right), \tag{2.2.22}$$

where there is 180 degrees lag between resonance frequencies corresponding to some modes when the beam is excited and evaluated at different points. The total magnitude is measured according

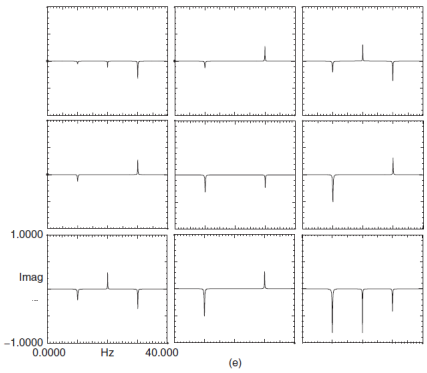
$$[Mag(\omega)] = \sqrt{Img([H(\omega)])^2 + Real([H(\omega)])^2}, \tag{2.2.23}$$

through the square root of the sum of the squares of the imaginary and real part (AVITABILE, 2017).

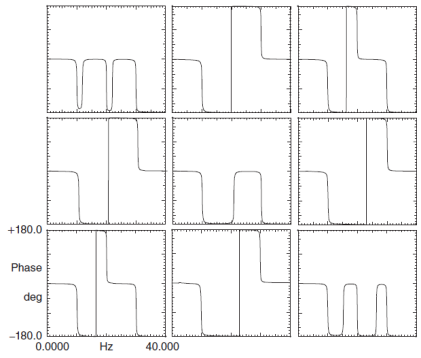
Figure 2.6 – Different measurements of Frequency Response Functions (FRFs).



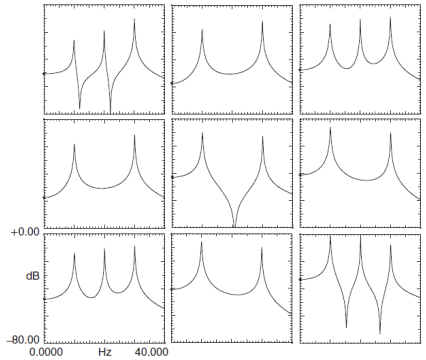
(a) Real part FRF.



(b) $Img([H(\omega)])$.



(c) Phase part



(d) Magnitude dB.

Source: (AVITABILE, 2017)

Chapter 3

Composite Materials

3.1 Introduction

Composite materials are a combination of at least two different materials on a macro-scale, generating another one with improved properties than each one alone (JONES, 1999). In the *fiber-reinforced composites*, for example, the fiber has the function of increasing strength and stiffness of the resulting compound. On the other hand, the matrix serves to keep the fibers together, promoting load transfer among them, and exercising protection against possible environmental weathering (REDDY, 2004).

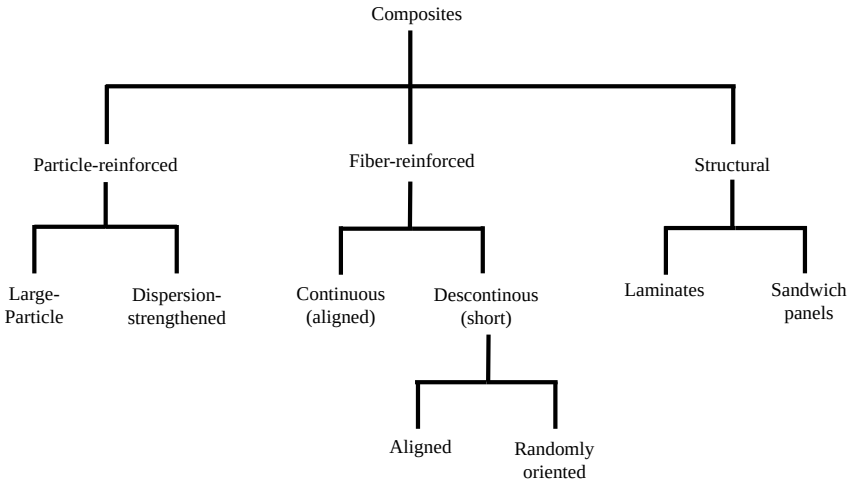
High-performance composites have been widely used in the aerospace industry with epoxy and aluminum matrices, as well as Kevlar, graphite, and boron fibers, due to their high strength and stiffness, combined with low specific density, enhancing the efficiency of these applications. Above all, in some cases, the use of composite materials allows for significant gains in financial terms. Taking into account the competitive aviation market, for example, where one of the most prominent difficulties is to reduce weight without impairing the strength and rigidity of the components (KAW, 2005).

3.2 Composite classification

In general, composites can be divided into natural and synthetic. Among the synthetics, which are most commonly used in engineering, it is possible to enumerate a series of other classifications resulting from the types of fiber and matrix, as well as their arrangement (NETO; PARDINI, 2016). Figure 3.1 shows different configurations, in which the reinforcement and matrix phases can be found in the composite materials. Usually found in the continuous phase, the matrices are made of polymeric, metallic, and ceramic materials (VÖLTZ, 2019). For reinforcements, different architectures with continuous and discontinuous fibers can be used, which significantly influences the mechanical behavior of the composite (IBRAHIM et al., 2015).

The particularities of each composite class and its consequences:

Figure 3.1 – Different classifications of composites found in engineering applications.



Adapted from (CALLISTER et al., 2007)

- **Particulate-reinforced composites**: This type of composite uses *large particles* and *dispersion* as reinforcement. In this context, the term *large particles* is associated with the interaction *reinforcement-matrix*. It can not be deal in the atomic and molecular level, such concrete, for example, where more rigid stones are distributed, locally restricting the displacements caused by the stresses transferred by the matrix. Dispersion follows the same principles of reinforcement than large particles, but the difference is the size of the particles used. In this case, the particles shall have between (0.001,0.1) μm. As an example, the metal alloys in which a volume of metallic or non-metallic particles are dispersed. The reinforcement efficacy depends strongly on the adhesion between the different phases present (CALLISTER et al., 2007).

- **Fiber-reinforced**: These are classified as continuum and non-continuum, depending on the fiber length, fundamentally, which influences the resultant strength. For composite with non-continuum fibers, the mechanics of strengthening the matrix occurs similarly to those of the reinforced by particles. However, the fiber can be aligned with the load axis and provides more specific strength (CALLISTER et al., 2007).

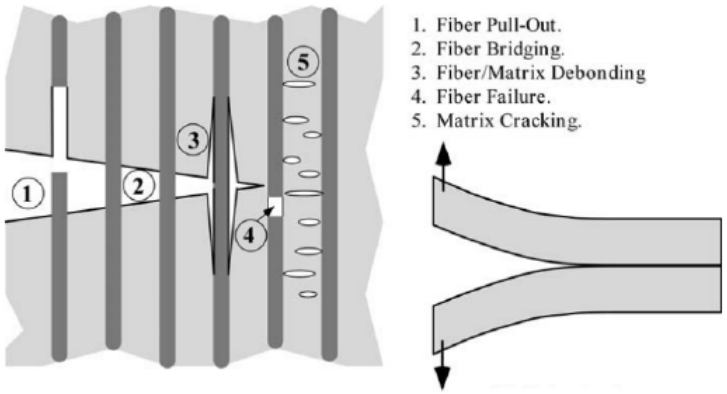
- **Structural**: These are normally used where better performance is mandatory. The properties do not depend on the materials used, but the geometry too. Thus, the principal examples of this class are *laminated* and *sandwich panels* (CALLISTER et al., 2007). According to Jones (1999), the laminate is composed of different materials lamina, with fibers continuum and oriented in a direction, which has a significant influence in its mechanical behavior. *Sandwich Panels* has two outer sheets, or faces, separated and bonded

with a thicker core. The fundamental idea is that most of the load is absorbed by the outer sheets, where commonly a more rigid material is used, such as reinforced plastics, for example. While the core has no structural function, so polymeric materials with low elastic modulus are often found (CALLISTER et al., 2007).

3.3 Damage and failure on composite materials

Even presenting several advantages in engineering applications, especially in the aeronautical and aerospace industry related to metals, this class of materials has the disadvantage of susceptibility to damage, losing much of its structural integrity when they occur. Therefore, damage can occur in different ways; during the processing of the raw material, part manufacture, handling, transportation, storage, maintenance, or in service. The possible damages in this type of material are fiber discontinuities, porosity, delamination, areas poor or rich resin, or in operations involving abrasion, erosion, the impact of hail, stones, and birds. Also, it must be taken into account that the damage is not always visible, but can reduce the resistance of a certain component significantly, and, in cases where damaged components are exposed to the effects of the environment such as temperature, environment, humidity and/or ultraviolet radiation, there may be a significant increase in the degradation of its physical and mechanical properties (REZENDE, 2007).

Figure 3.2 – Failures mechanism on laminated composite materials.



Source: (ANDERSON; ANDERSON, 2005)

Therefore, the main damages and failure mechanism are described:

1. **Pull-out**: It is expected that the laminated has good adhesion between fiber and matrix, providing higher load transfer, which yields strength along the cross-section. However, once these phases do not have suitable adhesion, it is inevitable the fiber slips, characterizing the pull-out phenomenon, causing loss of mechanical properties (TALREJA; SINGH, 2012).

2. **Fiber Bridging**: It occurs when the fiber is very strength. Hence, even the matrix is ductile and has a suitable interface, the crack caused by bubbles or void propagates, such that the fibers form a kind of bridge connecting the intermediate surfaces of the matrix causing stress concentrators (TALREJA; SINGH, 2012).

3. **Fiber/Matrix Debonding**: Thermomechanical loading is a significant contributor to this type of failure. Considering that the resin shrinks during the cooling phase, residual stresses are generated, resulting in the detachment between the phases. The performance in the load transfer along the section of the blade is impaired, which can cause failure. Such defect is extremely difficult to be detected by the naked eye, or by other visual tools (TALREJA; VARNA, 2015).

4. **Fiber Failure**: When failures occur in the weak points of the fibers, redistributing the load in the cross-section in a non-uniform manner, which can overload the neighborhood, causing the breakage of other fibers. As a consequence, the tensile strength of the laminate drops significantly (TALREJA; SINGH, 2012).

5. **Matrix Cracking**: It is one of the most severe types of damage that occur in the composite laminated, which can result in other types of damage such as fiber breakage and delamination. Therefore, cracks start in sharp corners, voids, distortions, or other stress concentrators, which coalesce until the properties offered by the matrix are definitively lost (TALREJA; VARNA, 2015).

6. **Delamination**: It corresponds to the separation of adjacent layers by shearing stresses in the layer plane is considered the most critical damage mechanism for composite design because it is barely vision. Some factors can lead to delamination initiation such as manufacturing by cutting and drilling operations, for example. A geometrical configuration such as curved segments, transitions, sudden changes of section and inclusion, even low-impact (SRIDHARAN, 2008).

3.4 Damage tolerance on composite materials structures

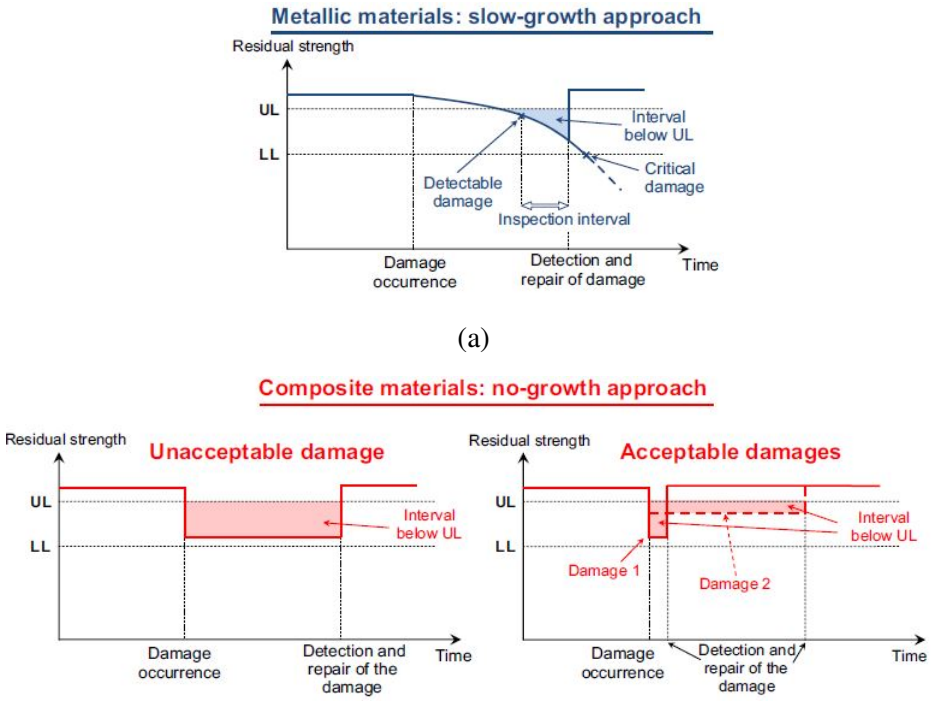
Damage tolerance was introduced in the 1970s for civil aircraft structures. According to Heida and Platenkamp (2012), the fundamental points are that flaws already exist in the structure as manufactured and that the structure may be inspectable or non-inspectable in service. Non-inspectable structures must be designed in such a way that the initial damage will not propagate to a critical size (causing failure) during the design service life. For inspectable structures, the initial damage must grow slowly and not reach a critical size in some predetermined inspection intervals. In this context, the projects are developed from the determination of Limit Load (LL) and Ultimate Load (UL). The first is defined as the load where the structure does not present any damage, while the second admits that damage occurs, without catastrophic failure.

Nevertheless, as the damage is inevitable and can evolve under fatigue solicitations, reducing resistance to a level below UL, the following recommendations must be followed in direction to increase the security of the application:

- It is recommended that even damaged the residual strength of the structure must remain above LL.
- Any damage that leads to a reduction in residual resistance below UL must be quickly detected and repaired so that the resistance stands above UL.

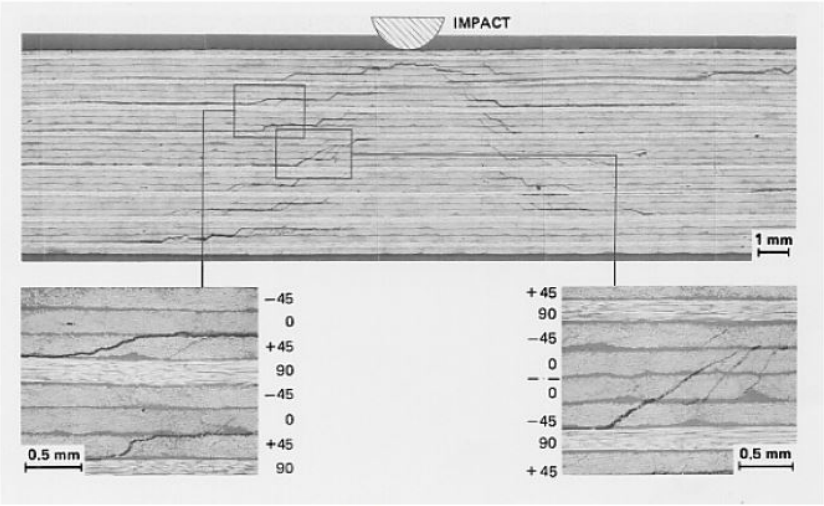
Figure. 3.3 shows the difference between damage tolerance in structures of metallic and composite materials. It can be seen that the damage progression. According to Fig. 3.3(a), it can be seen that the damage evolution in metallic materials occurs slowly, which allows monitoring through periodic inspections so that interventions are carried out before UL is less than LL, and reaches the critical level of damage. On another side, composite structures, when damaged present high residual strength loss instantly (Fig. 3.3(b)), which demands more conservative actions. FAA (2011) determines that the time of a composite material structure below UL should be less critical than for a metallic structure. It is necessary, therefore, to prove that UL always remains above the LL (except in some special cases). Additionally, it is essential that the damage is identified in its initial stage, so that repairs can be carried out, increasing its residual resistance, consequently, the safety of the structure (SILBERSCHMIDT, 2016).

Figure 3.3 – (a) General principle of damage growth and repair of metallic and (b) composite materials.



It becomes very important because, when subjected to impact, composite materials have high sensitivity to the appearance of sub-surface damage that is barely visible (BVID) or even without any visible mark impact, as Fig. 3.4 shows. Impact damage is the most significant type of in-service damage affecting the structural strength, such that a laminate can lose up to 65% of its undamaged static strength (HEIDA; PLATENKAMP, 2012). In this way, composites one up being normally over-sized, presenting residual resistance on the effects of damage.

Figure 3.4 – Cross-section of an impact damaged carbon-epoxy laminate.



Source:(HEIDA; PLATENKAMP, 2012)

Chapter 4

Rolling Bearings

4.1 Introduction

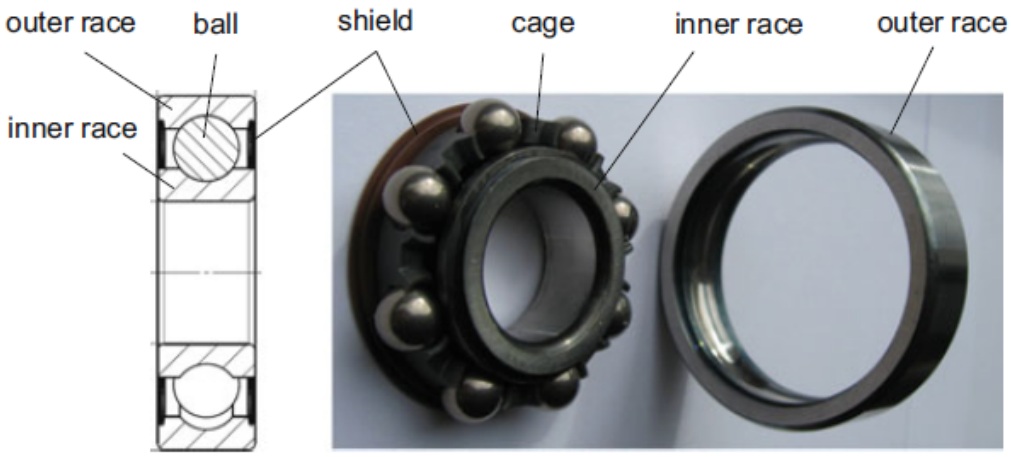
The term *rolling bearing* is used to describe the classes of mechanical components responsible for transfer the main load trough elements in rolling contact rather than in sliding contact. These elements must be designed to fit into space whose dimensions are specified, to receive a specific load, and to have a higher life when operated under the specified conditions. Thus, it can be found in different models, depending on the design load; pure thrust radial loads, pure thrust loads, or a combination of them (SHIGLEY, 2011). Therefore, the most used bearing is balls and roller bearings. It has different setups according to rolling element geometry such as deep groove ball bearings (BB), angular contact BB, cylindrical roller bearings (RB), spherical RB, needle RB and tempered RB. Figure 4.1 shows a ball bearing illustrating the main components. The balls are held in a polyamide cage and are supported by the inner and outer raceways. To keep lubricating grease inside the bearing and to protect the bearing from hard particle contaminants from outside during the operation, two lip seals and shields are installed at both sides of the bearings (NGUYEN-SCHÄFER, 2016).

Present in rotating machinery to perform a rotational or linear movement of various sub-components while reduces friction and stresses. Therefore, rolling bearings have several industrial applications in different fields as the petroleum industry, automotive, and food processing, for example. It is crucial inside the production chain, and when it fails, the higher financial loss can occur. Due to it, during the years, different maintenance techniques were developed to avoid catastrophic failures and control financial losses. There are three primary approaches for bearing fault diagnosis, namely Reactive maintenance, Preventive maintenance, Predictive maintenance, and Proactive maintenance (PATIL et al., 2016).

- **Reactive Maintenance**: This type of maintenance focuses on replacing the component only when it fails, which usually results in reduced safety, unforeseen downtime of production assembly lines, and higher cost.

- **Preventive Maintenance**: On the other hand, preventive maintenance focuses on strate-

Figure 4.1 – Components of deep-groove ball bearings.



Source: (NGUYEN-SCHÄFER, 2016)

gic planning, which imposes some corrective activities at pre-established time intervals, even without knowing the current condition of the bearing. This approach is better than the previous one, but it demands a high cost in maintenance activities and is not necessarily more efficient.

- **Predictive Maintenance**: The strategy is to take action when the rolling bearings show certain behaviors that usually result in a failure or degraded performance.
- **Proactive Maintenance**: As an addition to Predictive Maintenance, Proactive Maintenance focuses on not just identifying the failure, but the root causes. Industrial plants that have matured preventive and proactive maintenance approaches pay a lot of attention to real-time condition monitoring and diagnosis to implement corrective measures promptly, as early as possible.

However, if applied useful tools to damage detection in the rolling bearings, consequently, more effective and productive will be a system where it is inserted, suggesting a study more deeply about damage defects and their features.

4.2 Rolling bearings diagnosis

Bearing defects can be classified in localized and distributed. The localized defects include cracks, pits, and spalls caused by fatigue on the rolling surfaces, while distributed are surface roughness, waviness, misaligned races, and off-size rolling elements. These defects may result from manufacturing errors and abrasive wear from, sometimes, improper installation (CHOUDHURY; TANDON, 1998).

As mentioned, machinery condition is straightly related to defect presence on bearings, and their smooth performance is vital for the proper functioning of the machine. Therefore, several techniques for the detection of defects in rolling bearings are currently available based on changes caused in the mechanical properties of the component, which consequently affects its performance. These include vibration and acoustic measurements like overall level, statistical parameter as kurtosis, spectral analysis, high-frequency resonance, shock pulse, sound, acoustic emission, and oil monitoring such as spectrograph analysis, particle counting, ferrography, and chip detection. Above all, the most efficient technique used was vibration and acoustic measurements though sensibility in the signal response for damaged components (TANDON; NAKRA, 1992).

4.2.1 Time-domain approach

According to Tandon and Choudhury (1999), the simplest approach in time-domain used to rolling diagnosis are to measure the overall Root-Mean-Square (rms)

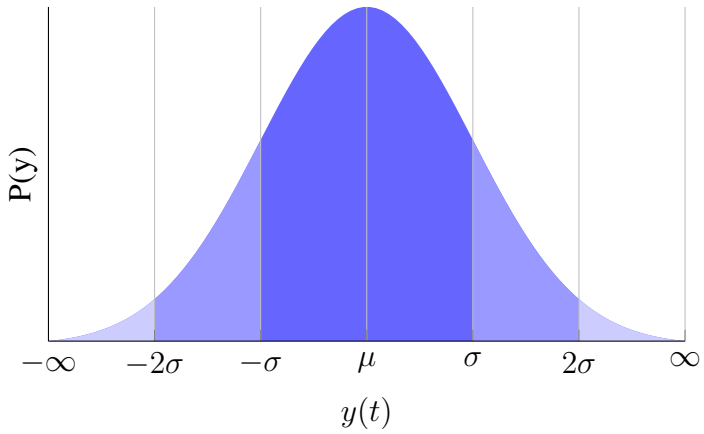
$$rms = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i)^2}, \tag{4.2.1}$$

and Crest Factor

$$X_{Cf} = \frac{y_{peak}}{rms}, \tag{4.2.2}$$

i.e., the ratio of peak value to RMS value of acceleration. However, even though it is widely applied, this method is not efficient in detecting localized defects. Farrar and Worden (2012) shown that vibratory response of a healthy structure excited in the time-domain approximates data generated randomly via normal or Gaussian distribution of probabilities shown in Fig. 4.2.

Figure 4.2 – Gaussian distribution.



Hence, when faults and other anomalies occur, the acceleration increases and the changes

the probability distribution. Such that, statistical moments of data can be used as feature extraction and damage sensitive

$$M_y = \int_{-\infty}^{+\infty} y^i P(y) dy \quad i = 1, 2, ..., m, \tag{4.2.3}$$

where $P(y)$ is the probability density function of instantaneous amplitude y, and i corresponds to the order of the moment. The first and the second moment are the mean

$$\overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i. \tag{4.2.4}$$

And variance,

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \overline{y})^2 . \tag{4.2.5}$$

The third is known as *skewness*

$$\gamma = \frac{\frac{1}{n} \sum_{i=1}^{n} (y_i - \overline{y})^3}{\sigma^3}, \tag{4.2.6}$$

and fourth is called *kurtosis*

$$\kappa = \frac{\frac{1}{n} \sum_{i=1}^{n} (y_i - \overline{y})^4}{\sigma^4}. \tag{4.2.7}$$

Mean and Root-Mean-Square measures the central tendency. When the response vibration signal present irregularities, alters both parameters. Mean is more sensitive to damage existence, while the other to the damage level because the mean value is sensitive to outliers. So a few extreme data points can influence this parameter, and sometimes, it suggested the use of median. The second moment in terms of variance and Standard deviation measures the dispersion around the mean of the times-series amplitudes. For a fixed level of excitation, it will increase with the level of damage. Skewness is related to the asymmetry of the random signal distribution. When this value moves away from zero means that there is non-linearity in the system, which may indicate plastic deformation and, consequently, the appearance of damage. In this way, Kurtosis measures the peaked nature of the response distribution, where 3 is the value expected for a Gaussian response. As with skewness, Kurtosis can also show an inconsistent tendency regarding the magnitude of the damage. Besides, the K-factor

$$X_K = y_{peak} \cdot rms, \tag{4.2.8}$$

is used to assess the deviation from the sinusoidal response in a rotating machine, which has the sensibility to the damage location (FARRAR; WORDEN, 2012).

4.2.2 Frequency-domain approach

An alternative to detect defects in bearings is using the frequency domain. Considering a specific failure type occurs, localized impulses happen significantly increasing the signal amplitude, at a determined frequency, *i.e.*, certain failure type is associated with a specific frequency. A bearing with several rolling elements generates dynamic forces with a single characteristic frequency, called rolling element passing frequency. Whenever a rolling element passes close to a machine part vibrations are excited in part. The damaged bearing spectrum and reference spectrum for undamaged bearing are required to be compared for increase amplitude at the defect frequency (DHAMANDE; CHAUDHARI, 2017). Hence formulas defining the characteristics frequencies for damages located on the rolling elements as well as on the inner and outer races can be found by

$$BPFO = \frac{N \times n}{2}\left(1 - \frac{d}{D}cos(b)\right), \tag{4.2.9}$$

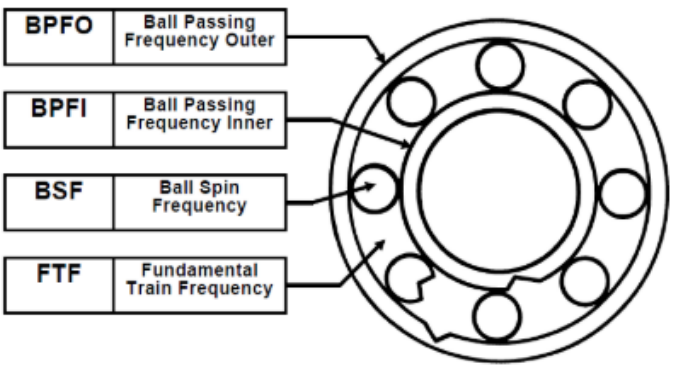
$$BPFI = \frac{N \times n}{2}\left(1 + \frac{d}{D}cos(b)\right), \tag{4.2.10}$$

$$BSF = \frac{N}{2}\frac{D}{d}\left(1 - \left(\frac{d}{D}cos(b)\right)^2\right), \tag{4.2.11}$$

$$FTF = \frac{N}{2}\left(1 - \frac{d}{D}cos(b)\right). \tag{4.2.12}$$

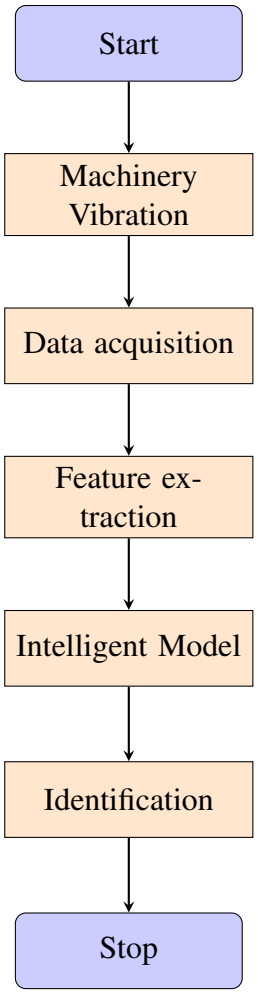
Figure. 4.3 shows the localization of each frequency, where *BFFO* and *BPFI* denote Ball Pass Frequency Outer and Ball Pass Frequency Inner, respectively. *BSF* and *FTF* denote Ball Spin Frequency and Fundamental Train Frequency, respectively. *N* is shaft speed, *n* is the number of rolling elements, *D* is the pitch circle diameter of rolling element, and *b* the contact angle.

Figure 4.3 – Frequencies for each ball bearing component.



Source: (DHAMANDE; CHAUDHARI, 2017)

Figure 4.4 – Main steps of machinery diagnosis.



Author's production

4.2.3 Automated bearing diagnosis

All the assessments showed above depending on the action of a human being for decision making, which requires training a technician who is capable of carrying out the process. Recently, with the advancement of technologies in the development of pattern recognition tools, automatic models have been developed, excluding the need for a human being in the process. Considering that the vibration signature has differences for intact and damaged bearings, the main idea is to use the features extracted via tools as Wavelet Transform (WT), Fast Fourier Transform (FFT) analysis, for example, to automated classify the bearings state. Therefore, it includes some tasks in Fig. 4.4, *i.e.*, given a machine in operation, showing vibration, an associated signal can be obtained, and from there, a feature extraction tool is applied, compressing the information so that it is used to train a computational model. After carrying out the appropriate training for this model, and reaching the acceptable level of learning, it can be used to diagnose failures automatically (SUN et al., 2017).

Chapter 5

Pattern Recognition

5.1 Introduction

Pattern recognition can be understood as a process of automatic identification of regularities in a data set, based on the characteristics presented by each class. This is a task that accompanies human beings for thousands of years on the planet and fundamental for its evolution since it is present in the voice, faces, food, weapons, and manuscripts recognition (DUDA et al., 2012).

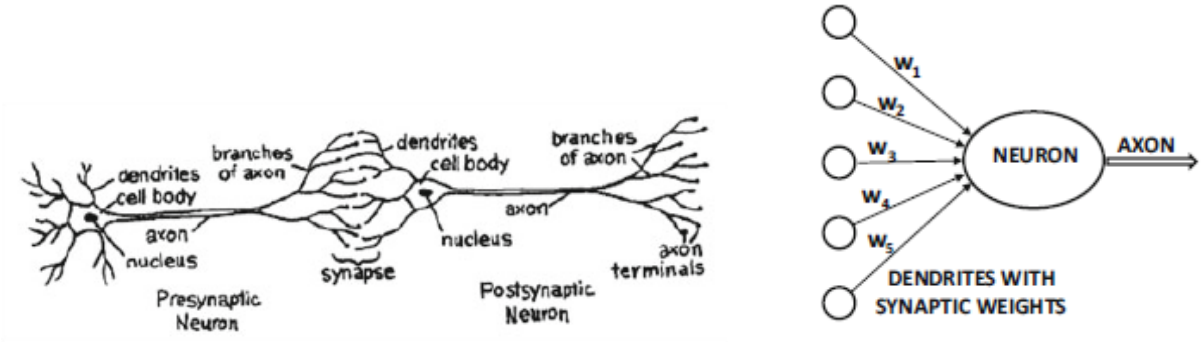
In recent years, the demand for higher productivity in the industry has served as propulsion for the plant automation process, replacing humans by machines in manual tasks. However, with the resources presented by modern computing concerning information processing, the concepts of pattern recognition were assimilated and started to incorporate statistical models capable of training machines to perform activities like humans. In this case, the machine does not only perform a programmed task as in automation but has the autonomy to make decisions based on the *learning* (PRASAD, 2014). According to Bishop et al. (1995), significant remarks are necessary: Data processing and feature extraction are crucial in solving pattern recognition problems. However, when an inappropriate tool is used, important information may be lost during this phase, making the classifier's task more difficult. Another consideration to use efficiently pattern recognition is that the data set must be large enough to identify the features. However, depending on the type of problem, the number of samples can be a limiting factor such as damaged or destroyed objects and medical tests, where evasive examinations are required.

Regarding the aspects pointed above, in this chapter, a brief review of Artificial Neural Networks will be carried out, in the context of pattern classification problems, showing their main properties, and how it can be used to solve classification problems, with emphasis on the simple *Perceptron* model and *Multilayer Perceptron* (MLP).

5.2 Artificial Neural Networks

Artificial Neural Networks have foundations in several areas such as neuroscience, mathematics, statistics, computer science, and engineering, so it has become a powerful tool for solving complex problems in several areas, which gives it a high degree of multidisciplinary. Modeling, pattern recognition, time-series analyses, signal processing, and control are problems that can be solved by the Neural Networks application, due to its ability to learning from examples of the input data. As studies on the human brain progressed discovers that its processing information happened from a network of units called *neurons*, connected one by another via *axons* (transmission lines) and *dendrites* (receptive zone), the connecting regions between axons and dendrite are called synapses, as shown in Fig. 5.1(a). The synapse is a key piece in information processing, and the most common is the *chemical synapse* that occurs when a pre-synaptic process releases a transmitting chemical substance that diffuses through the synaptic junction between neurons and acts resulting in an electrical output signal (HAYKIN, 2007). Following these ideas, the biological mechanism is represented by an artificial neuron, where weights are attributed for each signal component, simulating the synapses from the biological one (Fig. 5.1(b)). In this way, the input data is propagated to the network output, and the learning occurs from the adjustment of them (AGGARWAL, 2018).

Figure 5.1 – From the biological neuron model to the artificial used in the Neural Networks.



(a) Biological Neuron Network.

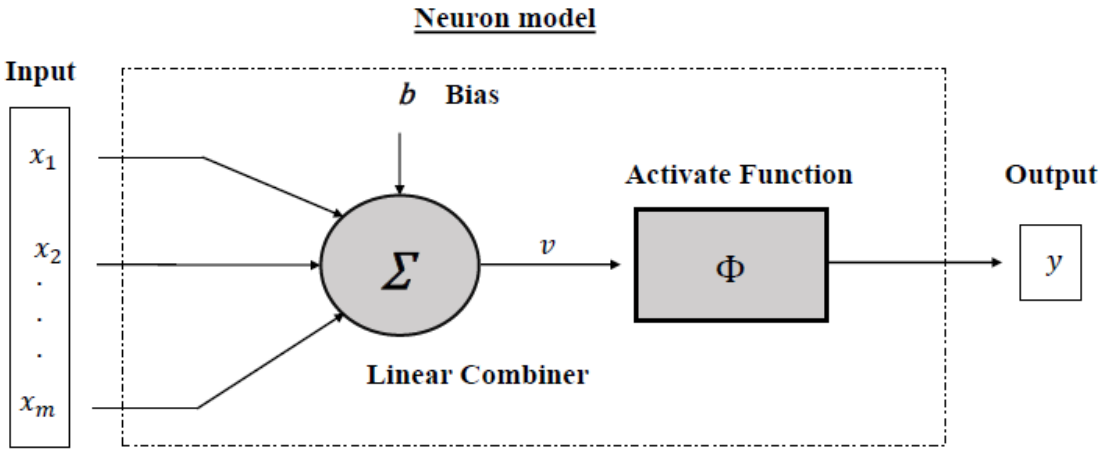
(b) Artificial Neuron Network.

Source: (AGGARWAL, 2018)

5.2.1 Perceptron Model

The simplest network model was proposed by Rosenblatt (1958), known as the *Perceptron Model* or *Rosenblat's Perceptron* shown in Fig. 5.2. It is a *feed-forward network* because the information is transferred from input to output layer, such that the neuron is considered a receptive field able to inhibit or activate the input signal, following the *on or off* theory developed by McCulloch and Pitts (1943). This study was fundamental for the development of more complex models because it introduces signal processing in a neuron.

Figure 5.2 – Perceptron model.



Author's production.

Then, the simple Perceptron works as follow: the input signal is a vector $\{x_1, x_2, ...x_m\}$, such that each element is multiplied to a correspondent synaptic weight and added by a *linear combiner*

$$u = \sum_{j=1}^{m} w_j x_j, \tag{5.2.1}$$

the result of linear combiner u is added to bias b, generating what is called *local induced field* v, providing the potential activation, once that it follows for activation function $\varphi(.)$, and it can be inhibited or excited depending on the output network y obtained

$$v = u + b, \tag{5.2.2}$$

$$y = \phi(v). \tag{5.2.3}$$

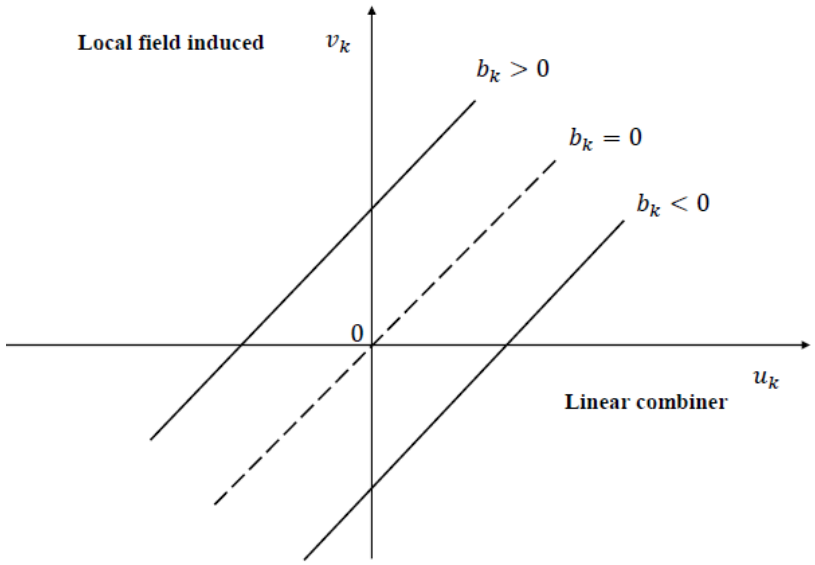
In this way, the weights and bias (parameters that make up the neuron model) can be adjusted, so each the input x_j is correctly mapped to an output y. The bias b is responsible to shift off the hyperplane, modifying the relation among locally induced field v and linear combiner output u as shown in Fig. 5.3. It represents constants in the system.

Mathematically, the neuron's output by *on* or *off* modeling is represented as

$$y_k = \begin{cases} 1, & \text{if } \sum_{j=1}^{m} w_{kj} x_j + b_k > 0, \\ 0, & \text{Otherwise} \end{cases}, \tag{5.2.4}$$

being 1 if its induced local field is non-negative, activating the neuron, and 0 otherwise, inhibiting the signal. It permits classifications between two linearly separable classes, such as AND

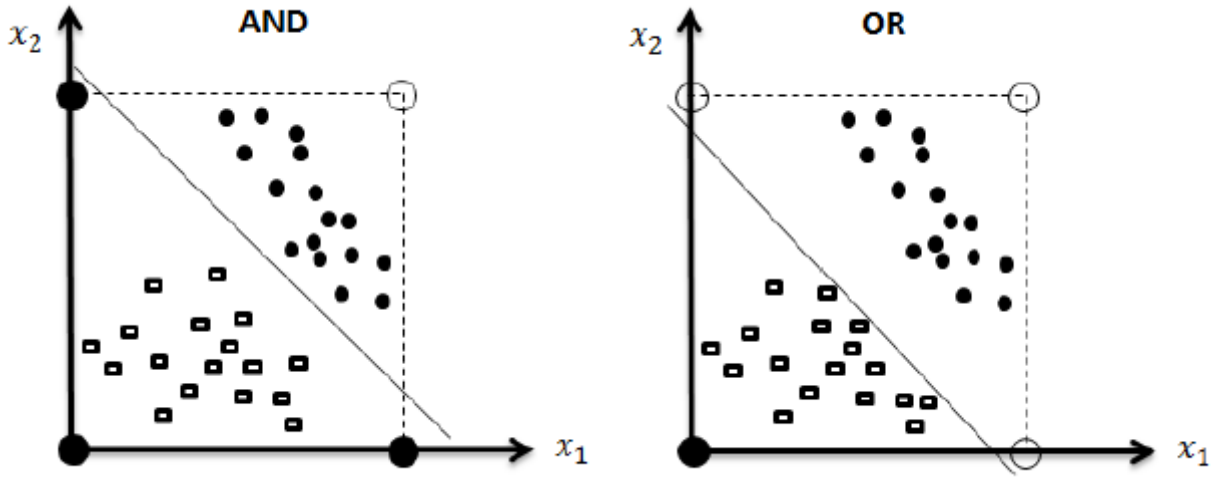
Figure 5.3 – The relation between the induced local field and linear combiners output.



Adapted from Haykin (2007)

and OR problems, presented in Fig. 5.4.

Figure 5.4 – AND/OR classification.



Source: (VÖLTZ, 2019)

5.2.1.1 XOR-exclusive problem

One of the limitations of the perceptron model is the inability to classify input patterns that are non-linearly separable. This situation arises, for example, in the problem of *OR*-

exclusive (XOR), which as a particular case of a more general problem, such as classifying points in a hypercube (HAYKIN, 2007).

Assume a unit square whose vertices correspond to the entries $(0,0), (0,1), (1,1), (1,0)$. Therefore, using the Boolean operator $\oplus$ representing the OR-exclusive function, the following results are obtained

$$0 \oplus 0 = 0,$$

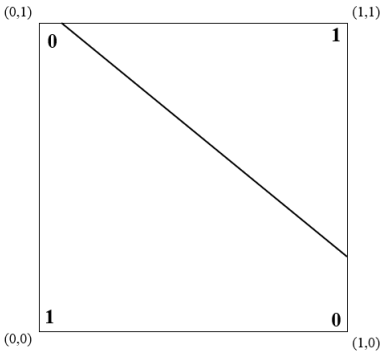
$$1 \oplus 1 = 0,$$

$$1 \oplus 0 = 1,$$

$$0 \oplus 1 = 1.$$

Thus, $(0,0)$ e $(1,1)$ belong to class 0, while $(0,1)$ and $(1,0)$, located at opposite vertices belong to class 1. As previously seen, in the case where two patterns are linearly separable, by adjusting the weights and bias a line establishes a decision boundary, separating the different classes. However, in the case of XOR, it is clear that with the location of points belonging to the same pattern at the vertices, only one line can not be able to correctly define appropriate decision boundary according to the (*c.f.* Fig. 5.5), where a straight line acts to separate different classes, but always there is an error because of the points on the vertex (HAYKIN, 2007).

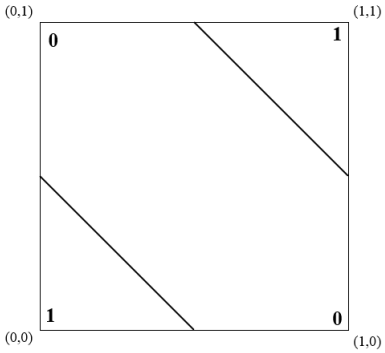
Figure 5.5 – Error in the classification of points located at the vertices of a unit square, characterizing the XOR-problem.



Author's Production. Adapted from (HAYKIN, 2007)

This problem can be solved by using at least one more layer of neurons in the network. In this way, as the information spreads from the input layer to the output layer, this intermediate neuron has the function of adding non-linearity to the system, allowing the following decision boundary in Fig. 5.6 to be formed by the network, solving the problem of XOR.

Figure 5.6 – Correct boundary decision obtained when an additional neuron layer is added to the network, solving the XOR-problem.



Author's Production. Adapted from (HAYKIN, 2007)

5.2.2 Activation Functions

Activation functions play a fundamental role in the exercise of neural networks, defining the neuron's output. However, there are different types of functions, each one with specific properties that must be considered to solve a problem. In general, they can be classified into partially differentiable and fully differentiable Silva et al. (2010).

5.2.2.1 Partially differentiable activation functions

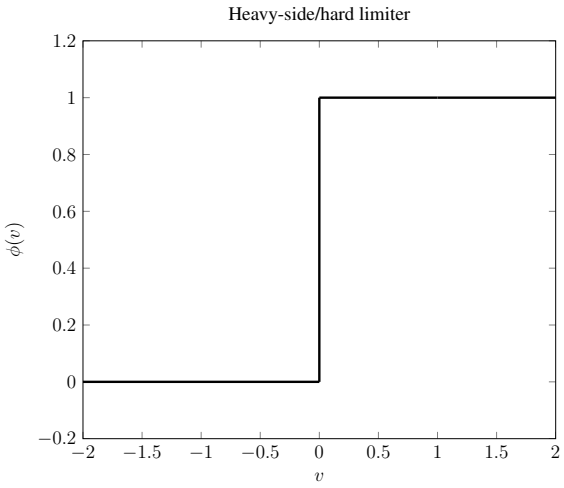
The main property of this type of function is that it does not have a first-order derivative at some points throughout its domain, as it has discontinuities. The most common are step, bipolar step, and a symmetric ramp function. A more detailed explanation of each is provided below:

- **Heavy-side/hard limiter**: For this function, positive unit values are assumed when the neuron activation potential is greater than or equal to zero. Otherwise, null results will be obtained as

$$\phi(v) = \begin{cases} 1, & \text{se } v \geq 0 \\ 0, & \text{se } v < 0 \end{cases}, \tag{5.2.5}$$

  corresponding to Fig. 5.7.

Figure 5.7 – Heavy-side/Hard limiter.



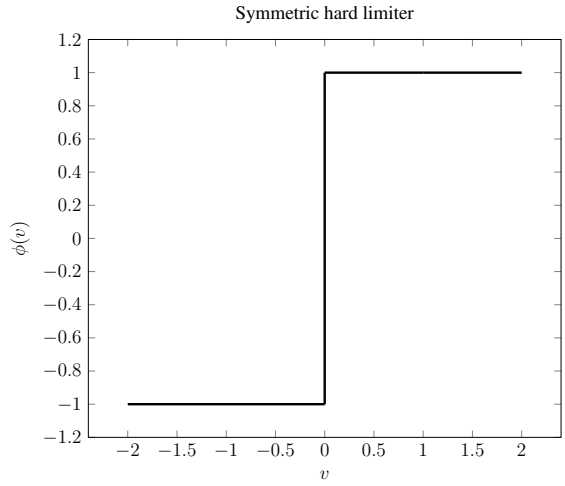
Author's Production

- **Symmetric hard limiter**: It is similar to the *heavy-side* function. However, it allows a negative unit to be evaluated when the function domain is also negative. Mathematically can be described as

$$\phi(v) = \begin{cases} 1, & \text{se } v > 0 \\ 0, & \text{se } v = 0 \\ -1, & \text{se } v < 0 \end{cases}, \tag{5.2.6}$$

  followed by the graphic representation in Fig. 5.8.

Figure 5.8 – Symmetric hard limiter



Author's Production

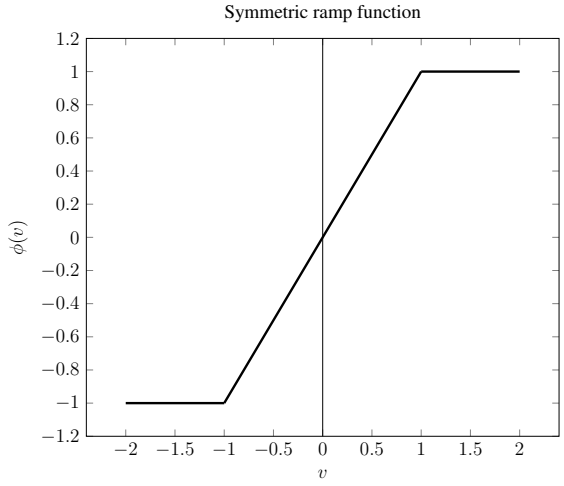
- **Symmetric ramp function**: The values returned are equal to the values of the activation

potentials themselves when they are defined in the range $[-1, 1]$, being restricted to the limit values otherwise. In mathematical notation, therefore:

$$\phi(v) = \begin{cases} 1, & \text{se } v > 1 \\ y, & \text{se } -1 \leq v \geq 1 \\ -1, & \text{se } v < -1 \end{cases} . \tag{5.2.7}$$

It can be seen in a graphic representation, as shown in Fig. 5.9.

Figure 5.9 – Symmetric ramp function.



Author's Production

5.2.2.2 Fully differentiable activation functions

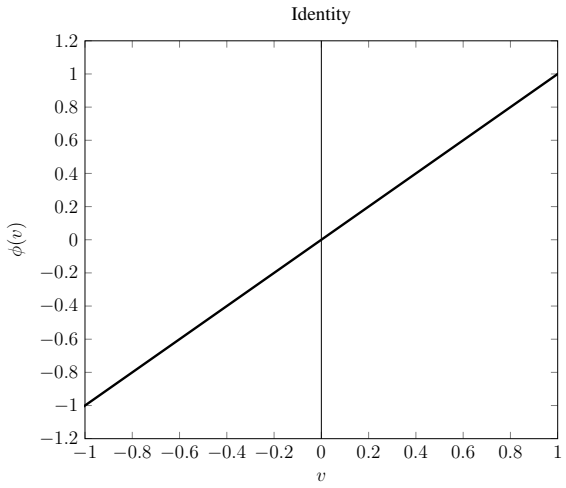
The fully differentiable activation functions are those that present first-order derivatives for all points in the domain and has the important role of the apply non-linearity to the system. Commonly, in applications in neural networks, the logistic, hyperbolic, and linear functions are employed.

- **Linear Function or identity**: The linear function results in output values identical to the activation potentials v. It is the most simple differentiable activate function, commonly used in output neurons when the target is a real value. Mathematically it is defined as,

$$\phi(v) = v, \tag{5.2.8}$$

and represented by a straight line in Fig. 5.10

Figure 5.10 – Identity.



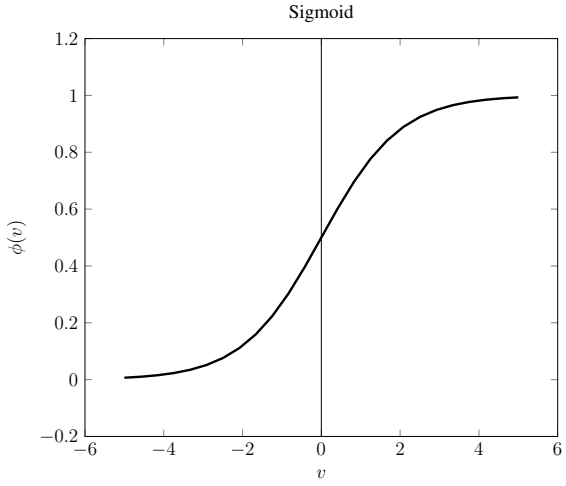
Author's Production

- **Sigmoid** (σ): The sigmoid activation has real numbers in an interval $[0, 1]$, helps perform computations that should be interpreted as a probability and has the sensibility to β used, then, for higher values, it will be closer to heavy-side. This function is written as

$$\phi(v) = \frac{1}{1 + e^{-\beta v}}, \tag{5.2.9}$$

  and represented by the curve seen in Fig. 5.11.

Figure 5.11 – Sigmoid (σ)



Author's Production

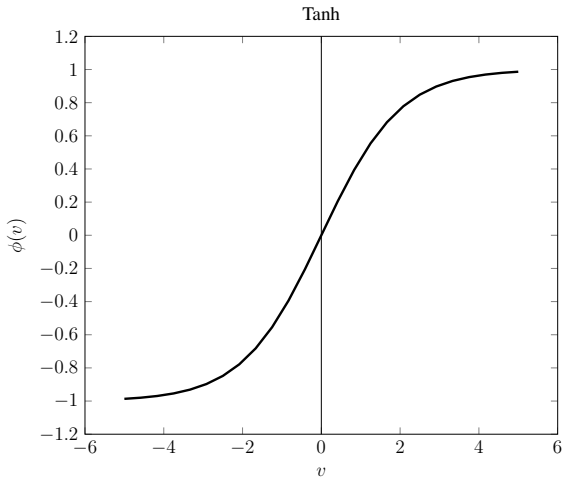
- **Tanh**: As an alternative to the logistic function, this type of function presents values in the interval $[-1, 1]$, where the parameter β influence the curve similar to the one previously demonstrated. However, in this case, it is closer to the *symmetric hard limiter*. The

$tanh$ function is recommended when are desired both positive and negative output. It is represented by the equation

$$\phi(v) = \frac{1 - e^{-\beta v}}{1 + e^{-\beta v}}, \tag{5.2.10}$$

and the following Fig. 5.12.

Figure 5.12 – Tanh.



Author's Production

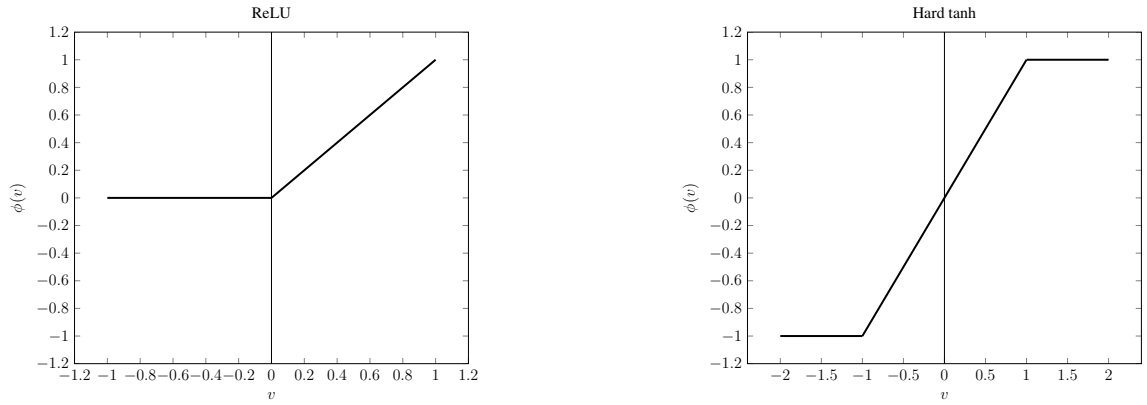
Sigmoid and *tanh* are the most used functions to insert non-linearity into artificial networks. However, in the last years, according to Aggarwal (2018), several piecewise linear activations functions have become popular as *ReLU* and *Hard tanh* because of the ease provided to train complex networks. Respectively, their functions are mathematically represented by

$$\phi(v) = max\{v, 0\}\text{(Rectfied Linear Unit [ReLU])}, \tag{5.2.11}$$

$$\phi(v) = max\{min[v, 1], -1\}\text{(Hard } \textit{tanh}\text{)}. \tag{5.2.12}$$

Their representations are exposed in Fig. 5.13.

Figure 5.13 – Activation functions that have become popular recently.



(a) ReLU. (b) Hard $tanh$

Author's Production

5.2.2.3 The Softmax Function

The softmax activation function forces the output of the neural network to represent the probability that the input falls into each class. Without the softmax, neuron's outputs are simply numeric values, with the highest indicating the winning class. In the softmax equation below, i represents the index of the output neuron calculated, and j all neurons in the level/group, while the variable z designates the array, which contains the information of other output neurons

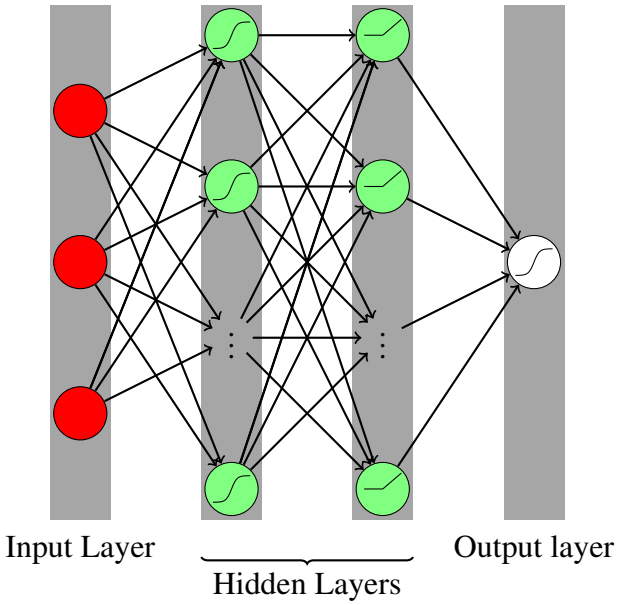
$$\phi_i = \frac{e^{zi}}{\sum_{j \exists e^{zj}}}. \tag{5.2.13}$$

In practice, to solve pattern classification problems using neural networks requires the use of this type of function in the neurons that make up the output layer, when partially differentiable functions are used. Differently from previously activation functions presented, softmax is only used in the output layer and uses the information of other neurons directly (HEATON, 2015).

5.2.3 Multilayer Perceptron (MLP)

As previously seen, Perceptron networks with only one neuron layer for processing information from the input signal are, in a way, limited for solving complex problems. Therefore, the approach used to make networks more robust and able to be applied for more complex problems is inserting more layers and neurons, defining a new concept called the *Multilayer Perceptron* (MLP) model. As a fundamental characteristic, this type of network has at least one hidden layer between the input and output layers. Also, it allows the existence of more than one neuron in the output layer, which gives the ability to solve extremely complex problems (SILVA

Figure 5.14 – Classical Multilayer Perceptron Model.



Author's Production.

et al., 2010).

MLP can be understood as a generalization of the Perceptron model, and thus, it has the characteristic of being a network of the type *feed-forward*. In addition, it is important to highlight three other characteristics:

1. The neurons of this network include a non-linear activation function, inserting non-linearity, which is an important property for learning.

2. The hidden layers in MLP enable the network to learn complex tasks, progressively extracting the most significant features from the input patterns.

3. The networks are highly connected, and the modification of the weights of a neuron changes the whole set.

The combination of these characteristics, leverage MLP networks to solve hard problems. However, they are responsible for some deficiencies presented by the model. The distributed non-linearity and high connectivity between neurons make a theoretical analysis of multilayer perceptron difficult. Second, the use of hidden layers considerably increases the size of the solution space, which can make it difficult to learn the network (HAYKIN, 2007).

5.2.4 Supervised Learning

This type of learning, also known as *learning with a teacher* is based on feeding the network in pairs of data, contenting the input, and correspondent output

$$[Data] = \left[\underbrace{\{x_1, ..., x_k\}}_{Input}; \underbrace{\{z_1, ..., z_k\}}_{Output} \right] \qquad k = 1, 2, ...n. \tag{5.2.14}$$

From Eq. 5.2.1-Eq.5.2.3, using the weights $\{w\}$ and bias $\{b\}$ through feed-forward network, the output a_n is evaluated for a MLP. Therefore, it is possible to establish an error function to measure the distance between predicted a_n and expected z_n values. Thus, supervised learning concerns to adjust the bias and weight values in direction to minimize the error evaluated.

There are different ways to evaluate the error. However, one of the simplest commonly used is the mean square error

$$MSE = \frac{1}{2}\sum_{k=1}^{n}(z_k - a_k)^2 = \frac{1}{2}\|\mathbf{z} - \mathbf{a}\|. \tag{5.2.15}$$

This function evaluates the distance among z and a in the euclidean field by the *norm-p* for $p = 2$. Then, with the parameters determined for each neuron, in each layer, the final output z is evaluated, and compared to the expected result a. The error obtained in the last layer returns to the previous, so the adjustments of the weights and bias are carried out in proportion to the sensitivity of the function, about to the weights and thresholds through the so-called *back-propagation* algorithm (BISHOP, 2006). This technique feeds back the error in the network, from the neurons of the output layers to the intermediate ones, allowing the weights and bias to be adjusted. Thus, the node delta δ_i is the value calculated for each node, and they take into account the error function for the neural network. Thus, for the mean-square function

$$\delta_k = (z_k - a_k)\,\phi_k'. \tag{5.2.16}$$

It is possible to observe that when the quadratic error function is used, the appropriate adjustment of the weights can become slow, due to the small steps coming from the very small gradient (ϕ_k') as it approaches the minimum error (BISHOP, 2006). Hence, Cross-entropy (CE) is an interesting alternative to fix this problem

$$CE = -\frac{1}{n}\sum_{k=1}^{n}(z_k ln a_k + (1 - z_k) ln(1 - a_k)). \tag{5.2.17}$$

Using CE, the node delta turns less complex than the mean-square error, because does not have an activation function derivative component

$$\delta_k = (z_k - a_k)\,. \tag{5.2.18}$$

Thus, this error function will require much fewer gradients calculations (HEATON, 2015).

5.2.4.1 Back-propagation algorithm

The learning algorithms are optimization-based mechanisms to minimize the error evaluated in the network, by updating the weights corresponding to the neurons. Therefore, the most used is the back-propagation. Thus, considering as an optimization problem, several methods can be applied to solve the problem, the most common among them is the Steep Descent method, which uses the negative direction of gradient function for minimization

$$\{X\}_{i+1} = \{X\}_i - \lambda_i\{\nabla f\,(X_i)\}, \tag{5.2.19}$$

where f is the function to be minimized, and λ_i, the step which $\{X\}_{i+1}$ is updated (RAO, 2009). In *deep learning*, the function will be the error, bias turns out a constant input $x_0 = +1$ with $w_{j0} = b_j(n)$, n is the number of the epoch which the weights were updated for a complete evaluation of the dataset and η called *learning rate*. Therefore, the Eq. (5.2.19) can be rewritten

$$\{w^{t+1}\} = \{w^t\} - \eta\underbrace{\{\frac{\partial E(n)}{\partial w}\}}_{\Delta w}, \tag{5.2.20}$$

in general, $\{\Delta w\}$ is the proportion of updating the weights, known as the delta rule. The partial derivative is treated as a sensitivity factor, and according to the Steepest method. Applying the chain rule,

$$\frac{\partial E(n)}{\partial w_{ki}} = \frac{\partial E(n)}{\partial e_k(n)}\frac{\partial e_k(n)}{\partial a_j k(n)}\frac{\partial a_k(n)}{\partial v_k(n)}\frac{\partial v_k(n)}{\partial w_{ki}(n)}, \tag{5.2.21}$$

whose $e_j(n)$is the signal of the error obtained at the output of the neuron k, defining the value of the instantaneous energy of the error,

$$e_k(n) = z_k(n) - a_k(n), \tag{5.2.22}$$

which is related to E in the output layer,

$$E(n) = \frac{1}{2}\sum_k e_k^2(n), \tag{5.2.23}$$

and determine the local gradient or node delta δ_k

$$\delta_k(n) = \frac{\partial E(n)}{\partial v_k}. \tag{5.2.24}$$

Therefore, by calculating the successive derivatives explained in the chain rule, the equation corresponding to the minimization direction is defined,

$$\{\frac{\partial E(n)}{\partial w}\} = -e(n)\phi'(v(n))a(n). \tag{5.2.25}$$

However, optimization methods based on the gradient have the disadvantage of being

sensitive to the existence of local minimums, consequently, can offer difficulties during the network training. Then, it becomes convenient to use the *momentum* α to reduce this effect (HAYKIN, 2007). Therefore, the weights update becomes

$$\{w^{t+1}\} = \{w^t\} + \alpha\{w^t - w^{t-1}\} - \eta\{\frac{\partial E(n)}{\partial w}\}. \tag{5.2.26}$$

The idea is to consider how much the synaptic weights change between two successive iterations. Considering the optimization method, many epochs can be taken, which increases the possibility of finding a minimum location along the way. When the current solution is far from the minimum, the difference between them becomes great, the momentary term will act as a brake on updating the weights, in an attempt to avoid convergence to a local minimum. Otherwise, as this value becomes small, approaching the minimum, the *momentum* influence will be almost zero. Thus, the use of this term tends to improve the training convergence of the network (SILVA et al., 2010).

5.2.4.2 Stochastic Gradient

Understood as an optimization process, the network weights adjustments searches for a set of parameters that minimize the error evaluated between the target and network output. However, as the complexity of the problem increases with the number of hidden layers, the amount of data and neurons used, the error is described by a more complex function, consequently, presenting non-convexity properties. In this case, algorithms used for training would have a significant influence on network performance. According to Goodfellow et al. (2016) the use of a stochastic gradient can improve it significantly. The insight is that the gradient can be estimated using a small set of samples. Thus, taken mini-batch of examples $B = \{z_1, z_2, ..., z_{m'}, \}$ from training set

$$\{g\} = \frac{1}{m'}\{\nabla w\}\sum_{k=1}^{m'} E(z_k, a_k, w), \tag{5.2.27}$$

such that $m' < m$, where m is the total samples. Thus, the weights adjustment follows the estimated gradient downhill

$$\{w\}_{k+1} \leftarrow \{w\}_k - \eta\{g\}_k. \tag{5.2.28}$$

5.2.4.3 Algorithms with Adaptative Learning Rate

One of the recent advances in neural network models is the use of adaptive learning, which has less influence on the value adopted for hyper-parameters, such as learning rate and momentum. The main idea of these algorithms is to separate learning rates for each term in the gradient so that they automatically adjust to the natural learning course. The most commonly

used are AdaGrad, RMSprop, and ADAM (GOODFELLOW et al., 2016). The particularities of each one is discussed below, as well as respective algorithms.

- **AdaGrad**: In this algorithm, the parameters are updated according to the square root of the summed squares of the gradient. Therefore, the directions that show the largest partial derivatives have a decrease in the learning rate. The AdaGrad enjoy some desirable technical properties to solve non-convex problems. However, empirically it has been found that - for training deep neural network models - the accumulation of squared gradients from the beginning of training can result in a premature and excessive decrease in the effective learning rate (GOODFELLOW et al., 2016).

Algorithm 1 AdaGrad algorithm.

Require: Global Learning rate η

Require: Initial parameters $\{w\}$

Initialize $\delta = 10^{-7}$ $r = 0$

while stopping criterion not met **do**

Sample a mini-batch of m examples from the training set $\left(\{x^{(1)}\}, ..., \{x^{(m)}\}\right)$ with corresponding targets $\{z^{(i)}\}$

Compute gradient: $\{g\} \leftarrow \frac{1}{m'}\{\nabla w\} \sum_{k=1}^{m'} E(z_k, a_k, w)$

Accumulated squared gradient: $\{r\} \leftarrow \{r\} + \{g\} \odot \{g\}$

Compute update: $\Delta w \leftarrow -\frac{\eta}{\delta + \sqrt{r}} \odot \{g\}$ (Division and square root applied element-wise)

Apply update $\{w\} \leftarrow \{w\} + \{\Delta w\}$

end while

- **RMSProp**: As a property, AdaGrad converges quickly in convex cases, and it has an impact on the performance of the deep neural network train. Thus, RMSProp uses an exponentially decaying average to discard history from the extreme past so that it can converge rapidly after finding a convex bowl as if it were an instance of the AdaGrad algorithm initialized within that bowl. RMSProp is an efficient and practical optimization algorithm for deep neural networks (GOODFELLOW et al., 2016).

Algorithm 2 RMSProp algorithm.

Require: Global Learning rate η, decay rate ρ

Require: Initial parameter $\{w\}$

Require: Small constant $\delta = 10^{-7}$ for numerical stability

Initialize gradient accumulation variable $r = 0$

while stopping criterion not met **do**

Sample a mini-batch of m examples from the training set $\left(\{x^{(1)}\}, ..., \{x^{(m)}\}\right)$ with corresponding targets $\{z^{(i)}\}$

Compute gradient: $\{g\} \leftarrow \frac{1}{m'}\{\nabla w\} \sum_{k=1}^{m'} E(z_k, a_k, w)$

Accumulated squared gradient: $r \leftarrow \rho r + (1-\rho)\{g\} \odot \{g\}$

Compute update: $\{\Delta w\} \leftarrow -\frac{\eta}{\sqrt{\delta + r}} \odot \{g\}$ (Division and square root applied element-wise)

Apply update: $\{w\} \leftarrow \{w\} + \{\Delta w\}$

end while

- **Adam**: Adam is an algorithm for first-order gradient-based optimization of stochastic objective functions, based on adaptative estimates of lower-order moments. The method is also appropriate for non-stationary objectives and problems with high noise and/or sparse gradients. If the function is differentiable, its parameters, gradient descent is a relatively efficient optimization method, since the computation of first-order partial derivatives all the parameters is of the same computational complexity as just evaluating the function (KINGMA; BA, 2014).

Algorithm 3 The Adam algorithm.

Require: Global Learning rate η (Suggested default: 0.001)

Require: Exponential decay rates for moment estimates, ρ_1 and ρ_2 in $[0, 1)$. (Suggested defaults: 0.9 and 0.999 respectively)

Require: Initial parameters $\{w\}$

Initialize small constant δ used for numerical stabilization. (Suggested default: 10^{-8}

Initialize 1st and 2nd moment variables $s = 0$, $r = 0$.

Initialize time step $t = 0$

while stopping criterion not met **do**

$\{g\} \leftarrow \frac{1}{m'}\nabla w \sum_{k=1}^{m'} E(z_k, a_k, w)$

$t \leftarrow t + 1$

Update biased first moment estimate: $\{s\} \leftarrow \rho_1\{s\} + (1-\rho_1)\{g\}$

Update biased second moment estimate: $r' \leftarrow \rho_2\{r\} + (1-\rho_2)\{r\}$

Correct bias in first moment: $s' \leftarrow \frac{s}{1-\rho_1^t}$

Correct bias in second moment:$r' \leftarrow \frac{r}{1-\rho_2^t}$

Compute update: $\Delta w = -\eta \frac{s'}{\sqrt{r'} + \delta}$ (operations applied element-wise)

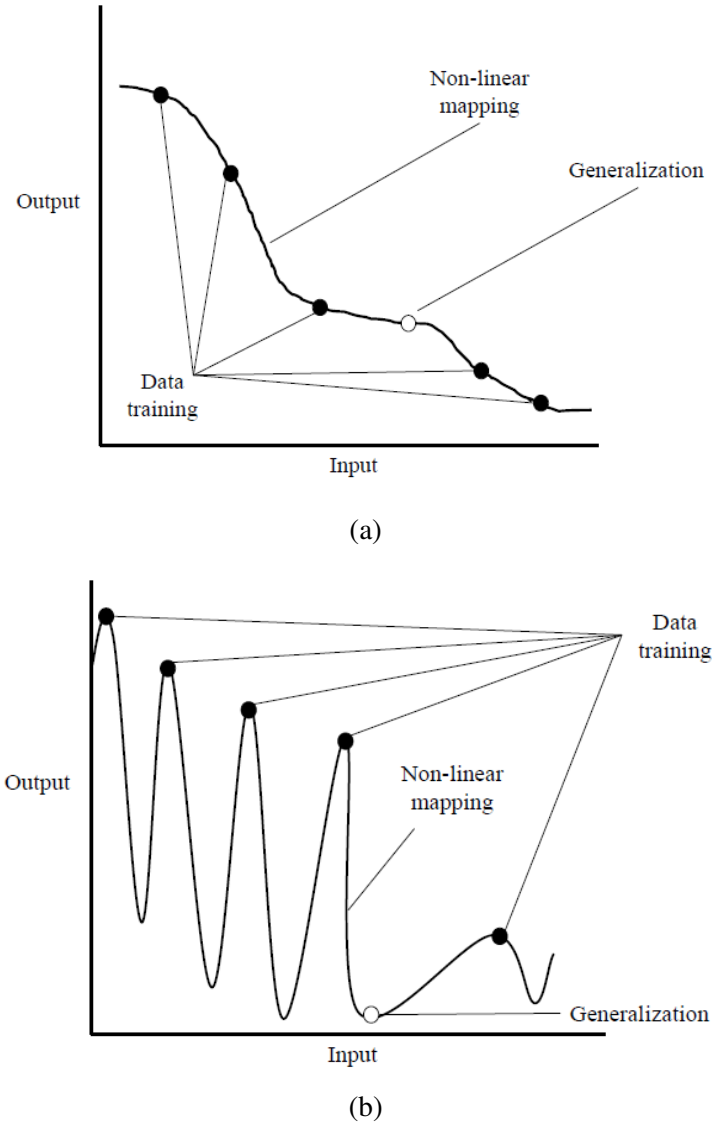
$\{w\} \leftarrow \{w\} + \{\Delta w\}$

end while

5.2.4.4 Generalization

In general, the main objective of a neural network is to make good predictions of a system, to make a *generalization* of the prediction model. The learning process can be seen as an adjustment of the curve that represents the *input/output* mapping over the network. In this way, it can be said that a network was generalized when samples that have not been used in the training step presents correct results. Figure 5.15(a) shows how generalization occurs in a neural network, fitting the training data along a curve, representing an interpolation problem. Therefore, in the case where this curve that represents a non-linear mapping is smooth, good predictions are realized, otherwise, such as Figure 5.15(b), the non-linear mapping is over-fitted, and cannot generalize. Generalization is influenced by factors such as: (1) the size of the training set, and how much it represents the system in question, (2) the network architecture, and (3) the physical complexity of the problem. It must be agreed, therefore, that there is no control over the latter term. Then, for a good generalization to be achieved it will be necessary to use good architecture and representative training set (HAYKIN, 2007).

Figure 5.15 – The generalization problem is seen as an interpolation that performs the mapping between input and output data.



Author's Production: Adapted from (HAYKIN, 2007).

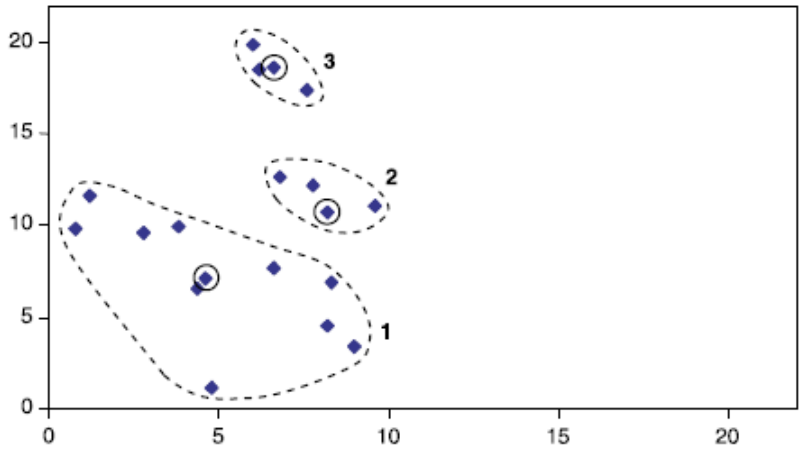
5.2.5 Unsupervised Learning

Different from the previous approach, where the data set to train the classification model used was labeled, defined as *learning with a teacher*, once that in the training phase each input has an associated output. Unsupervised procedures use unlabeled samples, without being told their classes before (DUDA et al., 2012). However, in engineering applications, this approach may be very useful. According to Verstraete et al. (2020), labeling output requests a significant investment, in addition to limiting the model's generalizability, because the solution is restricted to the engineer's knowledge.

Figure. 5.16 shows as *clustering* algorithm work in a dataset, finding groups of items

that are similar through their features. For example, an insurance company might group costumes according to income, age, types of policy purchase, or prior claims experience. In fault diagnosis, a different failure type might be grouped according to the values of certain variables keys (BRAMER, 2007).

Figure 5.16 – Clustering data.



Source: (BRAMER, 2007)

5.2.6 Bias and variance trade-off in ANN application

A common problem in the Artificial Neural Network (ANN) application is to determine the number of neurons and layers needed for a good fitting in the data set and, therefore, a suitable generalization. The key to understanding this problem and knowing how to get around it is to decompose the network error (*supervised-learning*) into two components called *bias* and *variance*. It is convenient to use the sum-of-square error without losing the generality of the conclusions to be taken. Considering that the network has only one output denoted by the function $y(x)$, the sum-of-squares error at the limit of an infinite set of data can be written in the form

$$E = \frac{1}{2}\int \{y(x) - \langle t|x\rangle\}^2 p(x)dx + \frac{1}{2}\int \{\langle t^2|x\rangle - \langle t|x\rangle^2\}p(x)dx, \tag{5.2.29}$$

where $p(x)$ is the unconditional density of the input data, and $\langle t|x\rangle$ denotes the conditional average, or regression of the target data given by

$$\langle t|x\rangle \approx \int tp(t|x)dt, \tag{5.2.30}$$

so that $p(t|x)$ is the conditional density of a target t about an input vector x. Similarly

$$\langle t^2|x\rangle \approx \int t^2 p(t|x)dt. \tag{5.2.31}$$

Looking to the Eq. (5.2.29), it is possible to observe that the second term is independent of the network function $y(x)$, therefore, independent of the network weights. The optimal network function, in terms of minimizing the error, is the one which makes the first term in Eq. 5.2.29 vanish, and is given by $y(x) = \langle t|x\rangle$. Such that, the second term represents the intrinsic noise in the data sets a lower limit on the error which can be achieved (BISHOP et al., 1995).

However, i the practice, pattern classification problems have a data set finite. Be D a training set with N patterns, and each taken from the some fixed joint distributions $p(x,t)$, the optimal networking making is given by the conditional average $\langle t|x\rangle$. Therefore, to measure how close the actual mapping function $y(x)$ is to the derived one is given by the integrand of the first term in Eq. 5.2.29. This quantity depends on the particular data set D. A manner of to eliminate this dependence is considering an average over all data set, which can be rewritten as

$$\tau_D\left[\{y(x) - \langle t|x\rangle\}^2\right], \tag{5.2.32}$$

such that $\tau_D[.]$ denotes the expectation or ensemble average. An important remark is that if network function $y(x)$ is always a perfect prediction of the regression function $\langle t|x\rangle$ then this error would be zero. However, a non-zero error can arise for essentially two distinct reasons. It may be that the network function is on average different from the regression function. This called *bias*. Alternatively, it may be that the network function is very sensitive to the particular data set D, so that, at a given x, it is larger the required value for some data sets and smaller for other data sets. This is called *variance*. Finally, the error decomposition can be explicit by writing

$$\{y(x) - \langle t|x\rangle\}^2 = \{y(x) - \tau_D[y(x)] + \tau_D[y(x)] - \langle t|x\rangle\}^2, \tag{5.2.33}$$

resulting in

$$\{y(x)-\langle t|x\rangle\}^2 = \{y(x)-\tau_D[y(x)]\}^2+\{\tau_D[y(x)]-\langle t|x\rangle\}^2+2\{y(x)-\tau_D[y(x)]\}\{\tau_D[y(x)]-\langle t|x\rangle\}. \tag{5.2.34}$$

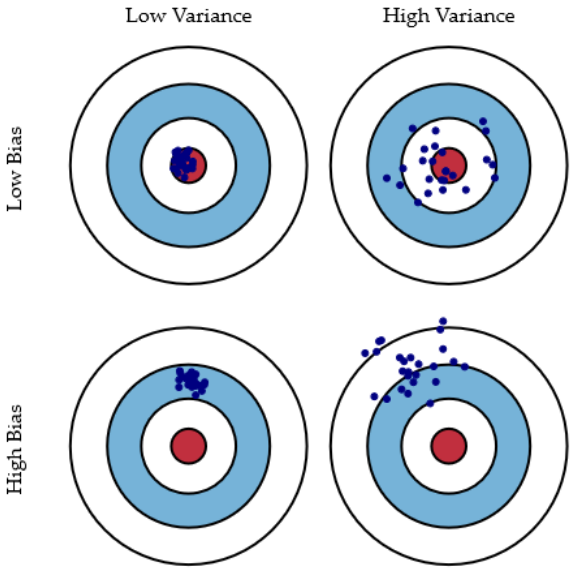
Taking the expectation in both sides over the ensemble of D, the third term on the right side vanishes, and

$$\tau_D\left[\{y(x) - \langle t|x\rangle\}^2\right] = \underbrace{\{\tau_D[y(x)] - \langle t|x\rangle\}^2}_{(bias)^2} + \underbrace{\tau_D[\{y(x) - \tau_D[y(x)]\}^2]}_{(variance)}. \tag{5.2.35}$$

The *bias* measures the extent, which average (over all data sets) of the network function differs

from the derived function $\langle t|x\rangle$. While the *variance* measures the extent to which the network is sensitive to the particular choice of data set (BISHOP et al., 1995). Figure. 5.17 shows how these affect the predictions, where the center of the target represents a perfect prediction.

Figure 5.17 – Graphical illustration of bias and variance.



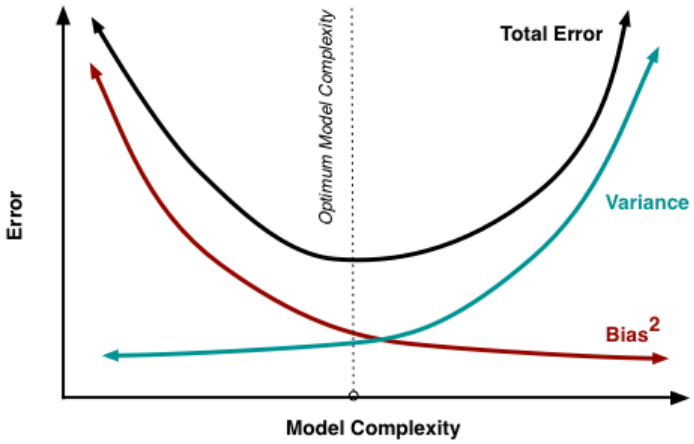
Source: (FORTMANN-ROE, 2012)

As parameters are added in the network, increasing the complexity of the system. It also increases the variance and reduces the bias, according to Fig. 5.18. Increasing the complexity of the model does not necessarily mean that the best predictions will be achieved. When the model has less complexity, $Bias^2$ will exert greater influence on the error, while for greater complexity, a $variance$ has a higher influence. Therefore, the sweet spot for any model is the level of complexity at which the increase of bias is equivalent to the variation reduction. Mathematically:

$$\frac{dBias}{dComplexity} = -\frac{dVariance}{dComplexity} \tag{5.2.36}$$

When the complexity model exceeds this sweet spot, the model is over-fitted; and under-fitting if the complexity falls short of the sweet spot. In practice, there is no analytical way to find this location and, it is recommended that many tests be made for different complexity levels, choosing which presents the best performance in the desirable work (FORTMANN-ROE, 2012).

Figure 5.18 – Bias and variance trade-off.



Source: (FORTMANN-ROE, 2012)

5.3 Confusion Matrix

A conventional manner to evaluate the pattern classifier performance is using the called *confusion matrix* (KUNCHEVA, 2004). This tool provides to verify the error distribution among different classes, evaluating the over and sub-estimation problem (HAY, 1988). Figure 5.19 shows an example of a confusion matrix. Considering a data set of healthy and damaged samples, subjected to a classifier, when a healthy sample is correctly predicted as healthy, it is a True Positive (TP) phenomenon. Otherwise, it corresponds to False Positive (FP) (ZHU et al., 2010). On the other side, for a damaged sample, True Negative (TN) is the correct classification, while False Negative (FN) is the misclassification.

Some important parameters to evaluate the classifier performance as *sensibility*, *sensitivity*, and *accuracy* can be described in terms of (TP), (TN), (FP), and (FN). Thus, sensitivity (S) is the proportion of true positives that are correctly identified by a diagnostic test

$$S = \frac{TP}{TP + FN}. \tag{5.3.37}$$

Specificity (E) is the proportion of the true negatives correctly identified by a diagnostic test

$$E = \frac{TN}{TN + FP}. \tag{5.3.38}$$

Accuracy (A) measures the degree of veracity of a diagnostic test on a condition

$$A = \frac{TP + TN}{TP + TN + FP + FN}. \tag{5.3.39}$$

Figure 5.19 – A simple confusion matrix example.

		Predicted Classes	
		P	N
Actual Classes	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

Author's production.

5.3.1 Convolution Neural Network

It is a type of Multilayer Perceptron network (MLP) where the input is not a vector as shown before, but a matrix with dimensions $(m \times n)$, such that a convolution operation is applied by *kernels* k with $p \times q$, where p and q have smaller size, resulting in features matrix $(m - p + 1) \times (n - q + 1)$. Figure. 5.20 shows an example a convolution operation by the kernel in an input matrix. Each term multiplies the elements of the matrix with specific weights, which permits that deep features are identified during the training. Thus, the output for a neuron is obtained by

$$y_i^l = \phi\left(b_j^l + \sum_{i \in M_j^l} y_i^{l-1} * k_{ij}^l\right), \tag{5.3.40}$$

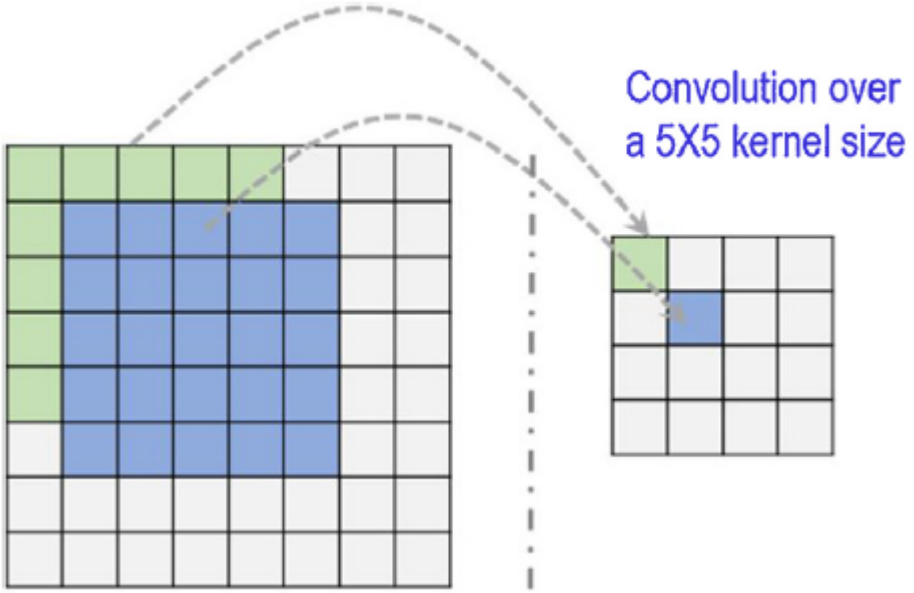
where ϕ is a non-linear activate function, ${b_j}^l$ the bias for l^{th} layer, kernel $[k]_{p \times q}$, and M_j^l the selected feature which maps i in the layer $(l-1)^{th}$ by the convolution $(*)$ applied to the input y_i^{l-1}. This step is followed by *pooling* operation, ensuring that important information is carried forward. Each feature map is subject to groupings by region, such as maximum or average non-overlapping elements. The output of this layer results in a dimensional reduction, depending on the step size choose

$$d_h^l = \Psi(a_j^l, N^l), \tag{5.3.41}$$

such that Ψ is a function to evaluate the pooling N^l and a_j^l the feature map resultant from the convolution (WANG et al., 2017).

Although the suitable results offered by this network type in pattern recognition problems, high computational efforts are demanded, becoming a limitation in their application.

Figure 5.20 – Map of a Convolution layer over a kernel size 5×5



Source: (WANG et al., 2017)

Chapter 6

Materials and Methods

In the previous chapters, a review of the essential subjects needed to develop this work was presented, fostering a theoretical background that will be used in further discussions. This chapter shows how they can be joined to provide a multidisciplinary methodology for detection and classification of damage in different structures, using Vibration-Based Model (VBM) response (time and frequency-magnitude) and Artificial Neural Networks (ANN). Initially, a data set consolidated by the researches community is used to verify the feasibility of the proposed technique. Therefore, after validated, the same strategy is used to evaluate delamination damage on composite laminae (responsible for most of the failures in structures of this type of material).

It will be seen that vibration responses have large size and feature extraction techniques are required to processing them, before feeding the neural network. It corresponds to an important step of the methodology because is not feasibly to use the total raw signal as input, considering the high computational cost demanded in relation to the number of neurons and hidden layers. However, the efficiency of damage detection/classification depends directly on the ability of such a technique to compress the largest amount of essential information for this task into a smaller feature vector. In this case, it will be used Statistical Parameters (SP) for rolling bearings time-domain data, and Principal Component Analysis (PCA) for composite beams both time-domain and frequency-domain. In addition, an alternative called Dislocated-Series (DS) is used, which does not apply statistical properties as the others, in order to evaluate how the quality of data-compression can improve the network performance.

The amount and quality of the data set are some of the most important factors to obtain good network generalization in problems of damage detection. Nevertheless, in a practice application, this methodology can be very expensive, because it is necessary to produce samples, insert damages in some of them, and extracting vibration response using data and acquisition system. In order to despite these problems, this work presents an alternative based on the finite element model able to smooth dimensional uncertainties and generate more samples.

6.1 Rolling bearings fault diagnosis

Rolling bearings data is provided by Case Western Reserve University (CWRU). According to Loparo (2013), experiments were conducted using 2 *horsepower* (hp) Reliance Electric motor, and acceleration data was measured. Motor bearings were seeded with faults from different sizes; 0.007" (0.177 8mm), 0.014" (0.3556 mm), and 0.021" (0.5334 mm), introduced separately at the inner raceway, rolling element (*i.e.*, ball). After each component was reinstalled and the vibration data was collected for different motor loads condition (motor speeds of 1797 to 1720 RPM).

6.1.1 Experimental procedures

Figure. 6.1 shows the experimental procedure to data extraction of the rolling bearings, which occurred through a system that contains a motor on the left and a dynamometer on the right a control system. The bearings support the motor shaft, and the damage was introduced to the bearing samples using electro-discharge (LOPARO, 2013).

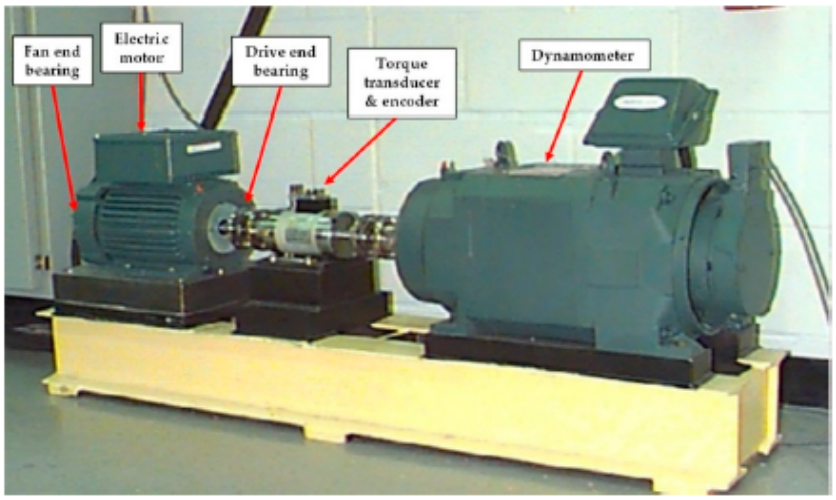
Vibration data was collected using accelerometers, attached to magnetic bases. Accelerometers were placed at the 12 o'clock position at both the Drive End (D_E) and Fan End (F_E) of the motor housing. During some experiments, an accelerometer was attached to the motor supporting base plate as well. Vibration signals were collected using a 16 channel DAT recorder and were post-processed in a Matlab environment. Digital data was collected at 12,000, and 48,000 samples per second for drive end bearing faults. Speed and horsepower data were collected using the torque transducer/encoder and were recorded by hand.

Outer raceway faults are stationary faults. Therefore, the fault location relative to the load zone of the bearing has a direct impact on the vibration response of the motor/bearing system. Experiments were carried out to quantify this effect for both fan and drive end bearings with outer raceway faults located at 3 o'clock (directly in the load zone), at 6 o'clock (orthogonal to the load zone), and 12 o'clock. Fan End (F_E) bearing: 6203-2RS JEM SKF, deep groove ball bearing. Drive End (D_E) bearing: 6205-2RS JEM SKF. Specifications of the rolling bearings used are shown in Tab. 6.1.

Table 6.1 – Bearings specifications.

Parameter	Drive End (DE)	Fan End (FE)
Inside Diameter [mm]	25.00	17.00
Outside Diameter [mm]	52.00	40.00
Thickiness [mm]	15.00	12.00
Ball Diameter [mm]	8.00	6.75
Pitch Diameter [mm]	39.00	28.50

Figure 6.1 – Test rig to measure vibration signals from damaged and undamaged bearings.



Source: (LI et al., 2019)

6.1.2 Fault Diagnosis in rolling bearings

Case studies will be developed to detect and classify damages of different sizes in the balls and inner race of the Fan End (FE) rolling bearing, using Artificial Neural Network (ANN). For the first time, the objective is only to know about the damage existence, as well as level 2 of SHM. After, the damage should be classified according to the corresponding magnitudes. The data used are the signals obtained from time-domain, with a sampling of 12.000 samples/second for both bearings choose, Fan End (F_E), and Drive End (D_E). It is worth mention that different loads were applied in the shaft by rotational velocity: Load 0 (1797 RPM), Motor Load HP 1 (1772 RPM), Motor Load HP 2 (1750 RPM), Motor Load HP 3 (1730 RPM). This simulates changes that can occur in the signal of the machines by external agent so that the classifier must be able to recognize them. Therefore, with 1920 samples for undamaged bearings, and 2880 for damaged, where each damaged state have the same number of 960 samples. All data is grouped and labeled to train, test, and validate the networks.

Before, as feature extraction whole an important step in pattern recognition problems, two techniques are applied; Statistical Parameters (SP) and Dislocated-Series (DS) for time-domain signals. In this work, the signals extracted in the time-domain are arranged in vectors, so each Statistical Parameter (SP) is evaluated, and the following matrix created, where m is the total number of signal samples and n is the number of SPs

$$\begin{bmatrix} SP_{11} & SP_{12} & \cdots & SP_{1k} \\ SP_{21} & SP_{22} & \cdots & SP_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ SP_{q1} & SP_{q2} & \cdots & SP_{qk} \end{bmatrix}_{m \times n} . \tag{6.1.1}$$

6.2 Damage detection in Glass/Epoxy beams

A study will be carried out on the use of dynamic responses (*time and frequency-domain*), aiming to detect damage in a glass-epoxy laminated beam with twelve laminae 0 degrees oriented $[0]_{12}$. The data used in this work was acquired by Völtz (2019) in the laboratory of Vibrations at Santa Catarina State University (UDESC), where the dynamic testing was performed.

6.2.1 Previous samples analysis

Different samples were manufactured for undamaged and damaged states. Figure. 6.2 presents the different delamination damage sizes for damaged beams using a *Teflon* between the middle plies, so it is classified into; D1 - 5 mm, D2 - 10 mm, and D3 - 19 mm. Thus, 73 specimens were manufactured in total, of these 25 for undamaged beams, 15-D1, 16-D2, 17-D3. Another important information is the variability presented by the samples, considering the limitations of the manufacturing process. Table 6.2, a statistical analysis of the data was performed, evaluating the mean, standard deviation, variance, maximums, and minimums for each variable. Then, normalizing the data to their respective maximum, fixing them on the same scale, with the help of a *box-plot* from Fig. B.1, it is noticed that the thickness has the highest variation between them.

Table 6.2 – Dimensional variation of samples.

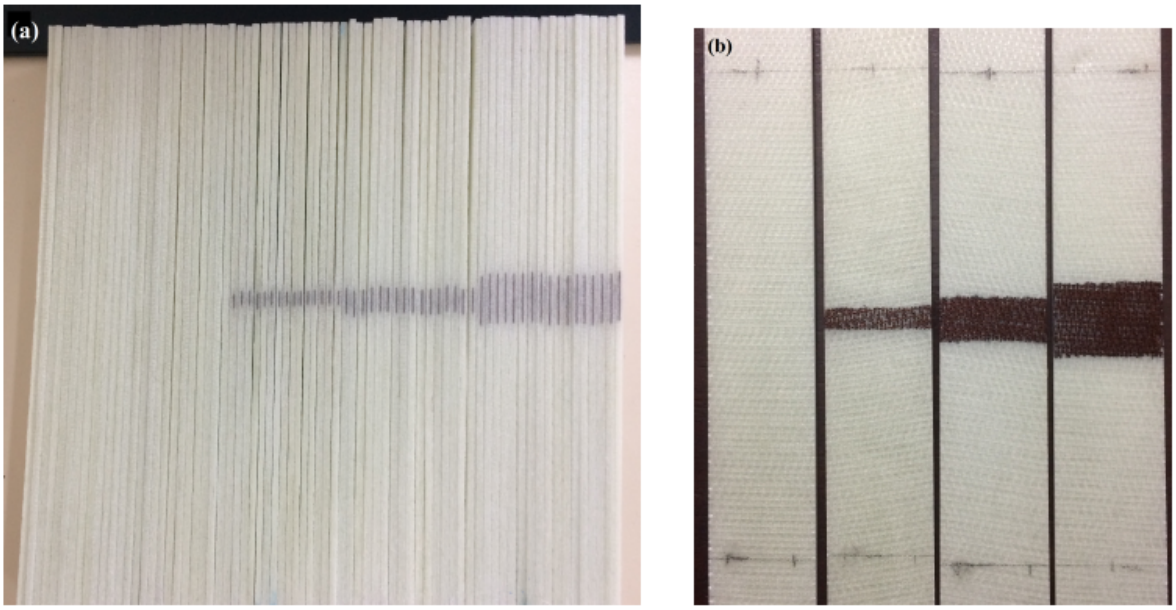
Parameters	**Width (mm)**	**Length (mm)**	**Thickness (mm)**
Mean (μ)	28.31	227.75	2.99
Maximum	29.40	230.00	3.35
Minimum	226.00	22.35	2.55
Standard deviation (σ)	1.24	0.74	0.18
Variance (σ^2)	1.54	0.56	0.03

It is necessary to consider these changes because all of these samples were subjected to modal analysis, and thus, these variations can significantly influence the collected dynamic responses.

6.2.2 Dynamic test: Experimental

The free-free boundary condition was chosen because it is the simplest form of boundary condition, and reduces effects from imperfect supports. The beams were suspended using wires to simulate a “free-free” boundary condition, as shown in Fig. 6.3. The natural frequencies and the Frequency Response Functions (FRFs) for undamaged and damaged beams were obtained using one accelerometer. The accelerometer model is 4517-C (sensitivity 0.18 pC=m2, and weight 0.6 g) lightweight structure, and was set in position 1 (*c.f.* Fig. 6.4). The accelerometer

Figure 6.2 – Healthy and Damaged composite beams.



Source: (VÖLTZ, 2019)

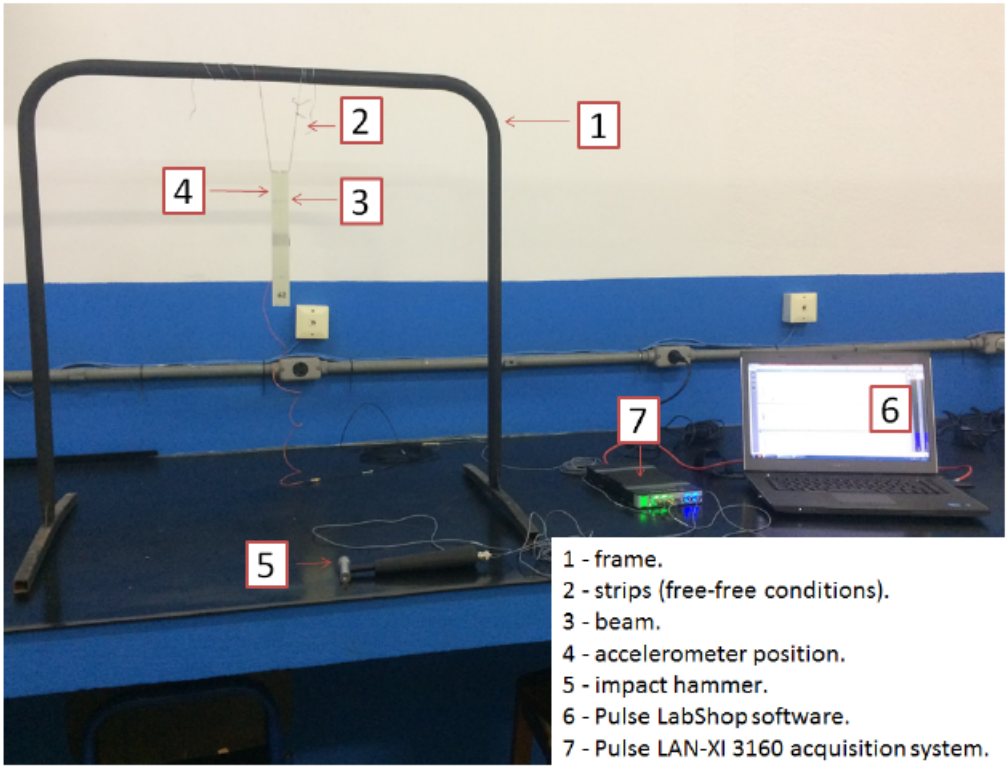
position was selected based on previous modal analyses, avoiding the presence of sensors in nodal lines. The excitation is provided by an impact hammer model 8206-003 (sensitivity 1.14 mV/N) with an aluminum tip from Brüel & Kjaer (B&K), and the position showed by the schematic representation of Fig. 6.4. Position 1 refers to H11, and position 2 refers to H21, concerning to the FRF matrix. The signals are processed by the data acquisition center via PULSE software. Note that the accelerometer position was dislocated from the centreline in the x-y plane, hence information about torsional modes could be evaluated. The analyzed frequency range was 0-3200 Hz, in a total of 6400 spectral points, as it shows in Tab. 6.3, and the values associated with the other setup parameters.

Table 6.3 – Experimental Setup.

Parameter	Value
Bandwidth	3.2 kHz
Spectral Lines	6400
Acquisition Time	2 seconds
Resolution	0.5 Hz
Averages	3
Window input	Force-Transient
Window response	Expotential
Frequency response	H_1
Test variance	2 days

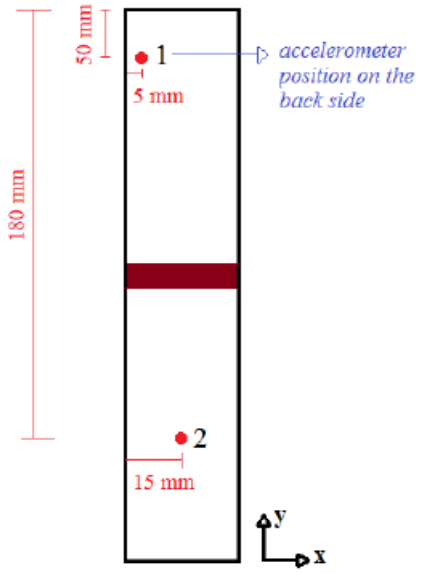
Source: (VÖLTZ, 2019)

Figure 6.3 – Beam experimental setup.



Source: (VÖLTZ, 2019)

Figure 6.4 – Schematic beam setup for test.



Source: (VÖLTZ, 2019)

6.2.3 Dynamic test: Computational

The case study addressed in this work aims to evaluate alternatives to reduce the number of samples manufactured, to generalize a neural network capable of detecting damage in structures of composite materials. As the works of Bakhary et al. (2007), Padil et al. (2017), and Padil et al. (2020) shows, the use of a consistent computational model, which can absorb the non-linearities of the properties, boundary conditions, and other uncertainties present in the phenomenon contributes significantly to the cost of damage detection methodologies. Therefore, several studies have been carried out in this direction. Thus, the approach discussed here has an overview exposed in Fig. 6.6. Initially, modal analysis is performed, based on all the responses presented by the experimental model, adjusting the properties of the laminate and the respective damping for each mode in the frequency range studied, isolating the first six rigid body modes.

The great advantage of developing this computational model is the higher number of specimens generated to train the neural network, allowing better learning for detecting damage on the structure, based on the fundamentals treated on the pattern classifiers. So, the strategy used to generate new samples from an adjusted model consists of determining a distribution of probabilities to obtain a new sample from the values presented for the dimensions: Length (L), width (w) e Thickness (t). At the first moment, the maximum and minimum values are taken into consideration, such that, it will result in $[L_{max}, L_{min}]$, $[w_{min}, w_{max}]$, $[t_{min}, t_{max}]$. Thus, fixed variation is adopted for each parameter. Then, it is possible to assess the frequency of

each value repeated, and, in this way, calculate the individual probability of a sample with that dimension. Finally, to obtain the discrete probability distribution as

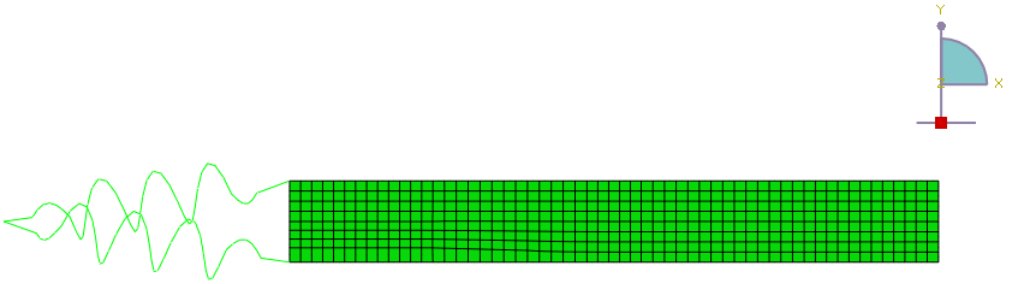
$$D_p = \frac{P_i}{N_s}, \tag{6.2.2}$$

where P_i is the individual dimension for a specific parameter, and N_s the number of samples. This allows some intrinsic uncertainties in the manufacturing process to be absolved and proceed to the computational model.

A numerical model using the Finite Element Method (FEM) of a simple composite beam in a free-free condition (using springs with 0.1 N/m to anchor the beam) was made on ABAQUS software, where the convergence was reached to 400 shell elements (S8R), with eight nodes and six-degree of freedom (3-displacement and 3-rotation) shown in Fig. 6.5. Geometry and mechanics properties used to the reference model are shown in Tab. 6.4, and they are chosen according to the previous analysis, comparing natural frequencies of the experimental model with a numerical one.

The excitation and signal collection positions are the same as the experimental process, presented in Fig. 6.4. Vibration signals are simulated with a sampling frequency of 0.5 Hz. The range studied is 1500 Hz using the direct steady-state dynamic analysis, and the state conditions are simulated as in the experimental phase. The damping factor used the pick-peak method, which evaluates the damping ratios for each peak and adjusting, in direction to approximate numerical response with experimental as much as possible.

Figure 6.5 – Computational model setup.



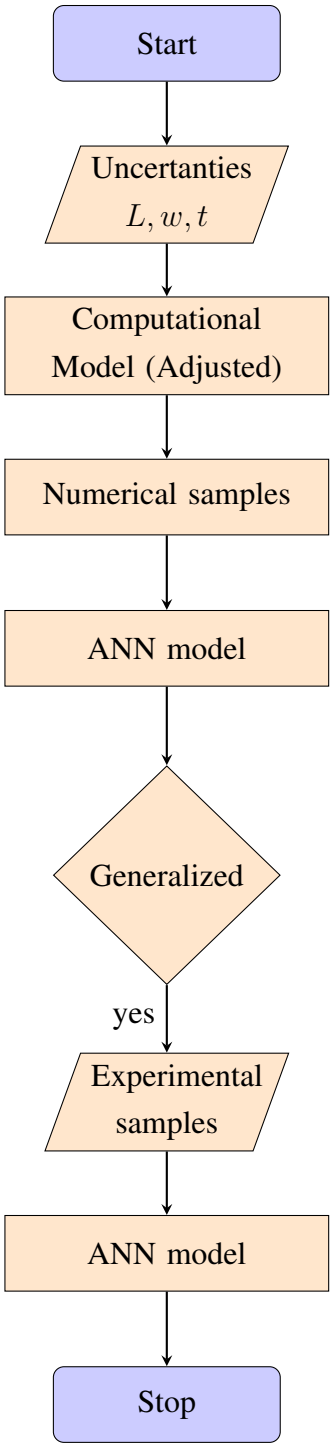
Author's production.

Table 6.4 – Composite beams $[0]_{12}$ properties adjusted in the first step of modal analysis.

Properties	Values
Length (L)	227 mm
Width (w)	28.35 mm
Thickness (t)	3.0 mm
Longitudinal Elasticity modulus (E_{11})	20.0 GPa
Transversal Elasticity modulus (E_{22})	7.3813 GPa
Poisson ratio (v_{12})	0.18
Shear modulus (G_{12})	2.1 GPa
Shear modulus (G_{13})	2.1 GPa
Shear modulus (G_{23})	4.0 GPa
Composite Density (ρ_c)	1260 kg/m^3

Figure. 6.6 summarizes in a flowchart the activities that will be made in this part of the work.

Figure 6.6 – Numerical approach schematization.



6.2.4 Damage Detection on composite beams

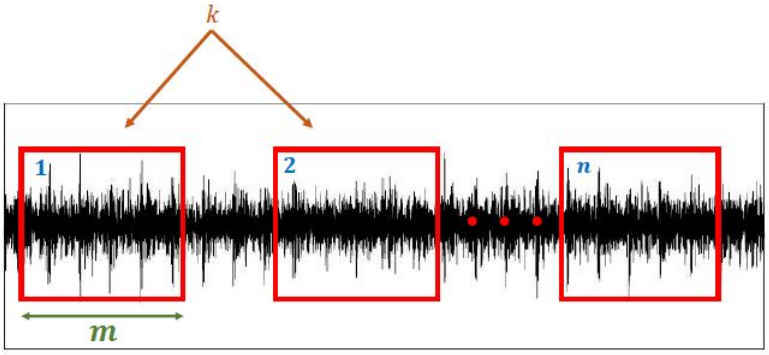
After extracting time-domain and frequency-domain responses from manufactured samples, feature extraction tools are used to reduce size-data without losing relevant information for damage sensitivity. After, it can be applied in neural networks to detect and classify healthy and damaged samples. Therefore, the objective is to use different tools for feature extractions in the pre-processing data step, like Principal Component Analysis (PCA) and Dislocated Series

(DS), evaluating which can be more efficient in SHM methodology.

6.2.5 Dislocated-Series (DS)

The main idea of this methodology is to cut out the raw signal in mini-batches with size m, and store all data in matrix the $[D]_{m\times n}$, such that n corresponds to the number of the mini-batches (Fig. 6.7 shows).

Figure 6.7 – Dislocated-Series (DS) applied to raw time-domain data.



Author's production.

a_1 and a_2 are calculated to shift out the mini-bathes, conforming the equation

$$a_1 = 0.5i\,(i+1)\,k \qquad\qquad i = 1, 2, ...n, \tag{6.2.3}$$

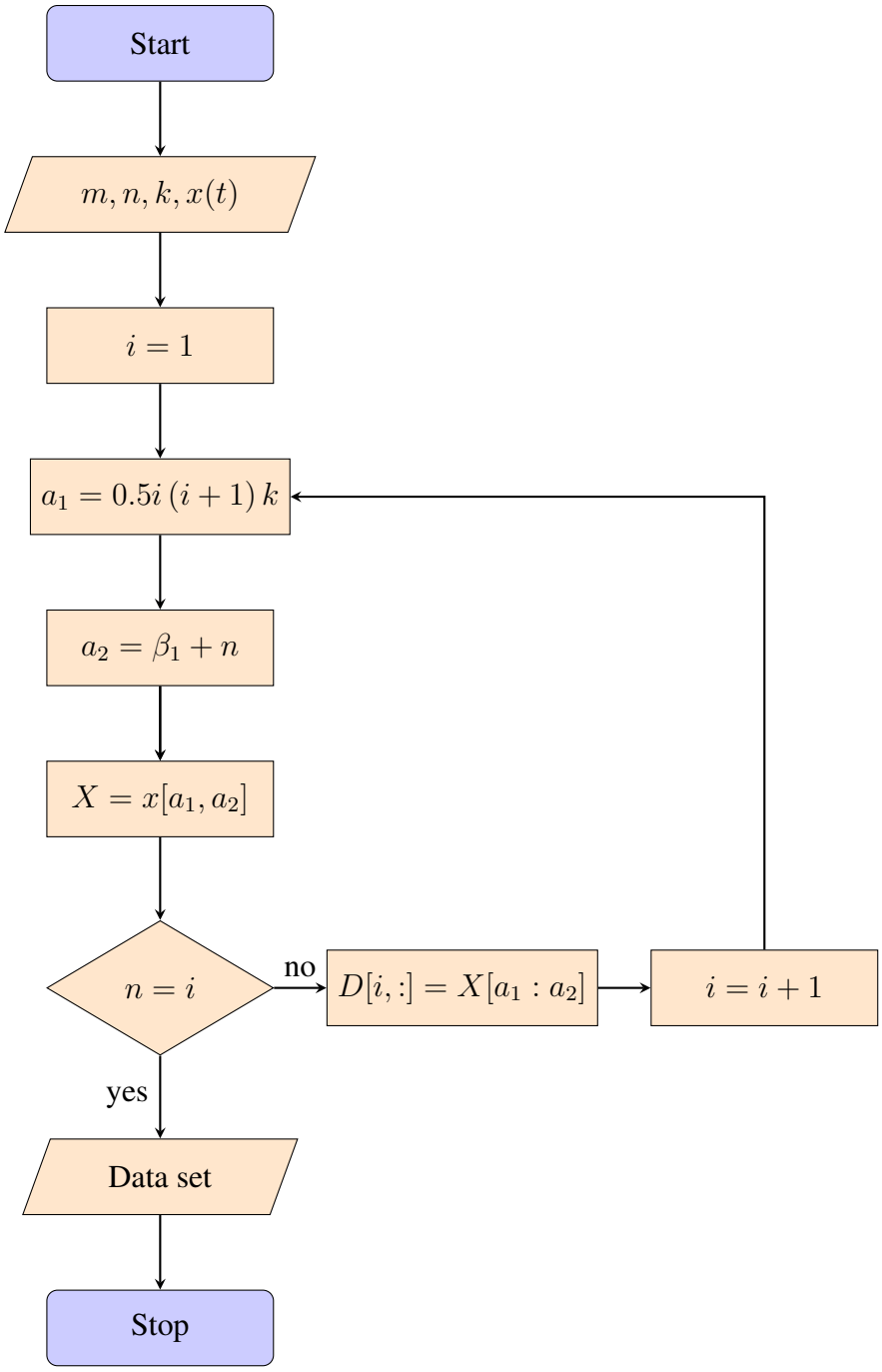
$$a_2 = a_1 + n, \tag{6.2.4}$$

where k is the dislocated step, and n the length of the intercepted signal from original. It means that the signal is cut out in n points from raw signal by m times.

$$D[i,:] = X[a_1 : a_2]. \tag{6.2.5}$$

Figure 6.8 shows a flowchart that represents the subroutine applied in the free programming language JuliaLang 1.2.0 version. In step 1, m, n, k is required to determine the matrix size resultant of the data D, and the signal x collected per sample. In Step 2, the counter starts $i = 1$, *i.e.*, serves a reference for the loop starting after evaluating the values of a_1 and a_2, extracting a piece of size m of x, allocating in a column of the matrix $[D]$. Finally, if $n = i$, the process has been completed, and the result will be $[D]$. Otherwise, the process has not yet been completed, redoing the previous steps.

Figure 6.8 – Flowchart for Dislocated-Series (DS).



Author's production.

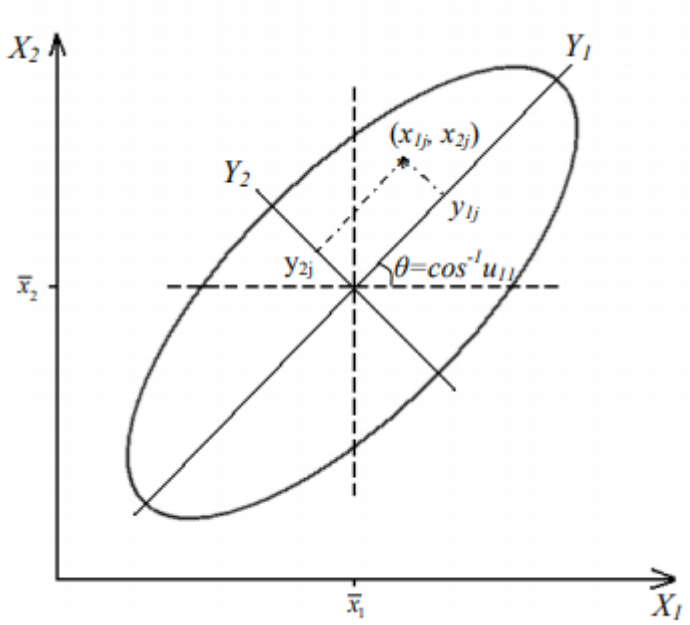
6.2.6 Principal Component Analysis (PCA)

The studies on Principal Component Analysis (PCA) started with Pearson (1901), reflecting on statistics as one of the most used tools for multivariate analysis of data in dimensional reduction. Hotelling (1933) studied the correlation between random variables. In general, the PCA is a linear transformation to transform an original set of k variables into a smaller set of n uncorrelated, such that $n < k$, so-called Principal Components (PCs) (TIPPING; BISHOP,

1999). These new variables are obtained from eigenvalue decomposition of the covariance matrix, which forms the basis for PCA. It is worth mentioning that each PC is a linear combination of original variables, and forms an orthogonal set. The full PCs set has the same size of original data, therefore, by removing those with lower power, a dimensional reduction is achieved with minimum information loss (WHITE et al., 2006).

Geometrically, PCA can be considered as a rotation applied to original data, where the new axes are known principal axes, such that the new directions represent the maximum variance of original data, as shown in Fig. 6.9 (DACKERMANN, 2009). In this example, some PCA properties could be observed. Considering an observation $x = (x_{1j}, x_{2j})$; Y_1 and Y_2 are the principal axes, where y_{1j} and y_{2j} are a projection of x onto the principal axes that gives the PCs score; the cosine of the angle θ between Y_1 and X_1 gives the first component of the eigenvector corresponding to Y_1. This axis has the property that the variance of projected points y_{1j}, with $j = 1, 2, ..., k$ is greater than the variance of the points when projected onto any other line or axis passing through $(\overline{x}_1, \overline{x}_2)$.

Figure 6.9 – Data rotation for the principle axis with max variance.



Source: (DACKERMANN, 2009)

Following the description of PCA, Jiong Tang (2006) used it as data reduction to Frequency Response Function (FRF) representation for damage detection applications via pattern recognition. Giving a matrix $X_{ij}(\omega)$ with FRFs evaluated in each k reference node, the mean of each column j and standard deviation are expressed respectively, as

$$\overline{x}_j = \frac{1}{m}\sum_{i=1}^{m} x_{ij}(\omega), \tag{6.2.6}$$

$$s_j = \sqrt{\frac{\sum_{i=1}^{m}(x_{ij} - \overline{x}_j)^2}{m}}. \tag{6.2.7}$$

Then, the data set is transformed into a standard normal space, yielding a new FRF matrix, expressed by $\tilde{x}_{ij}(\omega)$

$$\tilde{x}_{ij} = \frac{x_{ij} - \overline{x}_j}{s_j}. \tag{6.2.8}$$

Once that $\tilde{x}_{ij}$ is the variance matrix $\tilde{X}$, the covariance matrix

$$cov([\tilde{X}]) = \frac{[\tilde{X}]^T[\tilde{X}]}{m-1}. \tag{6.2.9}$$

By definition, solving the eigenproblem, (P_i) are the principal directions, and λ_i the correspondents variance

$$cov(\tilde{X})\{P_i\} = \lambda_i\{P_i\}. \tag{6.2.10}$$

The PCs determine the amount and direction of higher variability of the data, and, in most cases, the first ones can represent the properties of the data in the reduced space n-dimensional. Therefore, those of lesser value can be considered as noise contribution and can be discarded (HOYER; HYVÄRINEN, 2000).

With the use of PCs in the reduced space, FRFs can be reconstructed and compared with the original, according to a detailed procedure by Jiong Tang (2006), serving to evaluate the efficiency of the technique. Considering that an observation was made and $(x_{new})_{1\times k}$ be admitted. Using $\overline{x}_j$ e s_j

$$(\tilde{x}_{new})_{1j} = \frac{x_{new1j} - \overline{x}_j}{s_j}. \tag{6.2.11}$$

The projection of variance matrix

$$[A]_{1\times k} = [\tilde{x}_{new}]_{1\times k}[P]_{k\times k}, \tag{6.2.12}$$

$$[\tilde{x_R}]_{1\times k} = [A][P]^T. \tag{6.2.13}$$

Thus, the FRF for a new observation will be

$$(x_R)_{1j} = s_j(\tilde{x_R})_{1j} + \overline{x}_j. \tag{6.2.14}$$

6.2.7 Artificial Neural Network (ANN) to fault diagnosis

The neural network model chosen was Multilayer Perceptron (MLP), using the Flux toolbox of JuliaLang 1.2.0 version that receives pre-processed data by feature extraction techniques. This data is partitioned for the training, testing, and validation steps, such that Silva et al. (2010) proposed a percentage above 60% for training. After partitioned, the data for each corresponding step is shuffled, making it difficult for the network to memorize data. This model is based on supervised learning (with the help of a teacher), where the entire data set must have labels associated with the class belonging to the samples, defining the number of neurons that will be used in the output layer. Firstly, the objective is to classify the samples between healthy and damaged, establishing two classes, two neurons are used, following the method called neurons/classes, recommended facilitating classification. Thus, $[0.9999; 0.0001]$ (Healthy) and $[0.0001; 0.9999]$ (Damaged). The close values instead of 0 and 1 are used, because they represent the limits of the sigmoid activation function and can exert an influence on the training, causing the weights to increase a lot to reach that, according to Haykin (2007). The same methodology is adopted when more classes are required.

Determining hyper-parameters such as the number of hidden neurons and layers is a recurrent problem in the MLP networks. When too many of these are chosen, over-fitting can occur. While under-fitting when it is insufficient. Therefore, in this work, the idea is that some tests are carried out with different types of topologies, aiming to obtain the one that will present the best performance. Different types of the activation functions are used (tanh, sigmoid, relu) in the hidden neurons, and softmax in the output layer.

Chapter 7

Results and Discussion

This chapter is divided into case studies carried out for both damage detection in composite material beams and bearings using Artificial Neural Networks. However, before feeding the pattern classification tool, a feature extraction technique is necessary. Then, Statistical Parameters (SP), Dislocated-Series (DS), and Principal Component Analysis (PCA) are used, where their particularities are studied as:

- Number of principal components (PCs) for the PCA.
- Statistical parameters (SP) with higher sensitivity to damage.
- Size of the Dislocated Series (DS) and their influence on classifier performance.

Important parameters such as the number of neurons, layers, activate function, and learning algorithms will be studied for each case.

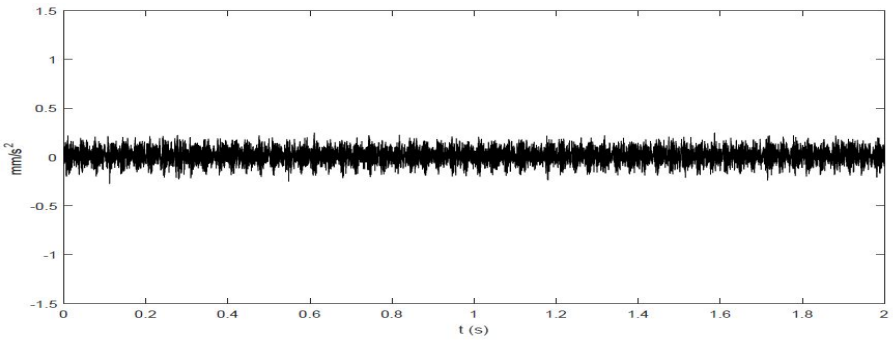
7.1 Damage detection on rolling bearings

The data used to detect damage was available by Loparo (2013), in the time-domain, for ball faults and inner race faults. Therefore, based on literature and preliminary studies, the Statistical Parameters (SP) as feature extraction were studied to this data set, and comparing with Dislocated Series (DS). After, the studies are expanded in terms of classification of different damage sizes, evaluating positive and negatives points for each one.

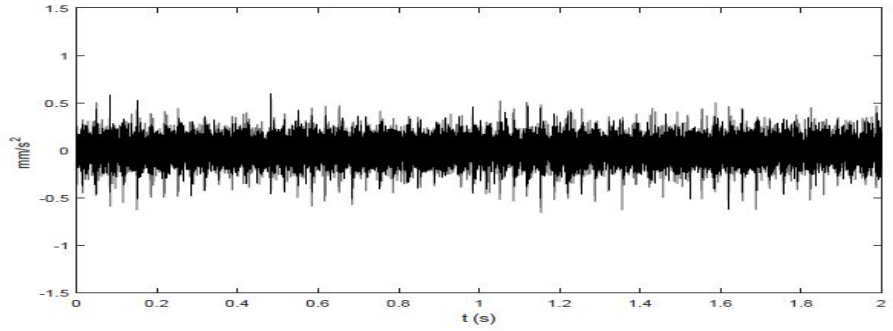
7.1.1 Inner race fault detection

Figure. 7.1 shows healthy and damaged signals for different damage sizes on the inner race. Looking at these vibration responses, first, it is possible to observe a notable difference between healthy and other signals. The healthy signal has a lower magnitude than the damaged one. However, when the size of damage improves, it is not easy to identify quickly features that differentiate them.

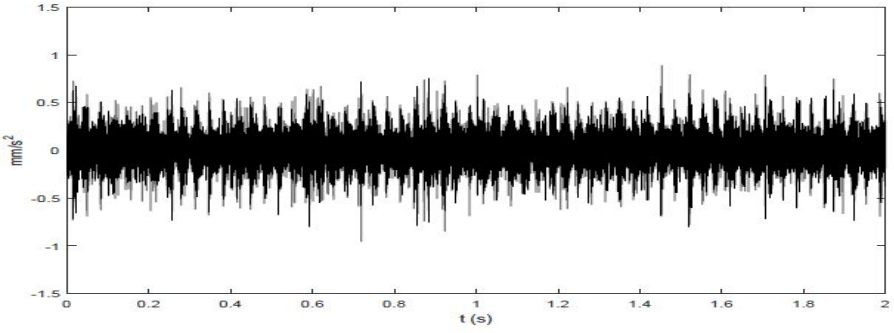
Figure 7.1 – Vibration signals data for both health and damaged bearings at 1797 RPM - Load 0.



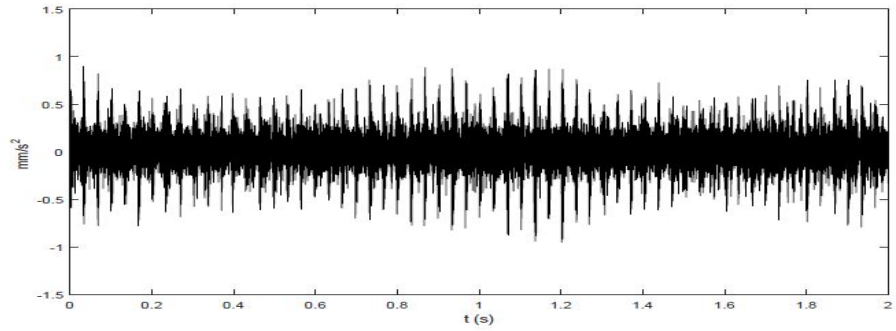
(a) Baseline raw signal.



(b) 0.007” Inner race fault raw signal.



(c) 0.014” Inner race fault raw signal.



(d) 0.021” Inner race fault raw signal.

Author’s production. Source: (LOPARO, 2013).

7.1.1.1 Statistical Parameters (SP)

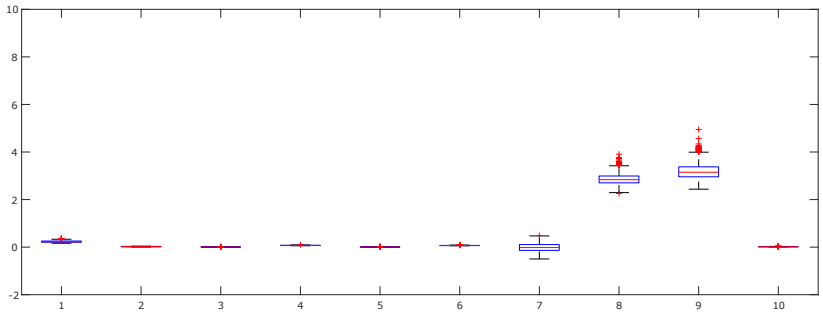
The data available for rolling bearings Drive End (DE) and Fan End (FE) has 240000 points for healthy samples, and 120000 fo each damage type. Therefore, as discussed, it is necessary to reduce the signal into smaller pieces to serve as input for the neural network. Thus, smaller vectors with a size of 1000 were employed, generating several Statistical Parameters (SP) for each sample, stored in a matrix. So, the next step consists to check the damage sensitivity for the parameters studied, which can be seen in Fig. 7.2, expressed by:

1. Peak amplitude (y_{peak}).
2. Mean ($\overline{y}$).
3. Mean square ($\overline{y}_{sq}$).
4. Root-mean-square (rms)
5. Variance (σ^2).
6. Standard Deviation (σ).
7. Skewness (γ).
8. Kurtosis (κ).
9. Crest Factor (X_{Cf}).
10. K-factor (X_k).

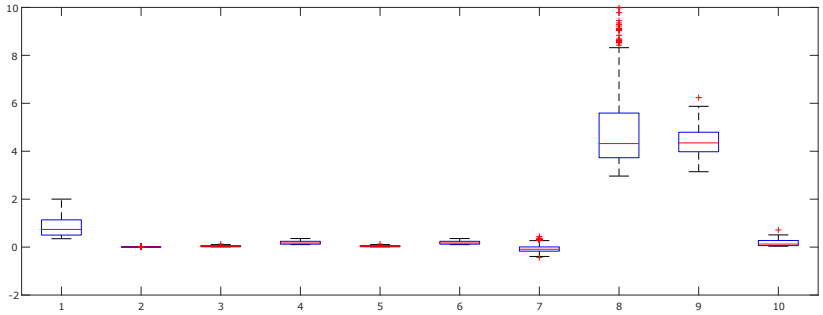
As seen in Fig. 7.2, the amplitude has increased the presence of faults in the inner race, explained by periodic localized impacts, inducing a vibratory excitation in the system. As *box-plots* from Fig. 7.2, it is possible to see how the damage influences each parameter specifically. Peak amplitude, Kurtosis, Crest Factor, and K-factor show high sensitivity to damage, but not to the size. Followed by Mean, Mean square, Root-mean-square, Variance, Standard Deviation, and Skewness. Taking attention to the significant presence of outliers in the D1 and D2 damages, notable to the parameters κ, and X_{Cf}. The red line in each box indicates the median evaluated for the samples of a given parameter. Thus, healthy samples have a tendency towards values close to 3, which is expected by theory, indicating that the signal reaches a Gaussian distribution. The small variation of the box for the Crest Factor and K-factor induce that the signal was collected close to the damage, which is confirmed because the signal used comes from the Fan End (FE).

Analyzing the damage sensibility, a Multilayer Perceptron (MLP) model with a back-propagation learning algorithm, using $\eta = 0.001$, and moment term $\alpha = 0.99$, aiming to answer a binary question regarding the existence only, as SHM level 2 required. The data were divided

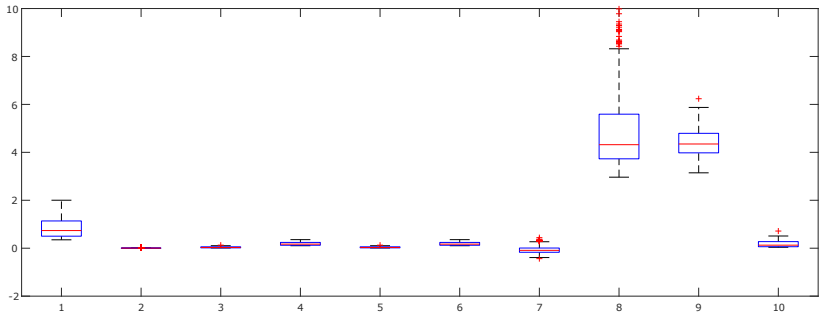
Figure 7.2 – Analysis of statistical parameters for damaged and intact rolling bearings.



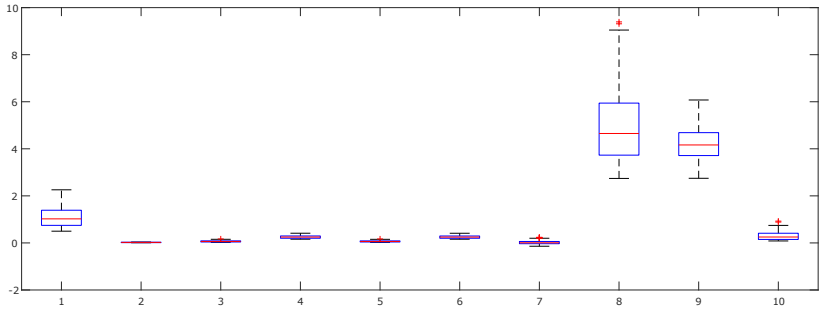
(a) Undamaged



(b) Damage - D1



(c) Damage - D2



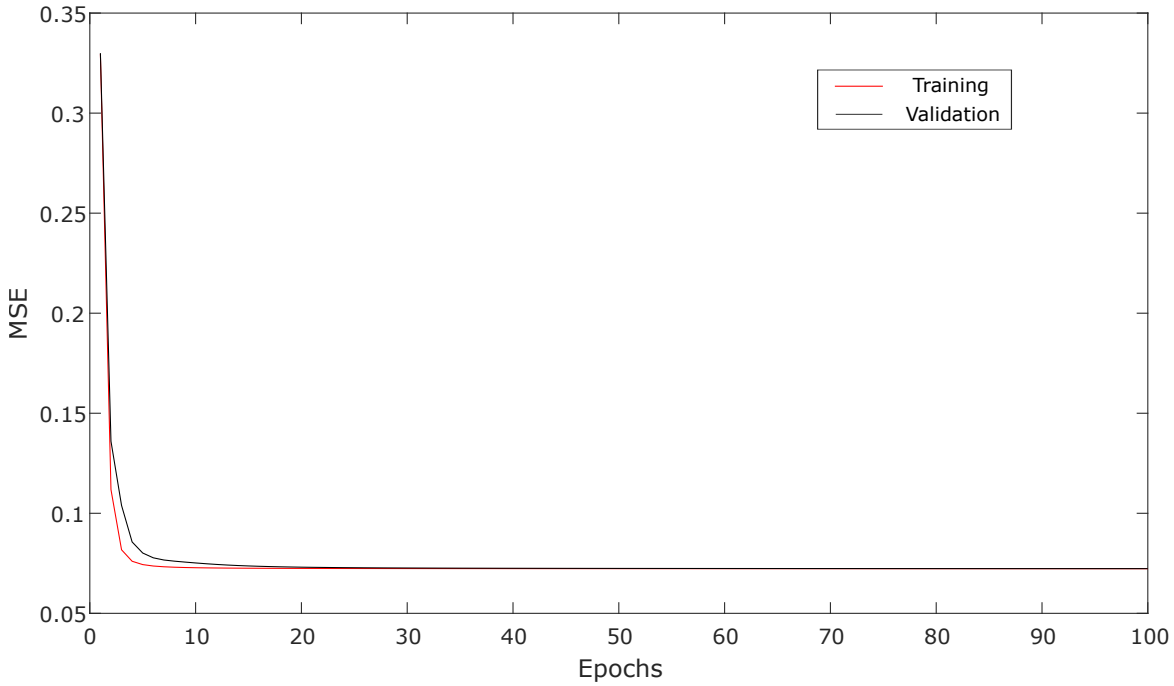
(d) Damage - D3.

Author's production.

into 80 % for training and 10 % for testing and validation. Network topology has two neurons in the hidden and output layer by the strategy of an output neuron by class, which facilitates the determination of a decision boundary between two sets of data. For training, the mean-square error function was used with the softmax in the last layer, allowing the maximum probabilities approximates for $[0.9999; 0.00001]$ (intact) and $[0.00001; 0.9999]$ (damaged), through 100 epochs to adjust the weights.

The results obtained were of 100.00 % for correct predictions in the training phase, and 100.0 % for the testing. Figure. 7.3 shows the convergence curves. Considering the data that were not previously used in the training or the testing phase, the validation phase presented 100 % accuracy. It shows that the detection problem has an excellent result.

Figure 7.3 – Convergence for training and validation phases of the network.



After successfully detecting the damage, the next task is to classify them according to the size of the damage present. Thus, instead of two neurons in the last layer, it is necessary to use four to follow the strategy of one neuron per class, while the input layer remains with ten neurons. In the hidden layer, eight neurons are used, followed by sigmoid activation functions. Figure. 7.4 shows the percentage of correct answers in the learning phase, using different algorithms. ADAM and RMSprop presented better results than Momentum and AdaMax, associated with the presence of local minimums and how each one updates the point in the weight field.

Table. 7.1 shows tests realized, modifying the number of hidden neurons, and activate functions. Using ADAM and RMSprop algorithms, it is presented the results for training, testing, and validation. Distribution error of the confusion matrix shown in Fig. 7.5 permits to

analysis that a healthy sample is correctly classified with 100% accuracy. However, it is relevant to note that, misclassification occurs for different damage sizes.

Figure 7.4 – Accuracy curve for different learning algorithms in training step.

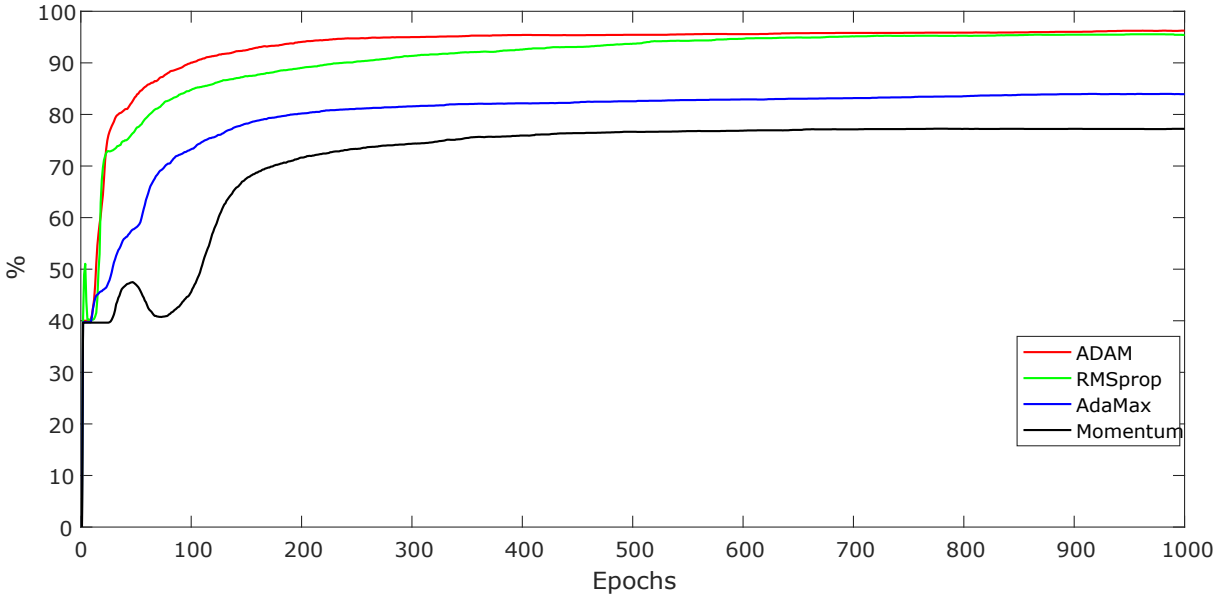
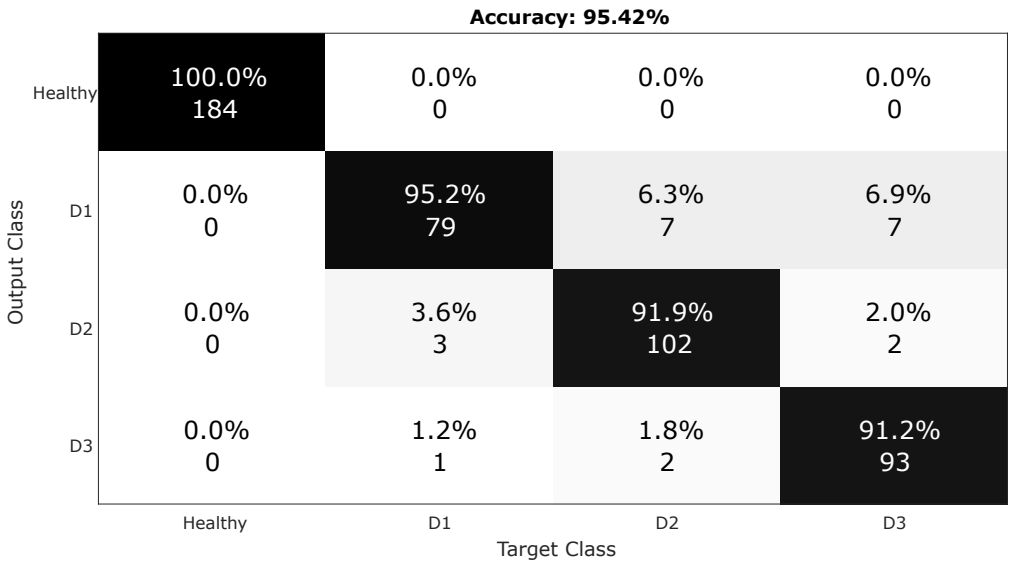


Table 7.1 – Results for different ANN topologies to detect damage on the inner race.

Topology	Activate Function	Algorithm	Training	Validation	Testing
10 - [8]- 4	[σ-σ]	ADAM	95.91%	95.20%	96.46%
		RMSprop	85.78 %	86.25 %	88.96 %
10 - [8]- 4	[$tanh$-σ]	ADAM	95.46%	94.16%	96.67%
		RMSprop	96.04%	96.67%	95.42 %
10 - [12]- 4	[σ-σ]	ADAM	96.09%	93.95%	96.88%
		RMSprop	96.04%	95.41%	96.46 %

Figure 7.5 – Confusion matrix for damage classification using RMSprop and 10-[8]-4 topology.



7.1.1.2 Dislocated-Time Series (DS)

This strategy is based on breaking the signal into n pieces with m sizes, aiming to increase the number of samples for the network so that it can detect the characteristics necessary to identify the damage. The idea is to evaluate an alternative to the statistical parameters, assuming that they can remove important information for the process.

However, taking into account that the data set to train the model must increase considerably about the statistical parameters, the complexity required for the model will also be higher. Thus, Tab. 7.2 shows different types of models used and respective performances, using softmax, cross-entropy, and $[tanh - \sigma]$ as activate functions.

Table 7.2 – Different models and their results for bearing damage detection.

Hidden Neurons	n	m	Algorithm	Training	Validation	Testing
35	100	10	Momentum	100 %	96.25 %	96.25%
	80	40	RMSprop	99.99%	98.81%	99.06%
50	100	10	ADAM	100 %	99.75 %	99.25 %
	80	40	RMSprop	100.0 %	99.31 %	99.06%
30	100	10	ADAM	100.0 %	99.45 %	99.71 %
	100	60	AdaMax	99.99 %	99.58 %	99.63%

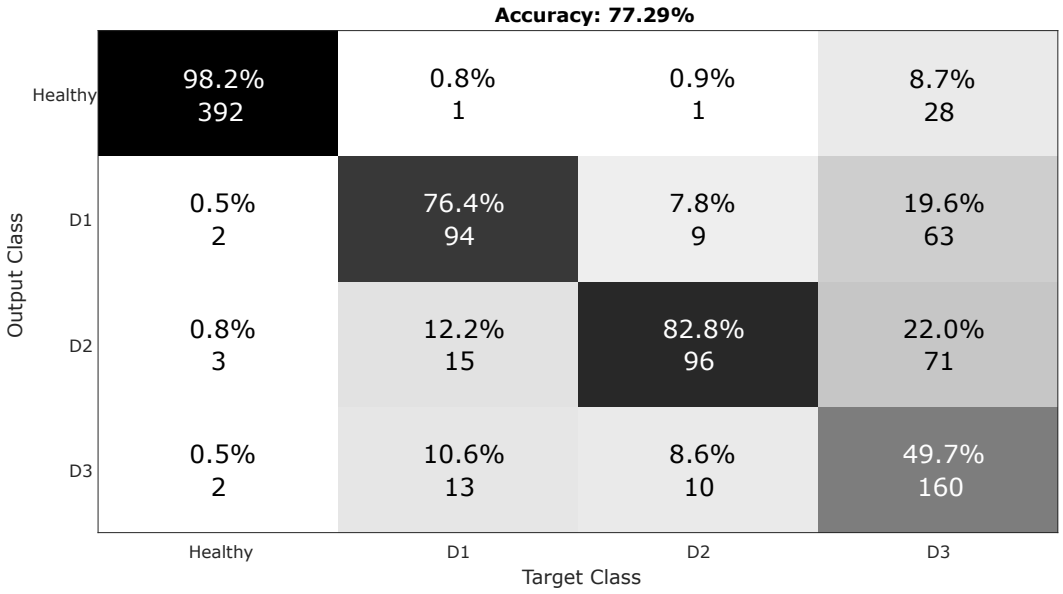
Besides, the training becomes slower because more data used, and the results obtained are satisfactory, with more than 99% for learning. It permits the application of the tool to classify damages in the next step. Thus, the network must have four neurons in the output layer to classify each state considered; Healthy, D_1, D_2, D_3. Table 7.3 shows the models used their respective results. However, in this approach, the time-dislocated series methodology does not show good results. The confusion matrix in Fig. 7.6 shows that the methodology keeps useful

to detect damage, classifying correctly healthy samples. However, different damage becomes hard work, needing a raw signal with large size to evaluate features and improves the results.

Table 7.3 – Inner race faults classification using time series dislocated.

Topology	(n,m)	Activate Functions	Algorithm	Training	Validation	Testing
100 - [10-5]- 4	(10,100)	[tanh-tanh-σ]	ADAM	85.85 %	46.25 %	50.31 %
100 -[20-15]-4	(10,100)	[tanh-tanh-σ]	RMSprop	84.68 %	50.62%	44.38 %
120-[30-15]-4	(30,120)	[tanh-tanh-σ]	RMSprop	96.94 %	65.72%	69.27 %
120-[25-10]-4	(60,120)	[relu-relu-σ]	AdaMax	98.94 %	76.04%	80.31%
120-[30]-4	(150,60)	[σ-σ]	AdaMax	98.25 %	72.5 %	77.29 %
150-[30]-4	(150,60)	[σ-σ]	AdaMax	98.24 %	73.33	77.29 %

Figure 7.6 – Decision Matrix for faults classification on rolling bearings with damage in inner race.



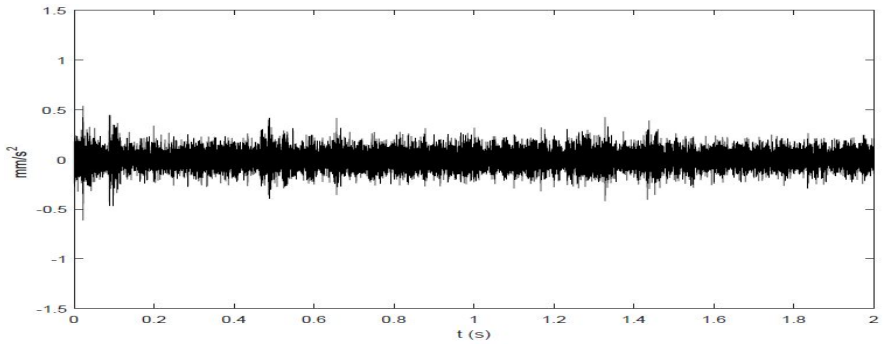
Both Dislocated-Series (DS) and Statistical Parameters (SP) show good generalization in damage detection problem. However, while DS demands more input neurons and, consequently, computational efforts, only ten parameters are needed for Statistical Parameters (SP). Thus, SPs become a better alternative in this case. Besides good results obtained for damage detection, it was not followed in the classification step. According to the confusion matrix, the damage size is not large sufficient to change the time-domain response and be detected by the neural network after feature extraction techniques.

7.1.2 Ball fault detection

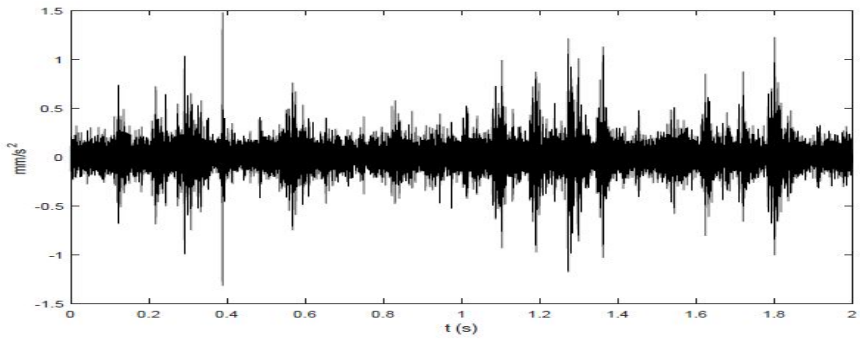
The next case study deals with the identification of damage, and classification of the different sizes of damage to the balls. According to the methodology, different signals are

exposed for the prior analysis, where can be noticed differences among in Fig. 7.7, for the data collected from the damaged bearings. Figure. 7.1(a) presents the intact signal samples which permit verify a larger amplitude increase due to impacts caused by damage, affecting the vibration excitation, and modifying the structural dynamic response. However, this data are treated both employing Statistical Parameters (SP) and by the Time Series Dislocated, analyzing benefits and negative points for each one.

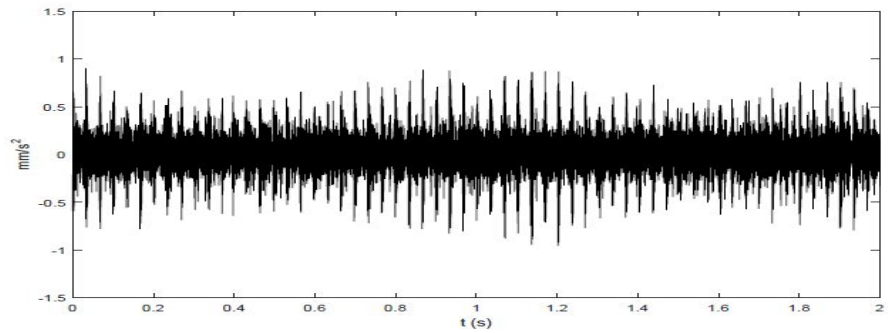
Figure 7.7 – Vibration signals data to damaged bearings for different sizes, collected at 1797 rpm.



(a) 0.007” ball fault raw signal.



(b) 0.014” ball fault raw signal.



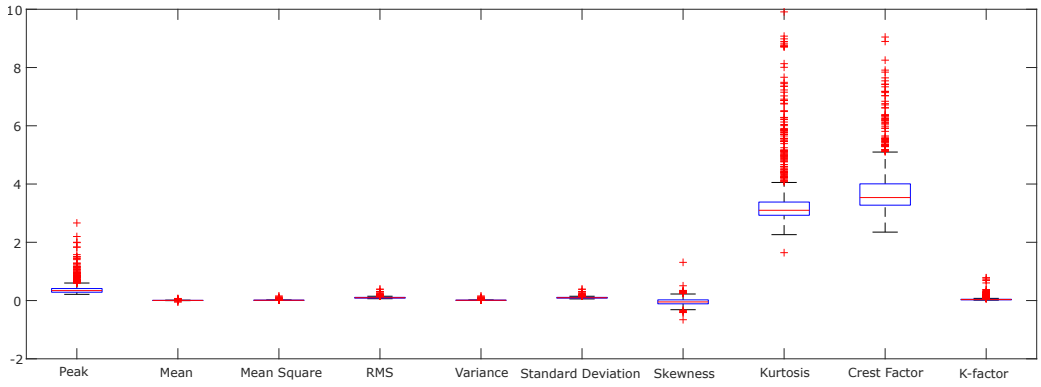
(c) 0.021” ball fault raw signal.

Author’s Production. Source: (LOPARO, 2013).

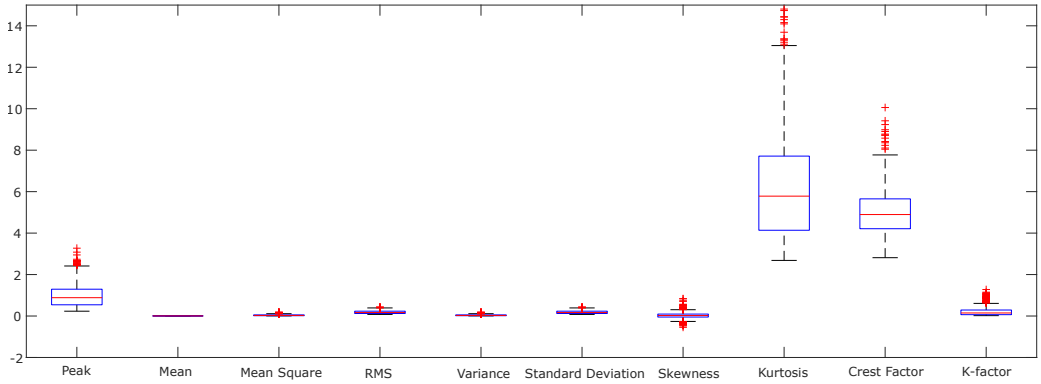
7.1.2.1 Statistical Parameters

As in the previous case, box-plots are used to visualize how damages on balls influence in the statistical parameters. It can be seen in Fig. 7.8 the difference between ball fault and inner race statistical parameters. In this case, maximum peak, Kurtosis, and Crest Factor have a higher sensitivity for D2, where the bounds of the box show higher amplitude. It is necessary to take into account the presence of outliers, represented by red markers.

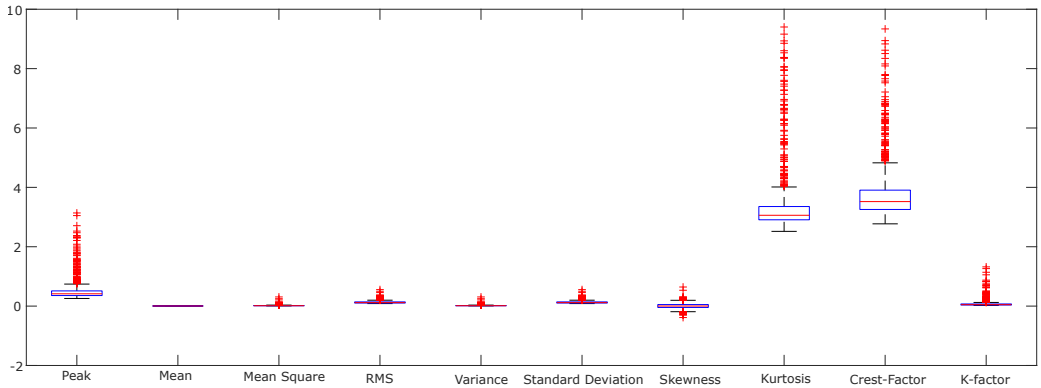
Figure 7.8 – Analysis of statistical parameters for rolling bearings with ball fault.



(a) D1



(b) D2



(c) D3.

Author's Production.

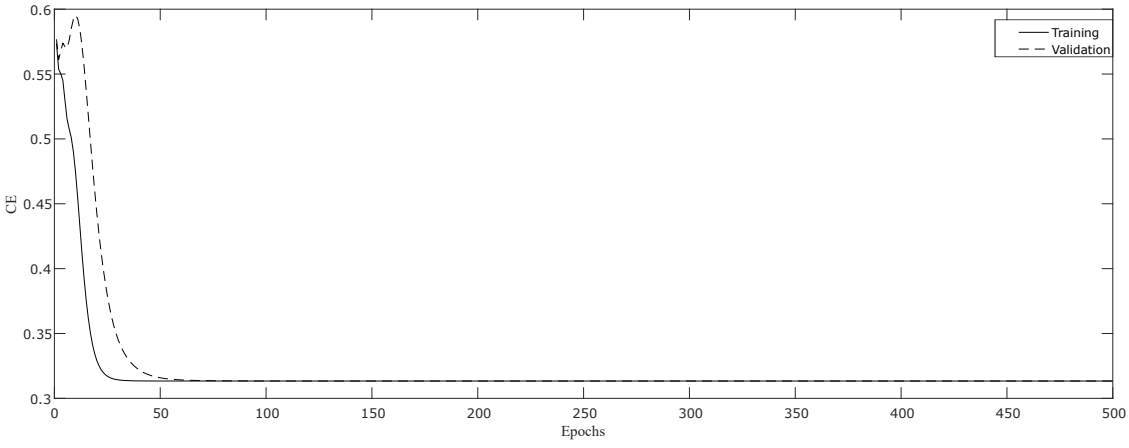
Following the methodology, at first, the proposed neural network should classify the samples only as damaged and undamaged, without having to worry about specifying the type of damage in question. Then, the Tab. 7.4 determines the different results obtained for the use of different training algorithms for a network with 10-[3]-2 topology, [$tanh$- σ] activation functions, in addition to softmax in the output layer with the error, evaluated via cross-entropy.

Table 7.4 – Models used to detect ball faults in the rolling bearings.

Algorithm	(η)	(α)	Training	Testing	Accuracy
Steepest Descent	0.00050	0.001	95.13%	96.25%	96.25%
ADAM	0.0001	-	99.92%	99.58%	99.79%

Using ADAM (without any modification) as a training algorithm, the network had a high generalization capacity with 99.79 % of the accuracy. Figure. 7.9 shows the training and test convergence curves.

Figure 7.9 – Curves for training and testing in damage detection model.



Although the damage detection was successfully carried out, as shown by the previous results, the same success was not achieved for the classification step, mainly about the damage.

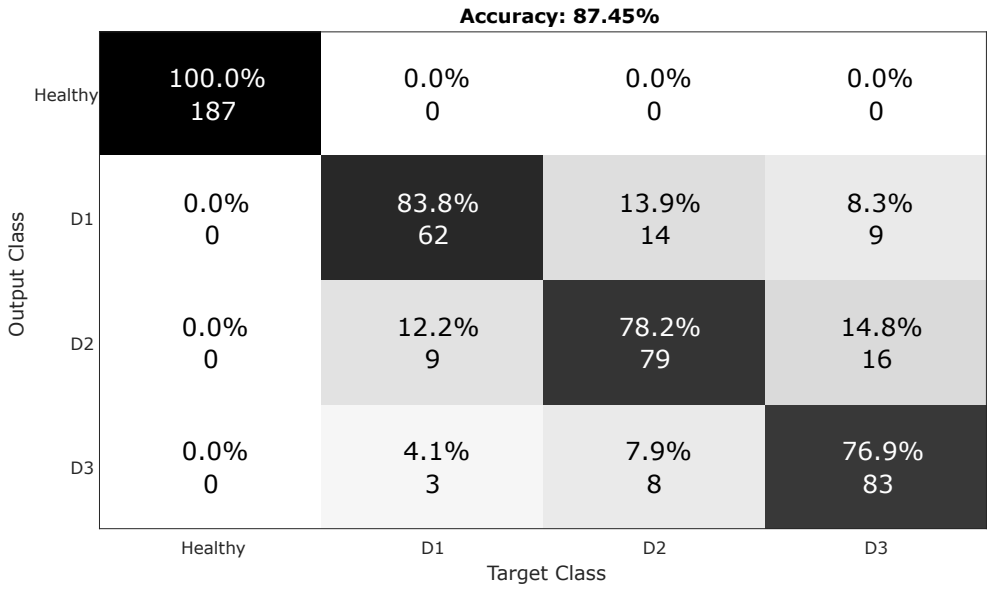
As there are four classes, four output neurons are used. After some tests were carried out, it was established the topology 10-[5-5]-4 for the network, considering different types of learning algorithms, as shown in Tab. 7.5, which ADAM once again showed superior generalization capacity for 1000 times.

Table 7.5 – Algorithms used to train the network and respective results.

Algorithm	(η)	(α)	Training	Testing	Accuracy
Steepest Descent	0.0015	0.0001	72.31%	73.33%	78.13%
Adam	0.00015	-	82.96%	84.16%	82.92%
RMSProp	0.002	0.95	84.50%	84.58%	83.75%
AMSGrad	0.001	0.995	85.83%	81.67%	86.25%
AdaMax	0.001	0.995	86.22%	87.5%	87.71%

As Fig. 7.10 shows, the use of statistical parameters in this problem shows good results when the interest is only to detect the damage. It is possible to observe that the first row and column of the decision matrix presented 100% of the intact samples correctly classified. However, different damages types result in misclassification, limiting the methodology, associated with the presence of outliers previously evaluated in the parameters, which make it difficult for the network to establish a boundary decision.

Figure 7.10 – Confusion matrix resulted in AdaMax algorithm testing.



7.1.2.2 Dislocated-Series (DS) Time

Regarding the change in the dynamic response for rolling bearing with ball faults, it is valid to use this strategy, aiming to observe their applicability. Thus, the signal is allocated as a package containing pieces of information about the whole. So, varying m and n that define the size of theses packages and input layer for neural network, as well as the total packages for a signal sampled. Therefore, Tab.7.6 shows some models used and their respective results,

which permit evaluate the technique to detect ball faults in rolling bearings using a time-domain response. As in the previous cases, the error function used was cross-entropy for 200 epochs.

Table 7.6 – Models of Neural Networks used to detect ball faults on rolling bearings and the respective performance.

Topology	(n,m)	Activate Functions	Algorithm	Training	Validating	Testing
60-[10]-2	(60,60)	$[\sigma - \sigma]$	ADAM	96.54 %	91.25 %	93.17%
30-[20]-2	(80,30)	$[relu - \sigma]$	ADAM	97.73%	94.56%	95.09 %
70-[25]-2	(80,70)	$[\sigma - relu]$	RMSprop	99.99 %	99.50%	99.56%
80-[35]-2	(80,80)	$[tanh - \sigma]$	Momentum	99.99%	99.34%	99.38%
100-[40]-2	(80,100)	$[tanh - \sigma]$	AdaMax	99.98%	99.65%	99.69%
100-[30]-2	(60,100)	$[tanh - \sigma]$	ADAM	100.0 %	99.75%	99.63%

With this, it is possible to infer the viability to use the technique for identifying Ball faults bases on the dynamic response, based on the tests that were performed varying the amount of data, topology, and learning algorithms. The better result was found for the ADAM algorithm with 30 neurons in the hidden layer, and a combination of the $tanh$ and σ as activating functions, but other models have presented good results too as Adamax, Momentum, and RMSprop.

To classify the damages in rolling bearings is not easy as seen in before cases because the signals are very polluted, and it becomes more difficult to extract important features, hindering effective learning. Testing several models and, therefore, as can be seen in the Tab. 7.7. All data uses cross-entropy error function and softmax in the output layer, where the data was partitioned in 80 % for training and 10 % for testing and validation.

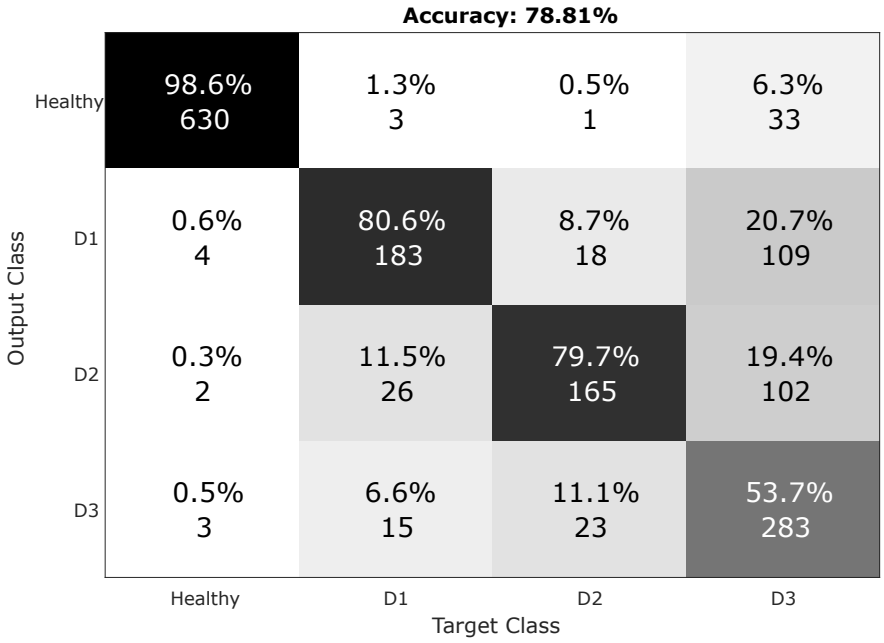
Table 7.7 – Models of Neural Networks used to classify ball faults on rolling bearings and their respective performance.

Topology	(n,m)	Activate Functions	Algorithm	Training	Validation	Testing
150-[80]-4	(60,150)	$[\sigma - \sigma]$	AdaMax	99.96 %	76.35 %	80.00%
120-[50]-4	(40,120)	$[tanh - \sigma]$	Momentum	98.24%	71.09%	72.19%
100-[35]-4	(100,80)	$[tanh - \sigma]$	ADAM	90.19%	58.90%	64.45%
80-[60]-4	(80,100)	$[relu - \sigma]$	RMSprop	91.78%	70.25%	76.19%
150-[60]-4	(100,150)	$[tanh - \sigma]$	AdaMax	99.59%	73.93%	78.81%

It is noticed that this technique is not suitable to classify different damage sizes due to the intrinsic noisy of the system. Figure 7.11 deals with the decision matrix obtained by the data from samples that were not used during the training and testing phase. Thus, proving that healthy samples are well classified, but the efficiency drops appreciably for the classification of damages, with worse results for Damage 3.

Statistical Parameters (SP) obtained from time-domain data have large number of outliers associated with noise and it makes difficult the damage classification. From two cases studied, it can be observed that DS has a good capacity to detect local features in the signal.

Figure 7.11 – Confusion matrix resulted for AdaMax algorithm training in a rolling bearing damage classification problem, using dislocated series.



7.2 Damage detection in a Glass/Epoxy Beam

As mentioned in the previous chapter, a data set was generated containing the dynamic responses extracted from the testing, i.e, data in the *time-domain* and *frequency-domain*. Thus, this section presents results for different techniques used in terms of extraction of properties and data compression, to establish a reliable technique that makes accurate diagnoses about the occurrence of damage in beams of composite material.

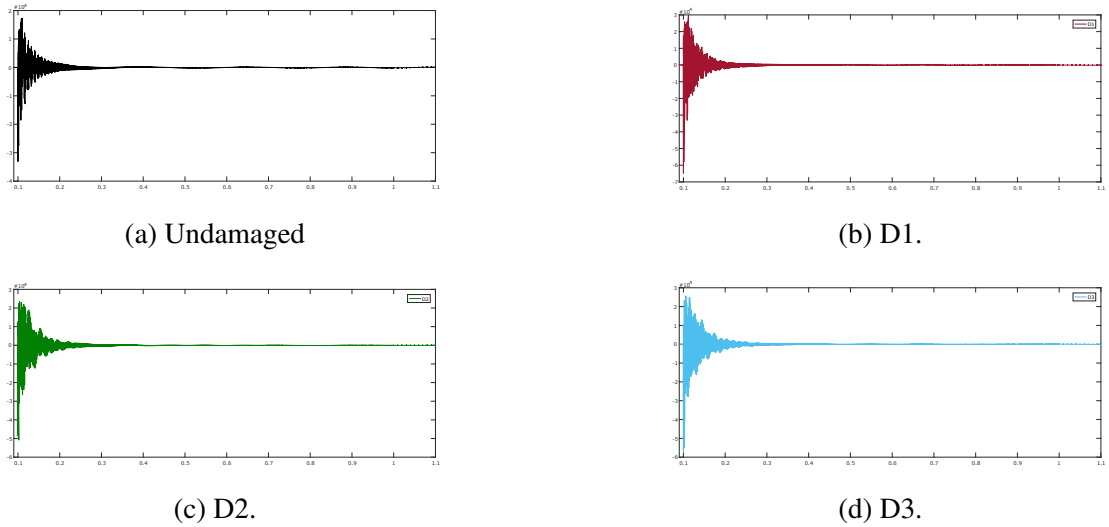
7.2.1 Time-domain approach

Dynamic responses for health and damaged beam in time-domain are shown in Fig. 7.12. All of them present the characteristic response of a damped system with *structural damping*. In this problem, it is complicated to observe with well precision changes between health and damaged samples responses, however, the feature extraction techniques must be able to identify damage sensitives, feeding the Artificial Neural Network (ANN) to detect and classify damages.

7.2.1.1 Dislocated-Series Time-domain

This proposal consists of subdividing the signals obtained for samples and using them directly in the neural network, leaving the task of selecting the necessary characteristics to classify certain types of damage, as previously applied to the rolling bearings. Therefore, the aim is to investigate the capacity of the technique in general, being used in a completely differ-

Figure 7.12 – Time-domain responses of samples corresponding to the different states.



(a) Undamaged

(b) D1.

(c) D2.

(d) D3.

Author's Production.

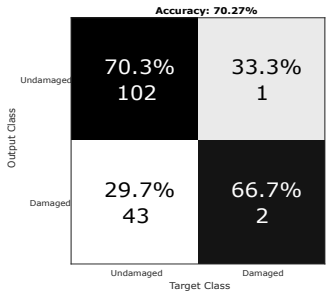
ent problem than the previous one, in order to test its generality. Initially, the algorithm for Dislocated-Series (DS), presented in the methodology must be applied, assuming values for the parameters k, n, and m. After that, a set of data will be obtained and sent to the neural network. In the first step, the objective will only detect the existence of damage to the structure, encompassing the first two steps of SHM. In the sequence, damages are classified in relation to their respective sizes.

Assuming $k = 2$, which determines the step to dislocate the signal. $n = 10$ for the number of mini-batches that will be stored in a matrix for a sample and $m = 20$ corresponding to the size of mini-batch, determining the number of input neurons to the network. At first, only two neurons were considered in the hidden layer (avoiding over-fitting) problems, which had the sigmoid (σ) activation function. The error function used was cross-entropy with softmax in the output layer. After performing some tests, using AdaMax as a training algorithm, the network presented the following results for 1000 epochs: 57.68% for training, 66.89% validation, and 70.27% in the testing phase. However, evaluating the confusion matrix for this case, according to Fig. 7.13(a), it presents high specificity, indicating that the model has poor generalization capacity. Hence, rather than modifying the number of neurons in the hidden layer, which could significantly improves the results, considering that this problem may be associated with under-fitting, the strategy adopted was to increase the number of input neurons before. Evaluating only the m influence. Therefore, as presented by the Tab. 7.8, when m increases the network performance improve. When small values are considered for m, the local features cannot be identified, making it difficult for the neural network determines am accurate boundary decision between the classes.

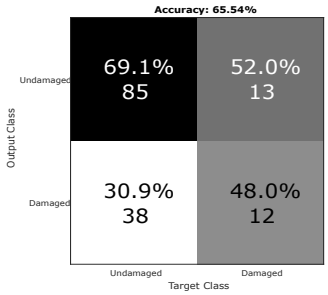
Table 7.8 – Network results for different numbers of input neurons, changing m.

m	Training	Validation	Testing
20	57.68%	66.89%	70.27%
40	69.34%	65.64%	65.54%
60	79.05%	68.91%	72.30%
80	82.60%	68.24%	72.97%
90	90.70%	81.08%	84.46%

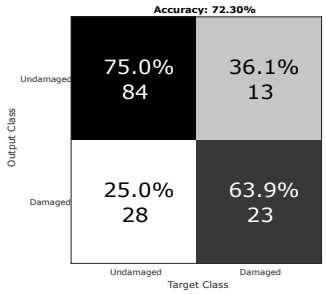
Figure 7.13 – Confusion matrix o the testing step for problems presented in tab. 7.8.



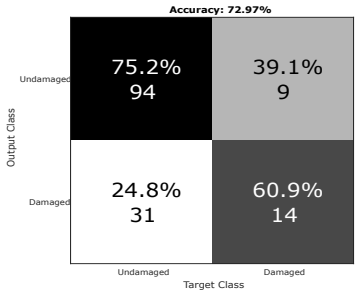
(a) $m = 20$.



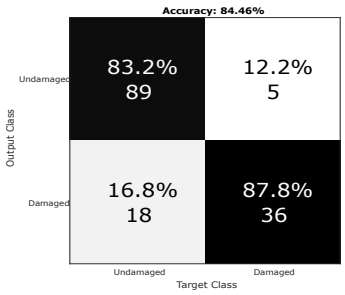
(b) $m = 40$.



(c) $m = 60$.



(d) $m = 80$.



(e) $m = 90$.

Author's Production.

From results presented by the confusion matrices, $m = 90$ was adopted, because, in addition to offering greater accuracy, it has more smoothed classification error between the two classes (healthy and damaged). Hence, the number of hidden neurons is increased to improve network generalization. It was made gradually as can be seen by Tab. 7.9. With 20, 25, and

30 neurons in the hidden layer, results not was satisfactory yet, until the next training and test results presented for 35 neurons establish an efficient choice in terms of reducing problems with under-fitting. When the highest increase was made in the number of neurons, the discrepancies between the results for training, validation, and testing show that the non-linear error regression curve is not so smooth, indicating a over-fitting problem.

After verifying that the model can generalize for damage detection, the next tests will be carried out to investigate its ability to classify them. Different models are shown in Tab. 7.10 were attributed, aiming to evaluate which offer the best generalization. Thus, with some tests performed, using $m = 80$, $n = 30$, and 45 neutrons in the hidden layer, 96.17 % accuracy was obtained in the testing, which represents an excellent result for the application in question. It can be seen in Fig. 7.14 the convergence curve for training and testing, using the cross-entropy (CE) error function. Figure 7.15 shows the distribution error through the confusion matrix, where it is observed that Healthy, D1, and D2 samples are classified accurately, while D3 presents some misclassification, reducing the efficiency of the model.

Dislocated-Series in a time-domain approach is useful to detect damage in composite materials. The training algorithm with the best results is RMSprop, ADAM, and AdaMAX that uses a stochastic gradient and adaptive learning, providing good convergence. However, the network working becomes slower than others because of the number of neurons required in the input layer, as well as in the hidden layer. In this case, the large difference between the different damages sizes permits that classification to be accurately. The errors observed mainly for D3 damage can be associated with uncertainties present in the system, which the tool used was not able to filter.

Table 7.9 – Network performance for a different amount of neuron used in the hidden layer.

Neurons	Training	Validation	Testing
20	98.90%	95.27%	97.30%
25	99.15%	94.59%	97.97%
30	98.64%	95.94%	95.25%
35	99.07%	99.32%	97.30%
40	99.07%	95.27%	92.27%
45	98.64%	96.62%	97.97%
50	99.40%	96.62%	94.55%
60	98.47%	93.91%	99.32%

7.2.1.2 Principal Component Analysis (PCA)

All data from time-domain response for the beams positions 1 and 2 are fully stored in a matrix $[G]_{m\times n}$, where m is the raw signal size evaluate to a unique sample, and n the sample number. Principal Component Analysis (PCA) is performed to obtain a orthogonal base, and make a projection, in order to a new reduced matrix $[G]_{m\times k}$, such as $k < n$. Therefore, in the

Table 7.10 – Models used to classify different damage types through Dislocated-Series in time-domain approach.

Topology	(n,m)	Activate Functions	Algorithm	Training	Testing	Validation
80-[25]-4	(30,80)	$[\sigma - \sigma]$	ADAM	99.74%	94.25%	93.98%
80-[30]-4	(30,80)	$[tanh - \sigma]$	RMSprop	99.32%	94.14%	93.92%
80-[45]-4	(30,80)	$[tanh - \sigma]$	AdaMax	99.29%	97.07%	96.17%
80-[35]-4	(30,80)	$[tanh - \sigma]$	RMSprop	99.29%	96.62%	96.17%

Figure 7.14 – Convergence to the classification model using AdaMax algorithm and 80-[45]-4 topology.

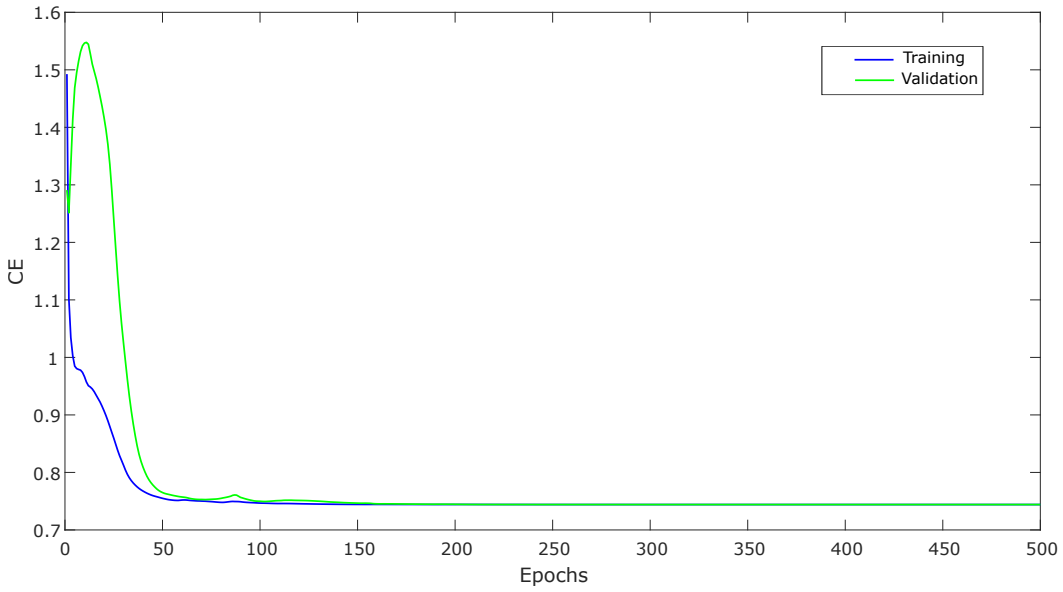
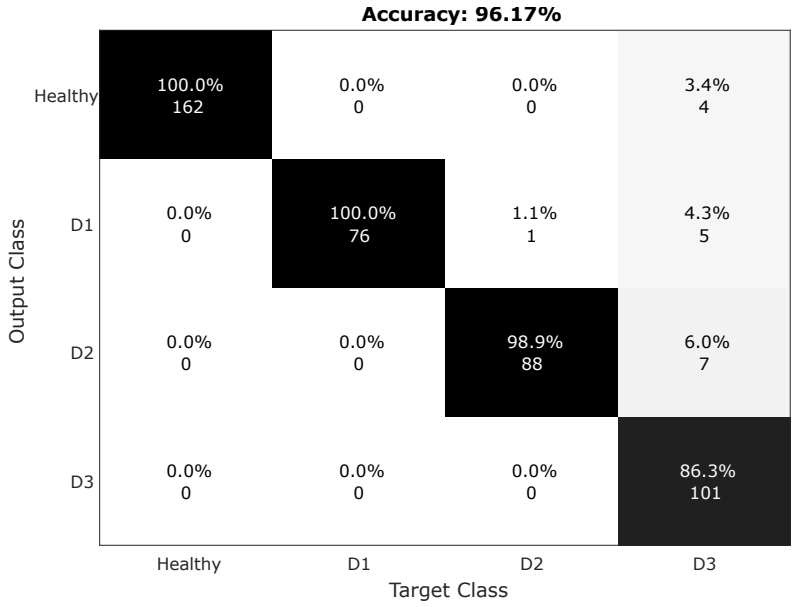
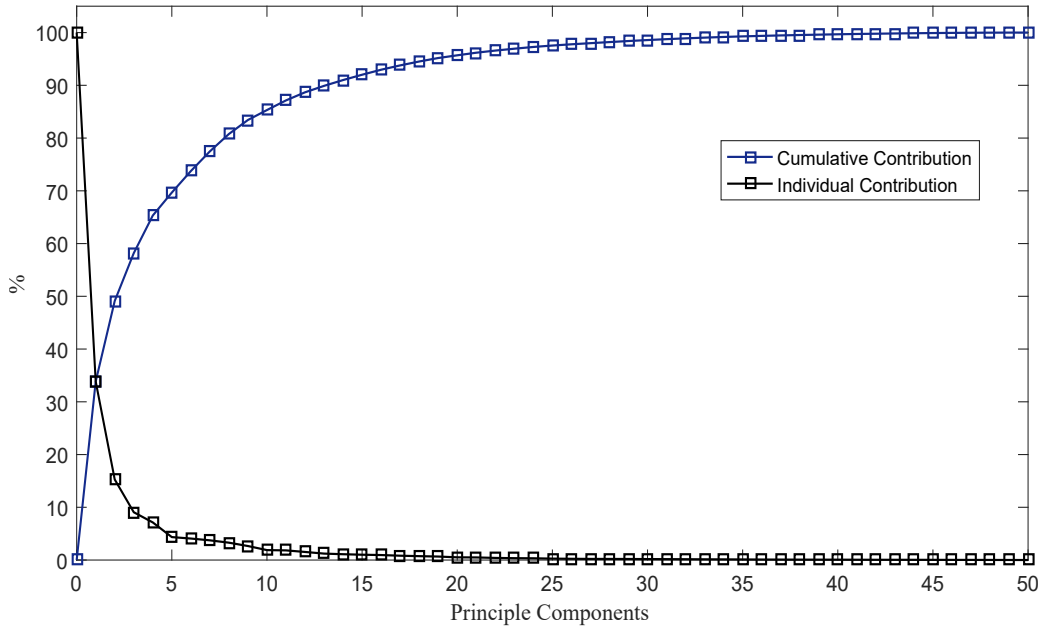


Figure 7.15 – Confusion matrix resulted in AdaMax algorithm training.



first time, it is needed to determine the number of PCs k that maximize the variance. Figure. 7.16 shows the accumulated and individual variance of up to 50 PCs, providing to look that, for 25 PCs, the variance reached approximately 100%, *i.e*, it is acceptable for the application.

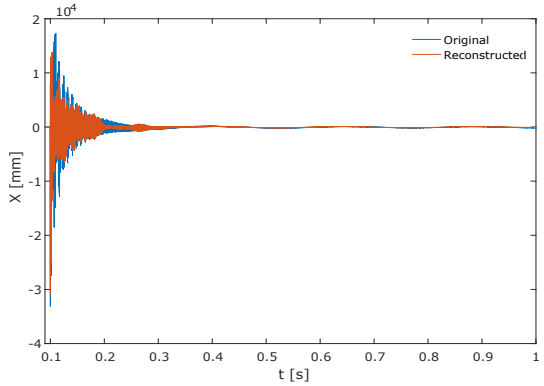
Figure 7.16 – Accumulated and individual variance for the principal components.



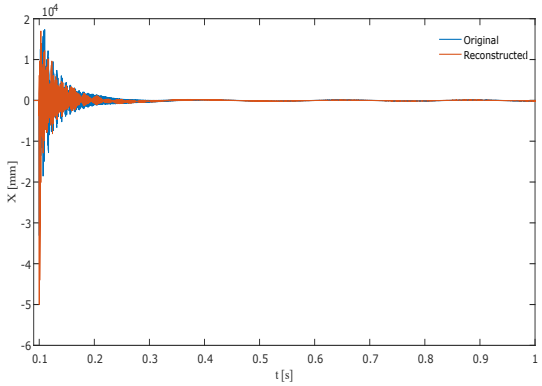
Following the theoretical background, 25 Principle Components (PCs) are sufficient is to make a signal reconstruction though an onto projection, as can be evaluated in Fig. 7.17. The reduction is performed for different PCs number, verifying its efficiency in the reconstruction, and validating the discussed in theory. It can be seen that for 5 PCs the variance accumulated is smaller than 70%, and, consequently, the reconstruction curve does not approach the original. For 20 PCs (*c.f.* Fig.7.17(b)), it is possible to see that the area in blue is filled with orange. Thus, it indicates better performance, until with 25 almost all the signal is already correctly reconstructed, demonstrating one of the main properties of the method, which is to represent an entire signal based on compressed information, os Principal Components (PCs).

After feature extraction is applied to the time-domain data set using PCA and signal reconstruction observation, it is required to analyze the damage sensibility for PCs to use as input in the pattern classification approach. So, taking 25 PCs for intact and damage samples, as shown in Fig. 7.18, it is possible to plot curves that correspond to the characteristics of each specific type of class. As can be analyzed, the curves have different amplitudes at each point, which is linked to the differences between dynamic responses. Therefore, the damage sensitivity analysis even allows us to verify that the PCA can make the differences between each signal more perceptible, and this can significantly influence the performance of the neural network. Then, different models were tested. Initially, using the learning rate $\eta = 0.001$ and moment term $\alpha = 0.99$, for a Momentum algorithm, with 15 neurons in the hidden layer. Data was divided into 80% for training, and 10% for testing and validation. The cross-entropy (CE)

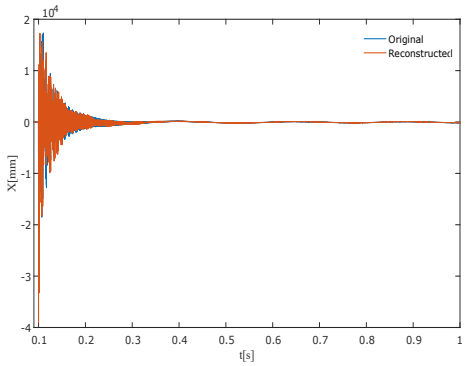
Figure 7.17 – Signals reconstruction via Principal Component Analysis in time-domain approach.



(a) 5 PCs.



(b) 20 PCs.



(c) 25 PCs.

Author's Production.

error function was used, as well as in the previous cases and the $tanh$ and sigmoid (σ) activation functions. The model was trained with 1000 epochs. For this first approach, good results were obtained, as shown in Fig. 7.19(a), 98.83%, validating the intelligent model to detect damage on composite beams. Another test was made with the AdaMax learning algorithm, and the good result was kept, this time 98.42% was reached in the validation. With the good results obtained, it can be inferred, therefore, that the established strategy works to detect damage in beams of composite materials. Thus, the size of the damage must be classified, specifically, according to the determined classes D1, D2, and D3.

Four neurons are considered in the output layer to classify the damage, corresponding each class to be classified through the network. Therefore, some models were considered, changing the number of neurons in the hidden layer, and learning algorithm, as presented in Tab. 7.11. The best performance was obtained for a neural network with $25-[15-10]-4$ topology, $[relu-relu-\sigma]$ for activation functions, and ADAM as algorithm learning by 500 epochs. Such that 93.39% is evaluated in the validation step, representing effective performance. In this case, specifically, the addition of another hidden layer with 10 neurons and the activation

Figure 7.18 – Damage sensitive using the PCA in time-domain.

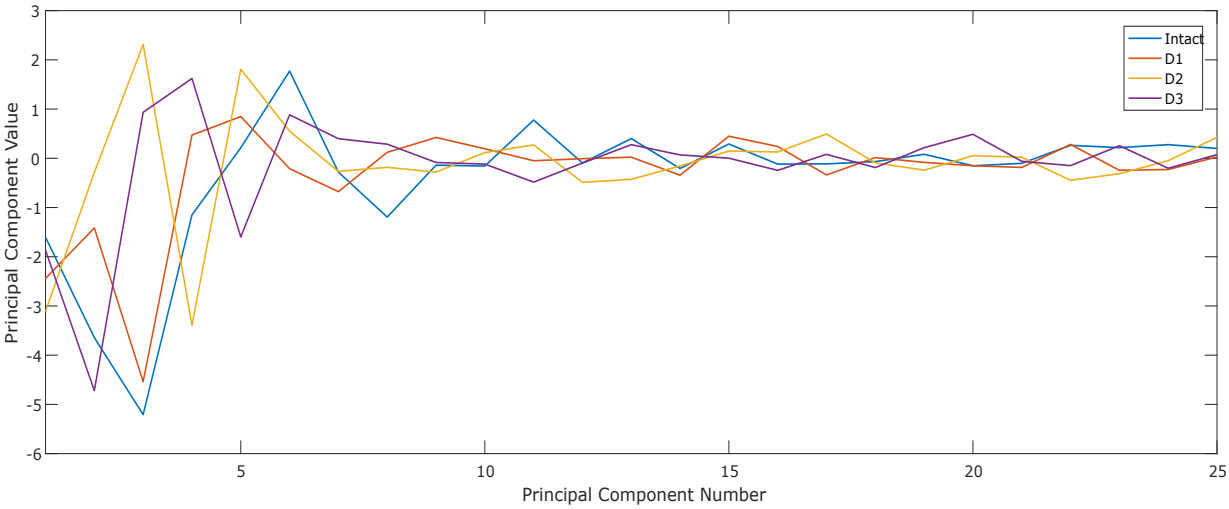
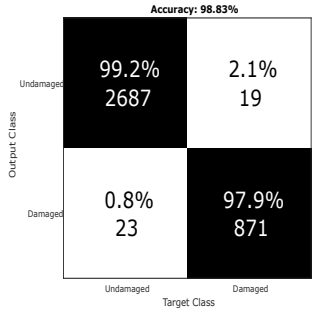
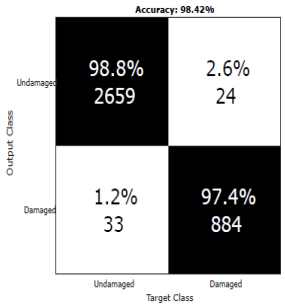


Figure 7.19 – Decision matrix for damage detection in composite beams using time-domain response through PCA data reducing.



(a) Confusion matrix for Steepest Descent with Momentum term.



(b) Confusion matrix for Adamax training algorithm.

Author's Production.

function $relu$ directly impact the improvements observed in the performance of the model, since this activation function does not saturate the gradient, allowing the adjustment of weights in minimizing error to achieve better results.

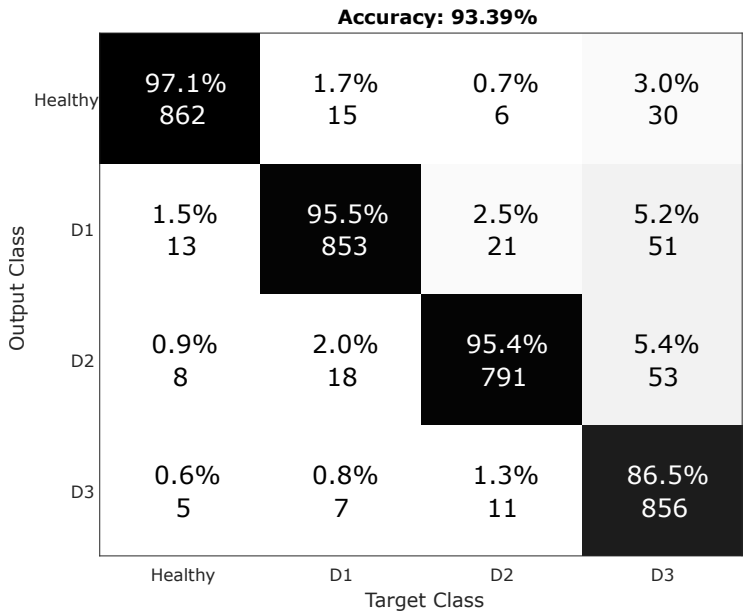
Table 7.11 – Models used to classify damage in the composite beams and its respective performance.

Topology	Activate Functions	Algorithm	Training	Validation	Testing
25-[15-10]-4	$[relu - relu - \sigma]$	ADAM	93.86%	92.75%	93.39%
25-[15]-4	$[tanh - \sigma]$	AdaMax	92.17%	90.05%	90.00%
25-[25]-4	$[tanh - \sigma]$	RMSprop	88.37%	86.52%	89.28%
25-[10-10]-4	$[tanh - tanh - \sigma]$	ADAM	87.43%	86.44%	87.06%

The confusion matrix of Fig. 7.20 shows the error distribution in classification, where

Healthy, D1, and D2 are classified with more than 95% accuracy. Most errors are found for the classification of D3 with false negatives for D2 and D3.

Figure 7.20 – Decision matrix to classify damage using 25 PCs in time-domain.

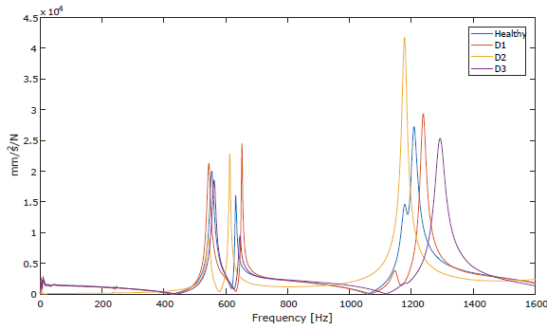


Author's Production.

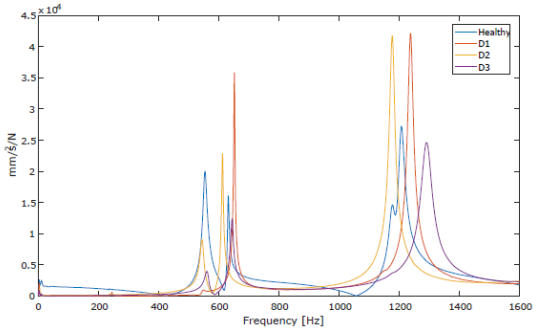
7.2.2 Frequency-domain approach

The frequency-domain approach corresponds to a good alternative in the context of the Vibration-Based Model (VBM) because it has more information about dynamic response of structures, which could increase ANN performance. Frequency Response Function (FRF) presents information about modal shapes, damping ratios, and resonance frequencies of the system. The signals extracted for the manufactured composite beams samples are shown in the Fig B.2-B.9. Taking some signals to verify the damage sensibility in the Fig. 7.21, it is possible see how damage affects the curves, associated to changes in the modal parameters. However, FRFs have large number of points and cannot be used directly in the Neural Networks. Hence two tools were used to extract the features; Dislocated-Series (DS) and Principal Component Analysis (PCA). In this case, the Statistical Parameters (SP) method has not applicability, because it was modeled for a random dynamic response generated by a Gaussian distribution of probabilities, which is not the case for signals in the frequency-domain.

Figure 7.21 – Frequency Response Functions for different composite beams state in the H_{11} and H_{21} position.



(a) Accelerometer position (H_{11}).



(b) Excitation position (H_{21}).

Author's Production.

7.2.3 Dislocated-Series Frequency-Domain

The same methodology used before in time-domain data is applied. The total signal evaluated is dislocated for n mini-batches with m points intercepted, which define the amount of input neurons. In this case, considering that undamaged and damaged samples has notable differences, best results for training and generalization are expected. Beginning, Table. 7.12 shows some models tested to detect damage on composite beams. As can be verified the model that used $90 - [40] - 2$ and $n = 40$ has obtained the best results, with 100.0% for training, 99.82% for testing, and 99.29% in the validation step. Cross-entropy (CE) as an error function and softmax in the output layer is used. It can be seen that when the input m improves, more hidden neurons are required, but the network presents a better generalization. Therefore, this application is useful for the first approach, such that the damage can be detected accurately.

Table 7.12 – Models used to detect damage on composite structures using series dislocated in frequency-domain.

Topology	n	Activate Functions	Algorithm	Training	Validation	Testing
60-[30]-2	40	$[tanh - \sigma]$	AdaMax	96.18%	94.68%	97.16%
60-[30]-2	40	$[tanh - \sigma]$	Momentum	98.89%	96.09%	97.16%
80-[30]-2	60	$[\sigma - \sigma]$	Momentum	99.95%	98.93%	99.53%
90-[40]-2	40	$[\sigma - \sigma]$	ADAM	100.0%	99.82	99.29%

In addition, to identify the damage, models must classify the different sizes corresponding to the induced delamination failures. In this way, the same strategy is used. Keeping the cross-entropy error function, placing 4 neurons in the last layer, and the softmax activation function, the number of hidden layers and neurons are modified, as well as the input m and n which determines the number of batches of the dislocated signal as show in Tab. 7.13.

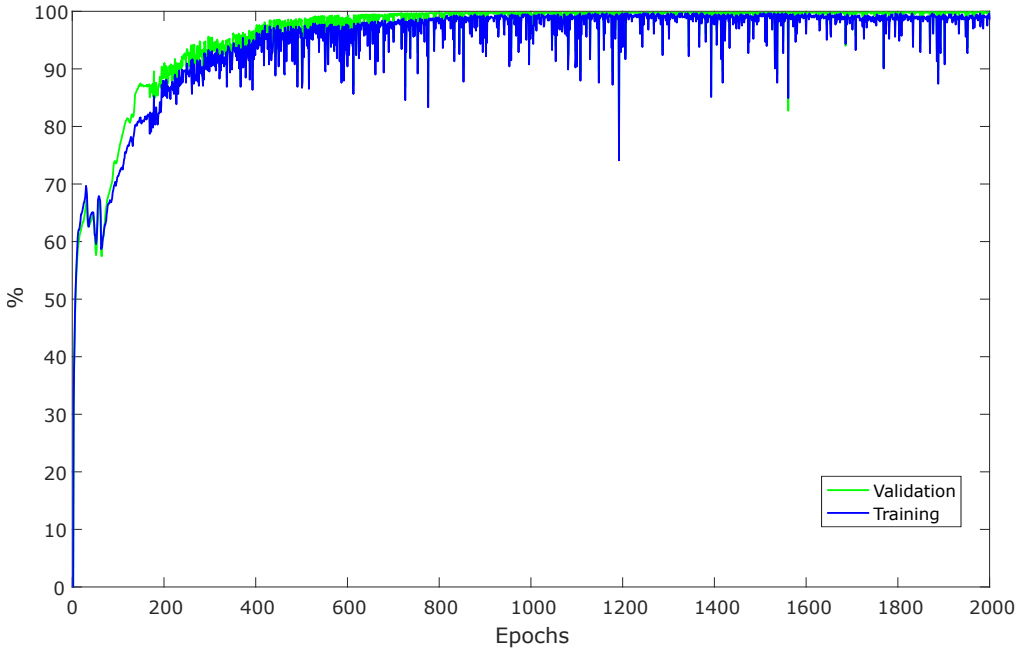
Table 7.13 – Models used to classify damages on composite beams using Series Dislocated in frequency-domain.

Topology	n	Activate Functions	Algorithm	Training	Validation	Testing
90-[35]-4	40	$[\sigma - \sigma]$	ADAM	99.84%	98.58%	99.29%
60-[20-15]-4	60	$[relu - relu - \sigma]$	AdaMax	77.86%	75.41%	75.18%
80-[20-15]-4	60	$[tanh - tanh - \sigma]$	RMSprop	96.30%	93.38%	93.26%
80[20-15]-4	60	$[tanh - tanh - \sigma]$	RMSprop	93.91%	92.08%	93.26%

The best result was found for the network with the highest m, indicating that the important properties necessary to assess the damage are extracted more precisely, whose distribution of the error is shown in the confusion matrix of the Fig.7.22. In this case, the damaged beams D1 and D2 are all correctly classified, while some classification errors arise for the Intact and D3, which may be associated with noise in the parts of the sampled signal. A Fig. 7.23 shows the convergence curves for training and testing the model, highlighting the presence of noise in the signal, which shows the complexity of the problem, however, good results were obtained at the end.

Figure 7.22 – Decision Matrix obtained to classify damages in the composite beams using Frequency-domain response.

Figure 7.23 – Training and testing curves for damage detection on composite beams using Frequency-domain response.



7.2.3.1 Principal Component Analysis (PCA)

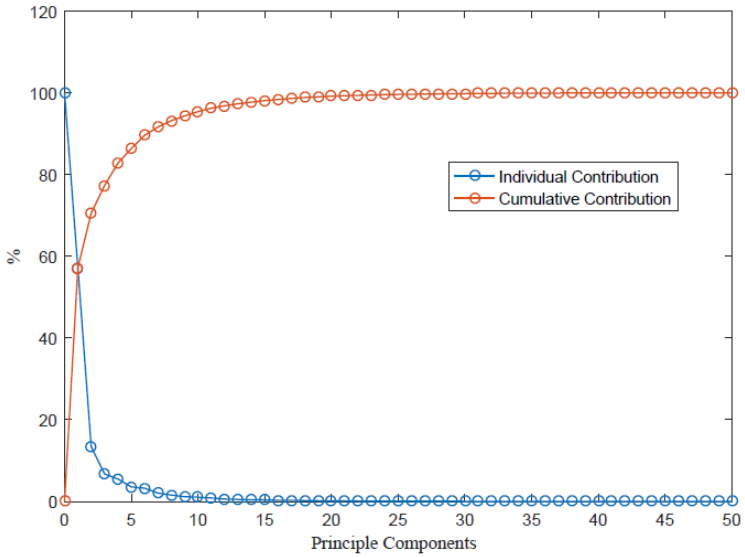
Each sample has FRF matrix of type $[H]_{2\times1600}$, because it was evaluated in two point (accelerometer and impact). As in the first time of the analyze is desirable only answer a binary question ”yes” or ”no” for the damage existence, the signals are divided in intact samples matrix $[HI]_{50\times1600}$, while damaged in a matrix $[HD]_{96\times1600}$. Therefore, PCA technique is performed, storing data information in correspondent PCs, making the dimension reduction through number of samples, keeping the amount of information to train the Neural Networks properly. So, the FRFs matrix resultant will be $H_{N_{PCs}\times1600}$.

After separating data for intact and damaged, it is necessary to evaluate the Principal Components (PCs) and check the associated variance, as can be seen in Fig. 7.24. It is noticed that for 30 PCs, the accumulated variance becomes very close to 100%, which determines that this is a good number to be used to compress the data.

One way to prove is by reconstructing the signals using different numbers of PCs, as shown in Fig. 7.25. It can be seen that for 10 PCs the accumulated variance is not sufficient to store the necessary information for reconstructing accurately the FRF. When 20 PCs were used, the variance improves and, the reconstructed signal become close to the original one. Therefore, the more PCs that are used, the reconstructed signal becomes closer to the original, as Fig. 7.25(c) illustrates for 30 PCs.

In this way, 30 PCs can carry with it all the information necessary to reconstruct the signal with maximum accuracy. Therefore, it will be used as input to the neural network, such that the numbers of hidden neurons activate the function, and the learning algorithm will be

Figure 7.24 – Accumulated and individual variance for the PCA in the frequency domain.



Author's Production.

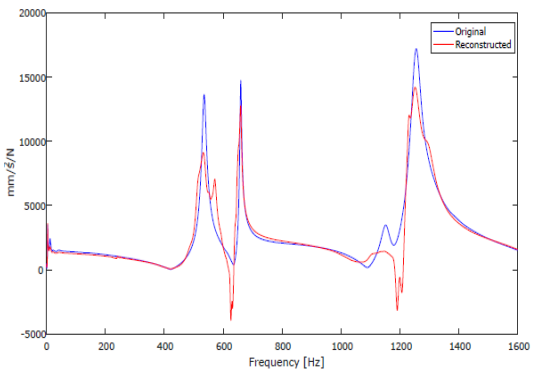
changed, in direction to obtain the best accurate model. Using cross-entropy as error function, softmax in the 2 output layer, the Tab. 7.14 shows the models and their respective performance. All of them have good results, with emphasis on what used ADAM as a learning algorithm, 8 neurons in the hidden and sigmoid layer for the activation functions (σ). The problem did not present great difficulties, with the convergence occurring for most of the tested models and good precision presented in its validation, which demonstrates the efficiency of the PCA technique for the treatment of FRFs, so that it is used in neural networks, requiring fewer models and faster to detect damage.

Table 7.14 – Models used to detect damage on composite beams for PCA approach and 200 epochs.

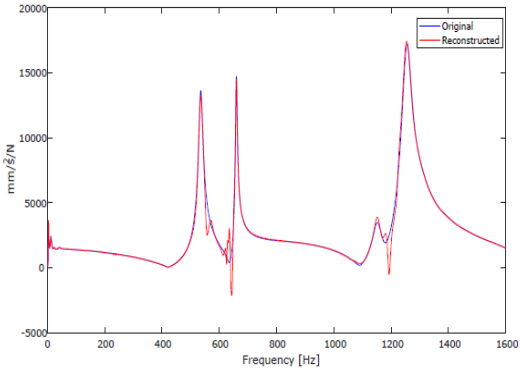
Hidden neurons	Activate Functions	Algorithm	Training	Validation	Testing
8	$[\sigma - \sigma]$	ADAM	99.78%	99.53%	99.53%
5	$[tanh - \sigma]$	AdaMax	99.57%	99.68%	99.45%
10	$[tanh - \sigma]$	Momentum	99.02%	98.75%	98.67%
7	$[\sigma - \sigma]$	RMSprop	99.14%	99.45%	98.98%

With the success of the methodology previously demonstrated for the damage detection problem, some necessary modifications are made to classified according to the different sizes of beam damage. Thus, in addition to 2 neurons in the output layer, 4 neurons must be used. Thus, Tab. 7.15 shows some models used in this step, such that the best result found was for the neural network with topology de 30-[30]-4, (σ) as an activation function, and ADAM in learning. Figure. 7.26 illustrates the decision matrix for this case, which has minimal classification errors in the model testing.

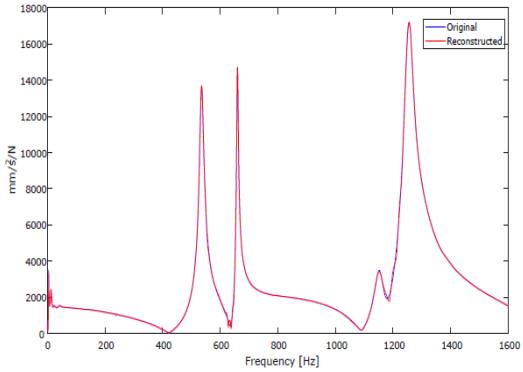
Figure 7.25 – Signals reconstruction using different PCs numbers.



(a) 10 PCs.



(b) 20 PCs.



(c) 30 PCs.

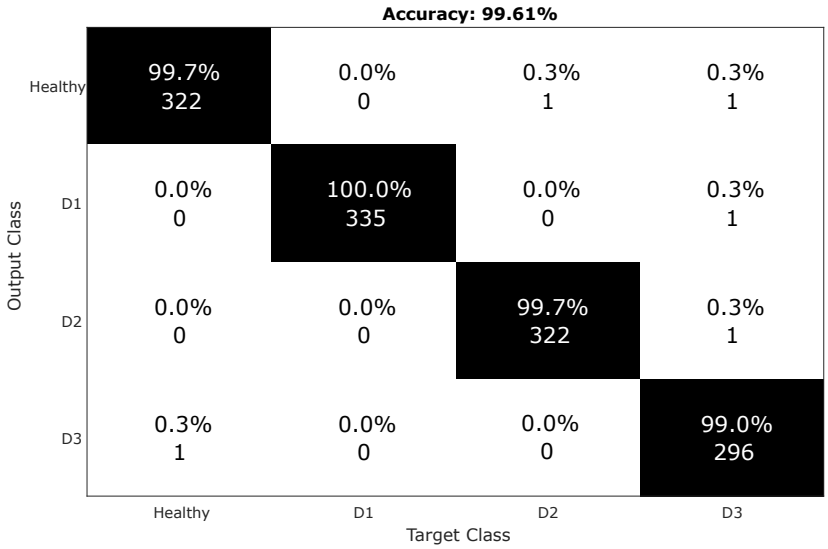
Author's Production.

Table 7.15 – Models used to detect damage on composite beams for the PCA approach and 1000 epochs.

Topology	Activate Functions	Algorithm	Training	Validation	Testing
30-[13]-4	$[\sigma - \sigma]$	RMSprop	97.02%	96.79%	97.03%
30-[20]-4	$[tanh - \sigma]$	AdaMax	99.90%	99.53%	99.53%
30-[25]-4	$[tanh - \sigma]$	Momentum	99.95%	99.84%	99.61%
30-[30]-4	$[\sigma - \sigma]$	ADAM	99.98%	99.68%	99.61%

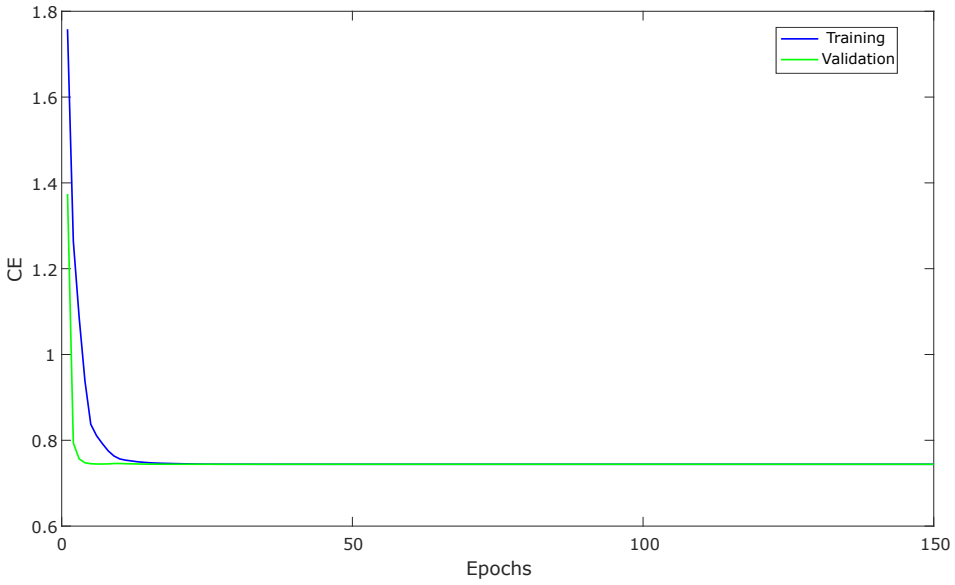
It is observed that using PCA, the noise is clear, and the efficiency of technique is improved, as can be seen by Fig. 7.27, where the training and testing curves are very smooth.

Figure 7.26 – Confusion matrix for classification problem using PCA in frequency-domain.



Author's Production.

Figure 7.27 – Network convergence curve with the topology of 30-[30]-4 using ADAM as a learning algorithm for 150 epochs.



Author's Production.

7.3 Numerical Model

Previous studies have shown that the use of neural networks in a supervised model is a good alternative for Structural Health Monitoring (SHM) applications because it provides damage detection efficiently in composite materials. However, this process requires a significant amount of data, what it presupposes manufactured samples, which can, often, increase the cost of the process. Thus, models based on vibration, modeled by the Finite Element Method (FEM),

based on modal analysis is commonly used. On the other hand, some researches have found a problem in its application, due to the effect of non-linearity on properties, boundary conditions, and dimensional uncertainties. The main idea of this section is to present a technique where dimensional uncertainties are considered to generate samples and train a neural network for damage detection.

7.3.1 Modal Analysis considering uncertainties

It was considered a numerical reference model that could have its properties adjusted according to the responses presented by the experimental one. A finite element model was created in the ABAQUS software, using around 400 SR8 shell elements, which have eight nodes in each element and a quadratic interpolation function, such that the boundary condition for the model was free-free following the experimental procedures. Initially, the properties used for the model were the same as those Völtz (2019) has used in their work, but the natural frequencies obtained were not close to the experimental range shown in Tab. 7.16. Nevertheless, after performing some analyses, the properties were adjusted, and natural frequencies within the ranges limited by the experiment were found, as can be seen in Tab. 7.17.

Table 7.16 – Experimental measurements for natural frequencies in composite beams.

State Beam	Mode 1 [Hz]	Mode 2 [Hz]	Mode 3 [Hz]	Mode 4 [Hz]	Mode 5 [Hz]
Healthy	223-251	467-572	594.5-659.5	1048-1201	1138-1300

Source: (VÖLTZ, 2019).

Table 7.17 – Numerical measurements for natural frequencies in composite beams.

State Beam	Mode 1 [Hz]	Mode 2 [Hz]	Mode 3 [Hz]	Mode 4 [Hz]	Mode 5 [Hz]
Healthy	238.11	546.21	639.86	1190.80	1218.30

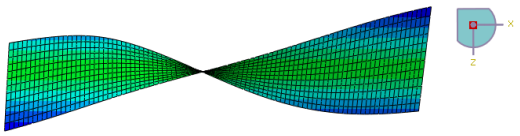
Author's production.

Complementing the results obtained in the modal analysis, Fig. 7.28 shown the five modal shapes considered, where the first and third are flexural modes, only. The second and fourth are torsional and, fifth a combination flexural-torsional.

However, to finalize the adjustment of the model and make it fit the presented methodology, the damping factors corresponding to the 5 modes must shape obtained in the 1500 Hz range considered be determined so that the numerical FRFs can be obtained. Therefore, some analyses were made by modifying the values corresponding to each damping ratio, and the setup considered was: $\zeta_1 = 0.0035$, $\zeta_2 = 0.025$, $\zeta_3 = 0.008$, $\zeta_4 = 0.025$, $\zeta_5 = 0.007$. These results were based on Fig. 7.29, where the red curve represents the response of the reference computational model while the others correspond to those obtained for manufactured samples.

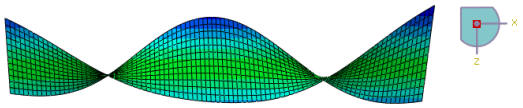
Figure 7.28 – Modal shapes for computational composite beams.

(a) Mode 1 (flexural).

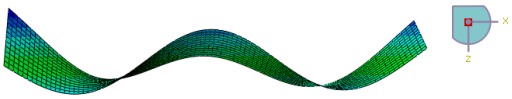


(b) Mode 2 (torsional).

(c) Mode 3 (flexural).



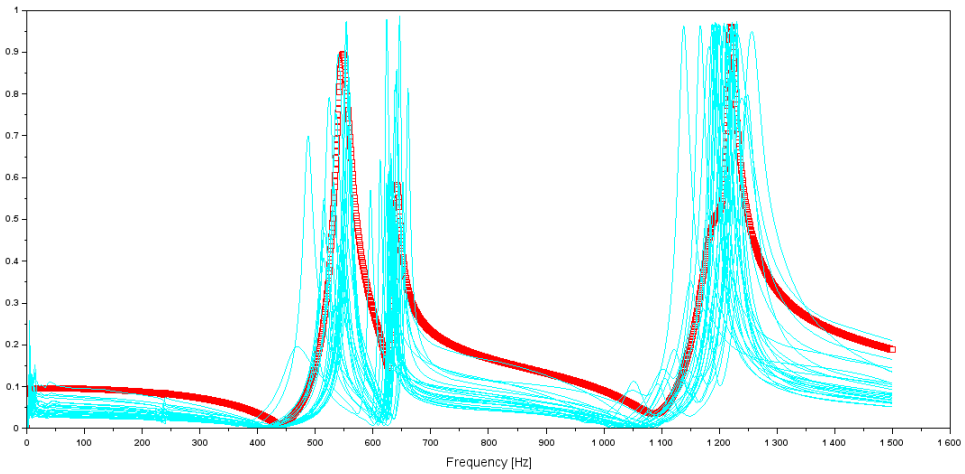
(d) Mode 4 (torsional).



(e) Mode 5 (flex-torsional).

Author's Production.

Figure 7.29 – Using a reference response for the numerical model to adjust the damping ratios of each mode.

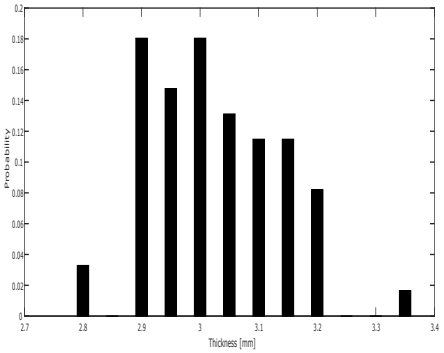


Author's Production.

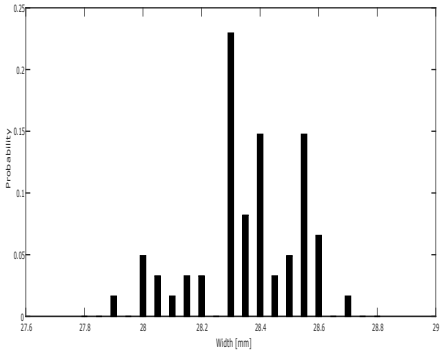
After completely model updating, adjusting its mechanical and dynamic properties so that its response is as close as possible to the experimental, the distribution of uncertainty obtained for each dimension is exposed in Fig. 7.30, and serve for samples to be created computationally and have their FRFs stored in a database to train neural networks in the job of identifying damage to composite materials. Thus 800 samples were generated, divided into 320

are Healthy, 160 D1, 160 D2, and 160 D3.

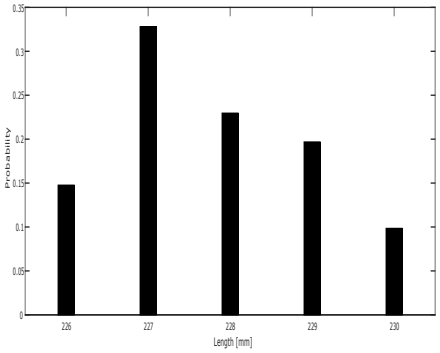
Figure 7.30 – Discrete distributions found for Width, Length and Thickness.



(a) Thickness distribution.



(b) Width distribution.



(c) Length distribution.

Author's Production.

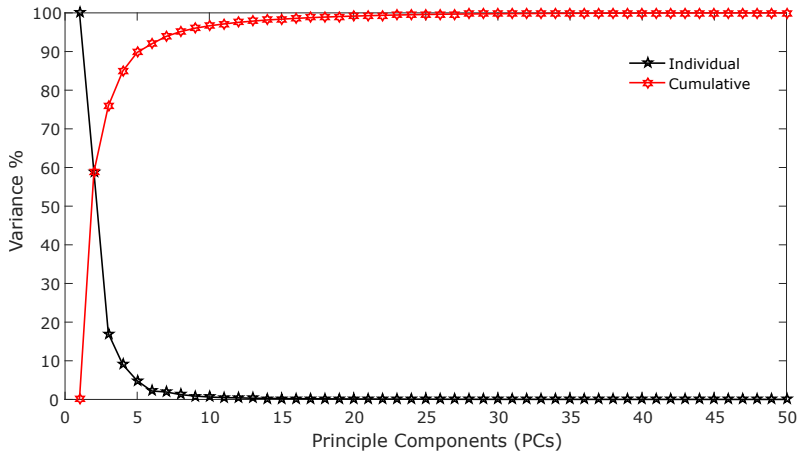
7.3.1.1 Principal Component Analysis (PCA)

For the FRFs matrix given by computational samples, after uncertainties application, there are two ways to apply Principal Component Analysis (PCA). The first one corresponds to a commonly used, where the size reduction is performed about the FRFs points. While the second possibility, used before in this work, the reduction occurs in terms of sample's number, maintaining the large FRF points, which help with more data than the first one to be used as neural network input.

Then, FRFs samples are grouped in a matrix, allocating each spectral line from H_{11} and H_{21} correspondent to individual samples, resulting in $[HI]_{1500\times320}$ (healthy), $[HD1]_{1500\times160}$ (damage 1), $[HD2]_{1500\times}160$ (damage 2), $[HD3]_{1500\times160}$ (damage 3). When the PCA is applied in order to reduce FRFs points, the matrix resultant will be $[H]_{PCs\times N_s}$, where PCs are the Principal Components used, and N_s number of samples.

Therefore, 30 PCs were chosen, determining the neural network input layer. The error function used was cross-entropy with softmax in the last layer, and sigmoid function for hidden

Figure 7.31 – Accumulated and individual variance for the PCA.



Author's production.

neurons. Table 7.18 shows two different topologies used, and the results presented.

Table 7.18 – Neural Networks performance using different setups.

Topology	Algorithm	Training	Validation	Testing
30 - [15] - 2	ADAM	100.0%	99.375%	98.75%
30 - [10] - 2	ADAM	99.68%	97.5%	97.5%

Through PCA alternative application, reducing the term corresponding to the number of samples, following the second approach mentioned above, obtaining the set of data that will be used as input for the neural network. Resulting in 6000 samples; $HI_{1500\times30}$, $HD1_{1500\times30}$, $HD2_{1500\times30}$, $HD3_{1500\times30}$ dividing it into 80% for the training phase, 10% for testing and validation. So, using cross-entropy as error function, softmax, and sigmoid for hidden neurons, the results obtained were better than the previous. ADAM algorithm presents the best performance than others tested for 2500 epochs, presented in Tab. 7.19, generalizes the model when only samples from computational models were used.

Table 7.19 – Result for the neural network in PCA alternative approach.

Topology	Algorithm	Training	Validation	Testing
30 - [15] - 2	ADAM	100.0%	100.0%	100.0%

Therefore, it is suggested that after the validation through computational samples, the same model be used to detect damage in a real structure, using vibration response from manufactured composite beams, investigating the uncertainties modeling. So a result was obtained corresponding to 70.95% correct classifications in the detection of damage, which is not a favorable result that allows the widespread use of this technique, but illustrates that there is feasibility,

as long as the model adjustment is best performed.

Chapter 8

Conclusions

It is possible to infer that Artificial Neural Networks can be used efficiently in damage detection problems, serving Structural Health Monitoring (SHM) applications. Promising results were obtained for all cases studied. It was analyzed that feature extraction tools have a significant influence on the neural network performance. In the rolling bearings fault diagnosis case, among the strategies used, Statistical parameters (SP) showed the best results, working as damage indicator for vibration data collected on time-domain. However, classifying damage not follow the results obtained for detection, revealing low sensitivity to the size of the damage, both for the ball and inner race faults. Another point to take into account is the correlation necessary among vibration response and data generated randomly by a Gaussian distribution, which limits the method utilization, and it cannot be applied to damage detection for composite beams.

On the other hand, Principal Component Analysis (PCA) shows good results for both detection and classification problems, using time-domain and frequency-domain data. Promising results were found, with more than 99% accuracy in the testing phase. Dislocated Series (DS) performed well too, but more neurons in the input layers were required to improve the results, adding computational costs, requiring more time to process the data. Therefore, the vibration data studied has intrinsic statistical properties and can be indexed on Principal Components (PCs) maximum variance-based.

After evaluating the functionality of this multidisciplinary methodology, their use on complex systems as plane's wings, industrial machines, and others, encompasses the characterization of main damages to generate reliable data and train the neural networks. However, manufacturing and labeling samples can be so expensive. In direction to offer an alternative, a promising strategy was approached using data from a numerical model, to reduce the quantity of the manufactured samples to train neural networks. To this end, uncertainties present in the dimensions were modeled using a discrete probability distribution for width, length, and thickness. After trained and generalized with numerical samples only, the same neural network receives data from the experiment and detects correctly 70% of the data set. It can be associated with the poor correlation between the experimental and numerical FRF verified on modal

analysis, because mechanical properties do not correctly determined.

With the results achieved, the studies carried out in this work encourage SHM as techniques for damage detection. However, it is suggested that tests be performed on structures with more complex geometries to verify the possible limitations of the model.

8.1 Future Works

- Verify the applicability of this method in more complex composite structures and damages.
- Study strategies based on unsupervised learning to eliminate the necessity of label the data set, which can be expensive in some cases.
- Apply probabilistic and non-probabilistic networks using the PCA as an alternative to the deterministic model used in this work.
- Apply Wavelet Transform (WT) in high frequencies, in direction to obtain more deep features.

Bibliography

ABDELJABER, O. et al. Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *Journal of Sound and Vibration*, Elsevier, v. 388, p. 154–170, 2017. ISSN 0022-460X. Disponível em: ⟨http://dx.doi.org/10.1016/j.jsv.2016.10.043⟩.

ACKLEY, D. H.; HINTON, G. E.; SEJNOWSKI, T. J. A learning algorithm for boltzmann machines. *Cognitive science*, Elsevier, v. 9, n. 1, p. 147–169, 1985.

AGGARWAL, C. C. Neural networks and deep learning. *Springer*, Springer, v. 10, p. 978–3, 2018.

ALLEMANG, R. J.; BROWN, D. L. A correlation coefficient for modal vector analysis. In: SEM ORLANDO. *Proceedings of the 1st international modal analysis conference*. [S.l.], 1982. v. 1, p. 110–116.

AMARI, S. A theory of adaptive pattern classifiers. *IEEE Transactions on Electronic Computers*, IEEE, n. 3, p. 299–307, 1967.

ANDERSON, T. L.; ANDERSON, T. L. *Fracture mechanics: fundamentals and applications*. [S.l.]: CRC press, 2005.

ANM. Parecer técnico n 07/2019 - gsbm/spm-esgj/lhpr/lpn/wan. Agência Nacional de Mineração, 2019.

AVITABILE, P. *Modal testing: a practitioner's guide*. [S.l.]: John Wiley & Sons, 2017.

BAKHARY, N.; HAO, H.; DEEKS, A. J. Damage detection using artificial neural network with consideration of uncertainties. *Engineering Structures*, v. 29, n. 11, p. 2806–2815, 2007. ISSN 01410296.

BALACHANDRAN, B.; MAGRAB, E. B. Vibrações mecânicas. *São Paulo: Cengage Learning*, 2011.

BISHOP, C. M. *Pattern recognition and machine learning*. [S.l.]: springer, 2006.

BISHOP, C. M. et al. *Neural networks for pattern recognition*. [S.l.]: Oxford university press, 1995.

BOSER, B. E.; GUYON, I. M.; VAPNIK, V. N. A training algorithm for optimal margin classifiers. In: ACM. *Proceedings of the fifth annual workshop on Computational learning theory*. [S.l.], 1992. p. 144–152.

BRAMER, M. *Principles of data mining*. [S.l.]: Springer, 2007. v. 180.

BROOMHEAD, D. S.; LOWE, D. *Radial basis functions, multi-variable functional interpolation and adaptive networks*. [S.l.], 1988.

CALLISTER, W. D.; RETHWISCH, D. G. et al. *Materials science and engineering: an introduction*. [S.l.]: John wiley & sons New York, 2007. v. 7.

CATBAS, F. N.; AKTAN, A. E. Condition and damage assessment: issues and some promising indices. *Journal of Structural Engineering*, American Society of Civil Engineers, v. 128, n. 8, p. 1026–1036, 2002.

CAWLEY, P.; ADAMS, R. D. The location of defects in structures from measurements of natural frequencies. *The Journal of Strain Analysis for Engineering Design*, SAGE Publications Sage UK: London, England, v. 14, n. 2, p. 49–57, 1979.

CHAKRABORTY, D. Artificial neural network based delamination prediction in laminated composites. *Materials & design*, Elsevier, v. 26, n. 1, p. 1–7, 2005.

CHOUDHURY, A.; TANDON, N. A theoretical model to predict vibration response of rolling bearings to distributed defects under radial load. 1998.

DACKERMANN, U. *Vibration-based damage identification methods for civil engineering structures using artificial neural networks*. Tese (Doutorado), 2009.

DHAMANDE, L.; CHAUDHARI, M. Bearing Fault Diagnosis Based on Statistical Feature Extraction in Time and Frequency Domain and Neural Network. *International Journal of Vehicle Structures and Systems*, v. 8, n. 4, p. 229–240, 2017. ISSN 0975-3060.

DOEBLING, S. W. et al. *Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review*. [S.l.], 1996.

DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern classification*. [S.l.]: John Wiley & Sons, 2012.

EWINS, D. J. *Modal testing: theory and practice*. [S.l.]: Research studies press Letchworth, 1984. v. 15.

FAA, A. 25.571-1d: Damage tolerance and fatigue evaluation of structure. *Federal Aviation Administration, Washington, DC*, v. 13, 2011.

FARRAR, C. R.; DOEBLING, S. W.; NIX, D. A. Vibration–based structural damage identification. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, The Royal Society, v. 359, n. 1778, p. 131–149, 2001.

FARRAR, C. R.; WORDEN, K. An introduction to structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, The Royal Society London, v. 365, n. 1851, p. 303–315, 2006.

FARRAR, C. R.; WORDEN, K. *Structural Health Monitoring.: A Machine Learning Perspective*. [S.l.]: John Wiley & Sons, 2012.

FERDINAND, P. The evolution of optical fiber sensors technologies during the 35 last years and their applications in structure health monitoring. In: . [S.l.: s.n.], 2014.

FORTMANN-ROE, S. *Understanding the Bias-Variance Tradeoff*. 2012. Disponível em: ⟨http://scott.fortmann-roe.com/docs/BiasVariance.html⟩.

FU, Z.-F.; HE, J. *Modal analysis*. [S.l.]: Elsevier, 2001.

GOMES, G. F. et al. The use of intelligent computational tools for damage detection and identification with an emphasis on composites – A review. *Composite Structures*, Elsevier, v. 196, n. April, p. 44–54, 2018. ISSN 02638223. Disponível em: ⟨https://doi.org/10.1016/j.compstruct.2018.05.002⟩.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. [S.l.]: MIT press, 2016.

GÜEMES, A. et al. Structural health monitoring for advanced composite structures: A review. *Journal of Composites Science*, Multidisciplinary Digital Publishing Institute, v. 4, n. 1, p. 13, 2020.

GUO, T. et al. Damage detection in a novel deep-learning framework: a robust method for feature extraction. *Structural Health Monitoring*, v. 19, n. 2, p. 424–442, 2020. ISSN 17413168.

HAY, A. The derivation of global estimates from a confusion matrix. *International Journal of Remote Sensing*, Taylor & Francis, v. 9, n. 8, p. 1395–1398, 1988.

HAYKIN, S. *Redes neurais: princípios e prática*. [S.l.]: Bookman Editora, 2007.

HEATON, J. *AIFH, Volume 3: Deep Learning and Neural Networks*. [S.l.: s.n.], 2015.

HEBB, D. O. *The organization of behavior: a neuropsychological theory*. [S.l.]: J. Wiley; Chapman & Hall, 1949.

HEIDA, J. H.; PLATENKAMP, D. J. In-service inspection guidelines for composite aerospace structures. In: *18th World Conference on Nondestructive Testing*. [S.l.: s.n.], 2012. p. 16–20.

HINTON, G. E.; OSINDERO, S.; TEH, Y.-W. A fast learning algorithm for deep belief nets. *Neural computation*, MIT Press, v. 18, n. 7, p. 1527–1554, 2006.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, Warwick & York, v. 24, n. 6, p. 417, 1933.

HOYER, P. O.; HYVÄRINEN, A. Independent component analysis applied to feature extraction from colour and stereo images. *Network: computation in neural systems*, Taylor & Francis, v. 11, n. 3, p. 191–210, 2000.

HU, W.; YANG, J.; HOU, Y. Damage Detection Method of Ancient Timber Structure Based on BP Neural Network and Total Wavelet Energy Rate. *IOP Conference Series: Materials Science and Engineering*, v. 780, n. 3, p. 0–6, 2020. ISSN 1757899X.

HUMAR, J.; BAGCHI, A.; XU, H. Performance of vibration-based techniques for the identification of structural damage. *Structural Health Monitoring*, Sage Publications Sage CA: Thousand Oaks, CA, v. 5, n. 3, p. 215–241, 2006.

IBRAHIM, I. D. et al. The use of polypropylene in bamboo fibre composites and their mechanical properties–a review. *Journal of Reinforced Plastics and Composites*, SAGE Publications Sage UK: London, England, v. 34, n. 16, p. 1347–1356, 2015.

INMAN, D. J.; SINGH, R. C. *Engineering vibration*. [S.l.]: Prentice Hall Englewood Cliffs, NJ, 1994. v. 3.

Jiong Tang. Frequency response based damage detection using principal component analysis. *2005 IEEE International Conference on Information Acquisition*, IEEE, p. 407–412, 2006.

JONES, R. M. *Mechanics of composite materials*. [S.l.]: CRC press, 1999.

KATUNIN, A.; DRAGAN, K.; DZIENDZIKOWSKI, M. Damage identification in aircraft composite structures: A case study using various non-destructive testing techniques. *Composite structures*, Elsevier, v. 127, p. 1–9, 2015.

KAW, A. K. *Mechanics of composite materials*. [S.l.]: CRC press, 2005.

KELLY, S. G. *Mechanical vibrations: theory and applications*. [S.l.]: Cengage learning Stamford, CT, 2012.

KESSLER, S. S. et al. Damage detection in composite materials using frequency response methods. *Composites Part B:Engineering*, v. 33, n. 1, p. 87–95, 2002. ISSN 13598368.

KHOSHNOUDIAN, F.; TALAEI, S.; FALLAHIAN, M. Structural damage detection using frf data, 2d-pca, artificial neural networks and imperialist competitive algorithm simultaneously. *International Journal of Structural Stability and Dynamics*, World Scientific, v. 17, n. 07, p. 1750073, 2017.

KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

KUMAR, A. et al. Bearing defect size assessment using wavelet transform based Deep Convolutional Neural Network (DCNN). *Alexandria Engineering Journal*, Faculty of Engineering, Alexandria University, v. 59, n. 2, p. 999–1012, 2020. ISSN 11100168. Disponível em: ⟨https://doi.org/10.1016/j.aej.2020.03.034⟩.

KUNCHEVA, L. Combining pattern classifiers methods and algorithms. john wiley&sons. *Inc. Publication, Hoboken*, 2004.

LI, H.; HUANG, J.; JI, S. Bearing fault diagnosis with a feature fusion method based on an ensemble convolutional neural network and deep neural network. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 19, n. 9, p. 2034, 2019.

LIU, R. et al. Dislocated Time Series Convolutional Neural Architecture: An Intelligent Fault Diagnosis Approach for Electric Machine. *IEEE Transactions on Industrial Informatics*, IEEE, v. 13, n. 3, p. 1310–1320, 2017. ISSN 15513203.

LOPARO, K. A. *Western Reserve University Bearing Data Center Website*. 2013.

MARR, B. A short history of machine learning–every manager should read. *Forbes. http://tinyurl. com/gslvr6k*, 2016.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *Bulletin of mathematical biophysics*, v. 5, n. 4, p. 115–133, 1943.

MEDEIROS, R. D.; VANDEPITTE, D.; TITA, V. Structural health monitoring for impact damaged composite: a new methodology based on a combination of techniques. *Structural Health Monitoring*, SAGE Publications Sage UK: London, England, v. 17, n. 2, p. 185–200, 2018.

MEDEIROS, R. de et al. A new methodology for structural health monitoring applications. *Procedia Engineering*, Elsevier; AMSTERDAM, v. 114, p. 54–61, 2015.

MEIROVITCH, L. *Fundamentals of vibrations*. [S.l.]: Waveland Press, 2010.

MESSINA, A.; WILLIAMS, E.; CONTURSI, T. Structural damage detection by a sensitivity and statistical-based method. *Journal of sound and vibration*, Elsevier, v. 216, n. 5, p. 791–808, 1998.

MICHAELS, K. *Opinion: OEMs Focus On Mature Aircraft For Aftermarket Growth*. 2018. Disponível em: ⟨http://aviationweek.com/commercial-aviation/opinion-oems-focus-mature-aircraft-aftermarket-growth⟩.

MONTALVAO, D.; MAIA, N. M. M.; RIBEIRO, A. M. R. A review of vibration-based structural health monitoring with special emphasis on composite materials. *Shock and vibration digest*, Bournemouth University, Fern Barrow, Poole, Dorset, BH12 5BB, UK, v. 38, n. 4, p. 295–324, 2006.

NETO, F. L.; PARDINI, L. C. *Compósitos estruturais: ciência e tecnologia*. [S.l.]: Editora Blucher, 2016.

NGUYEN-SCHÄFER, H. *Computational design of rolling bearings*. [S.l.]: Springer, 2016.

OKAFOR, A. C.; CHANDRASHEKHARA, Y. P. J. K. Delamination prediction in composite beams with built-in piezoelectric devices using modal analysis and neural network. v. 338, 1996.

OOIJEVAAR, T. *Vibration based structural health monitoring of composite skin-stiffener structures*. Tese (Doutorado) — University of Twente, 3 2014.

PADIL, K. H. et al. Non-probabilistic method to consider uncertainties in frequency response function for vibration-based damage detection using artificial neural network. *Journal of Sound and Vibration*, Elsevier, v. 467, p. 115069, 2020.

PADIL, K. H.; BAKHARY, N.; HAO, H. The use of a non-probabilistic artificial neural network to consider uncertainties in vibration-based-damage detection. *Mechanical Systems and Signal Processing*, Elsevier, v. 83, p. 194–209, 2017.

PATIL, A. B.; GAIKWAD, J. A.; KULKARNI, J. V. Bearing fault diagnosis using discrete wavelet transform and artificial neural network. In: IEEE. *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*. [S.l.], 2016. p. 399–405.

PEARSON, K. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, Taylor & Francis, v. 2, n. 11, p. 559–572, 1901.

PIAZZAROLI, R. An SHM approach using machine learning and statistical indicators extracted from raw dynamic measurements. v. 16, n. 1998, p. 1–17, 2019.

PRASAD, J. R. Pattern recognition: possible research areas and issues. *International Journal of Computer Science and Network*, v. 3, n. 5, 2014.

RAO, S. S. Mechanical vibrations(laboratory manual). Year, Edition Addison-Wesley Publishing Company, ISBN0-201-52686-7, 5th Edition., 1995.

RAO, S. S. Engineering optimization: Theory and practice. John Wiley and Sons, 2009.

REDDY, J. N. *Mechanics of Laminated Composite Plates and Shells*. CRC Press, 2004. Disponível em: ⟨https://doi.org/10.1201/b12409⟩.

REZENDE, M. C. Fractografia de compósitos estruturais. *Polímeros*, SciELO Brasil, v. 17, n. 3, p. E4–E11, 2007.

RHIM, J.; LEE, S. W. A neural network approach for damage detection. v. 16, p. 437–443, 1995.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, American Psychological Association, v. 65, n. 6, p. 386, 1958.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *nature*, Nature Publishing Group, v. 323, n. 6088, p. 533–536, 1986.

SHIGLEY, J. E. *Shigley's mechanical engineering design*. [S.l.]: Tata McGraw-Hill Education, 2011.

SILBERSCHMIDT, V. V. *Dynamic deformation, damage and fracture in composite materials and structures*. [S.l.]: Woodhead Publishing, 2016.

SILVA, I. d.; SPATTI, D. H.; FLAUZINO, R. A. Redes neurais artificiais para engenharia e ciências aplicadas. *São Paulo: Artliber*, v. 23, n. 5, p. 33–111, 2010.

SRIDHARAN, S. *Delamination behaviour of composites*. [S.l.]: Elsevier, 2008.

SUN, J.; YAN, C.; WEN, J. Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning. *IEEE Transactions on Instrumentation and Measurement*, IEEE, v. 67, n. 1, p. 185–195, 2017.

SZEGEDY, C. et al. Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2015. p. 1–9.

TALREJA, R.; SINGH, C. V. *Damage and failure of composite materials*. [S.l.]: Cambridge University Press, 2012.

TALREJA, R.; VARNA, J. *Modeling damage, fatigue and failure of composite materials*. [S.l.]: Elsevier, 2015.

TANDON, N.; CHOUDHURY, A. A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. *Tribology international*, Elsevier, v. 32, n. 8, p. 469–480, 1999.

TANDON, N.; NAKRA, B. Comparison of vibration and acoustic measurement techniques for the condition monitoring of rolling element bearings. *Tribology International*, Elsevier, v. 25, n. 3, p. 205–212, 1992.

TANG, Z. et al. Convolutional neural network-based data anomaly detection method using multiple information for structural health monitoring. *Structural Control and Health Monitoring*, v. 26, n. 1, p. 1–22, 2019. ISSN 15452263.

TENG, S. et al. Structural damage detection using convolutional neural networks combining strain energy and dynamic response. *Meccanica*, Springer Netherlands, v. 55, n. 4, p. 945–959, 2020. ISSN 15729648. Disponível em: ⟨https://doi.org/10.1007/s11012-019-01052-w⟩.

TIPPING, M. E.; BISHOP, C. M. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Wiley Online Library, v. 61, n. 3, p. 611–622, 1999.

UN. World population prospects 2019: Highlights (st/esa/ser. a/423). United Nations (UN), Department of Economic and Social Affairs, Population . . . , 2019.

VAPNIK, V.; CORTES, C. *Support Vector Networks, machine learning 20, 273–297*. [S.l.]: Kunwer Acedemic Publisher, 1995.

VAPNIK, V. N. An overview of statistical learning theory. *IEEE transactions on neural networks*, Citeseer, v. 10, n. 5, p. 988–999, 1999.

VERSTRAETE, D. B. et al. Deep semi-supervised generative adversarial fault diagnostics of rolling element bearings. *Structural Health Monitoring*, SAGE Publications Sage UK: London, England, v. 19, n. 2, p. 390–411, 2020.

VÖLTZ, L. R. *Fault Diagnosis in composite structures using Artificial Neural Networks and Principal Component Analysis*. Dissertação (Mestrado) — State University of Santa Catarina, Joinville, Santa Catarina, Brazil, 3 2019.

WANG, P. et al. Virtualization and deep recognition for system fault classification. *Journal of Manufacturing Systems*, The Society of Manufacturing Engineers, v. 44, p. 310–316, 2017. ISSN 02786125. Disponível em: ⟨http://dx.doi.org/10.1016/j.jmsy.2017.04.012⟩.

WHITE, P.; TAN, M.; HAMMOND, J. Analysis of the maximum likelihood, total least squares and principal component approaches for frequency response function estimation. *Journal of Sound and Vibration*, Elsevier, v. 290, n. 3-5, p. 676–689, 2006.

WIDROW, B.; HOFF, M. E. *Adaptive switching circuits*. [S.l.], 1960.

XIAO, W. Structural health monitoring and fault diagnosis based on artificial immune system. Worcester Polytechnic Institute, 2012.

YAM, L. H.; YAN, Y. J.; JIANG, J. S. Vibration-based damage detection for composite structures using wavelet transform and neural network identification. *Composite Structures*, v. 60, n. 4, p. 403–412, 2003. ISSN 02638223.

ZANATTA, R. *MATERIAIS COMPÓSITOS NA AVIAÇÃO*. 2012. Disponível em: ⟨http://www.aviacao.org/article/materiais-compositos/⟩.

ZHAO, R. et al. Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing*, Elsevier, v. 115, p. 213–237, 2019.

ZHU, W. et al. Sensitivity, specificity, accuracy, associated confidence interval and roc analysis with practical sas implementations. *NESUG proceedings: health care and life sciences, Baltimore, Maryland*, v. 19, p. 67, 2010.

ZOU, Y.; TONG, L.; STEVEN, G. P. Vibration-based model-dependent damage (delamination) identification and health monitoring for composite structures—a review. *Journal of Sound and vibration*, Elsevier, v. 230, n. 2, p. 357–378, 2000.

Appendix A

Auxiliary discussion

A.1 Orthogonality of modal vectors

The vibration modes of the structure have an important property in relation to mass matrices $[M]$ and stiffness $[K]$ for solving problems with multiple degrees of freedom (DOF). Considering the eigenvalue problem established in the (2.2.8). Two solutions are assumed

$$[K]\{\phi_r\} = \omega_r^2[M]\{\phi_r\}, \quad [K]\{\phi_s\} = \omega_r^2[M]\{\phi_s\}. \tag{A.1}$$

Premultiplying both sides of the first and second equation from Eq. (A.1) by $\{\phi_s\}^T$ e $\{\phi_r\}^T$,

$$\{\phi_s\}^T[K]\{\phi_r\} = \omega_r^2\{\phi_s\}^T[M]\{\phi_r\}, \quad \{\phi_r\}^T[K]\{\phi_s\} = \omega_r^2\{\phi_r\}^T[M]\{\phi_s\}. \tag{A.2}$$

Transposing the second equation of (A.2) and replacing in the first, the relation can be obtained

$$\left(\omega_r^2 - \omega_s^2\right)\{\phi_s\}^T[M]\{\phi_r\} = 0. \tag{A.3}$$

As in general the natural frequencies are distinct $\omega_s \neq \omega_r$, Eq. (A.3) is satisfied provided

$$\{\phi_s\}^T[M]\{\phi_r\} = 0, \quad r \neq s, \tag{A.4}$$

which represent the *orthogonality relation* for the modal vectors $\{\phi_r\}$ and $\{\phi_s\}$ with the inertia matrix. Inserting Eq. (A.4) into the first of Eq. (A.3), it can be see that modal vectors are orthogonal with respect to the stiffness matrix $[K]$,

$$\{\phi_s\}^T[K]\{\phi_r\} = 0, \quad r \neq s. \tag{A.5}$$

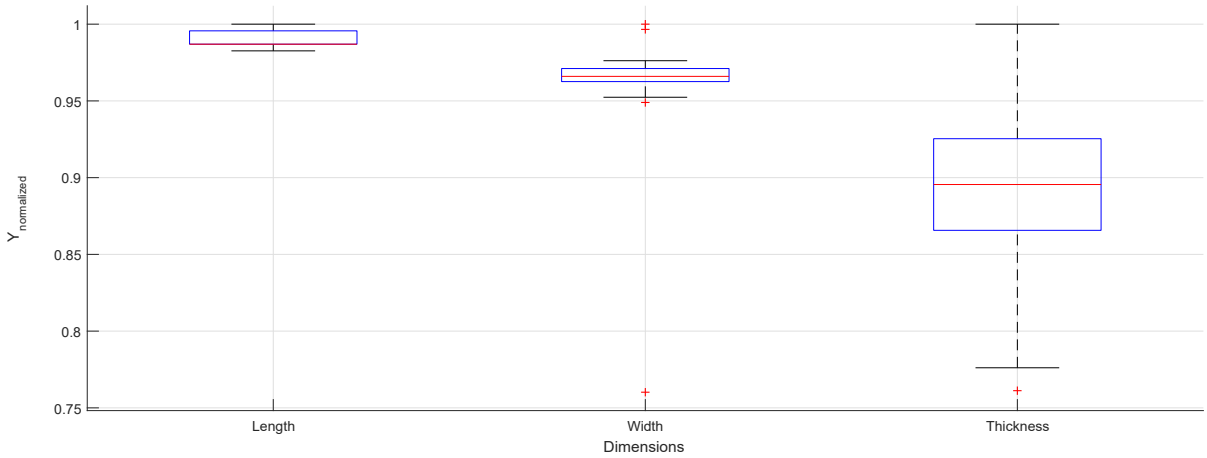
It is worth mentioning that the orthogonality property is valid only when the matrix $[M]$ and

$[K]$ are real positive definite and symmetric (MEIROVITCH, 2010).

Appendix B

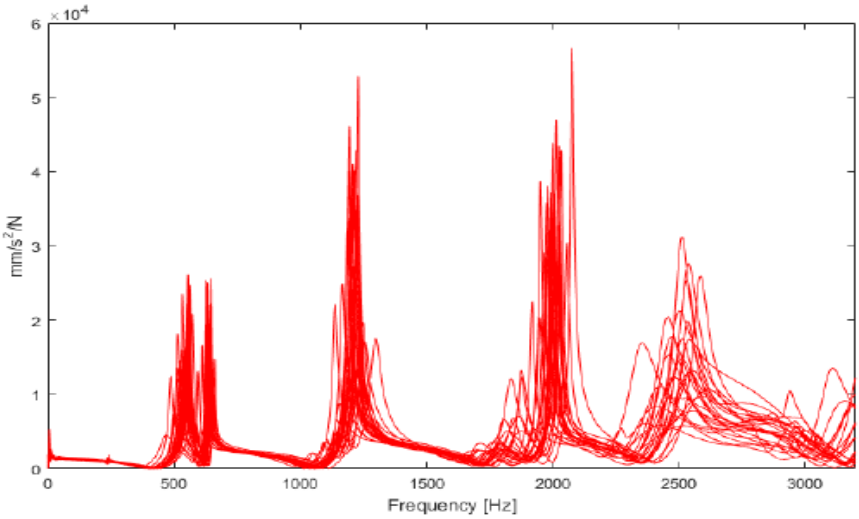
Graphics

Figure B.1 – Individual analysis of sample dimensions.



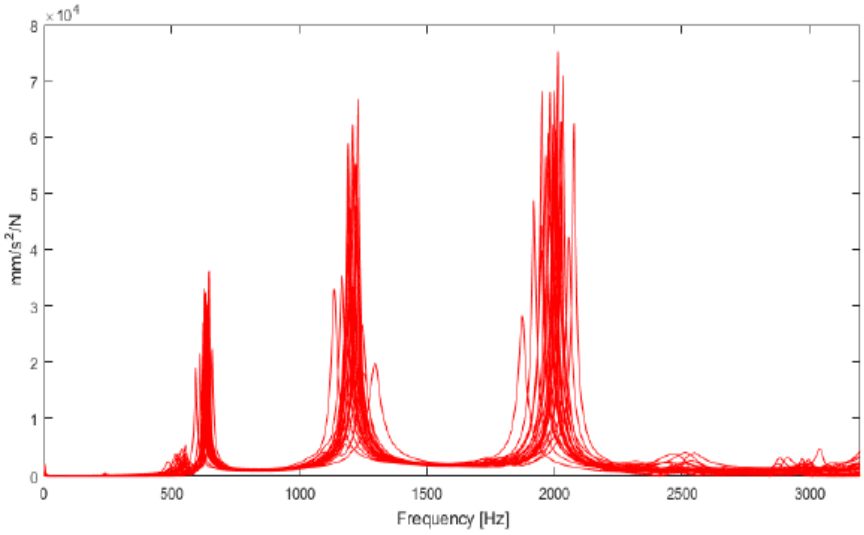
Source: Author's production

Figure B.2 – Frequency response functions for all Healthy samples in the H_{11} position.



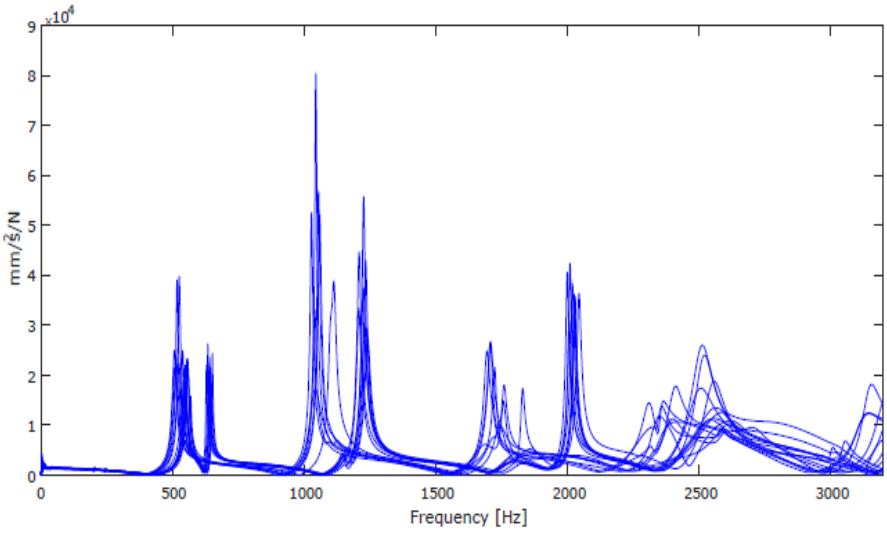
Source: (VÖLTZ, 2019).

Figure B.3 – Frequency response functions for all Healthy samples in H_{21} position.



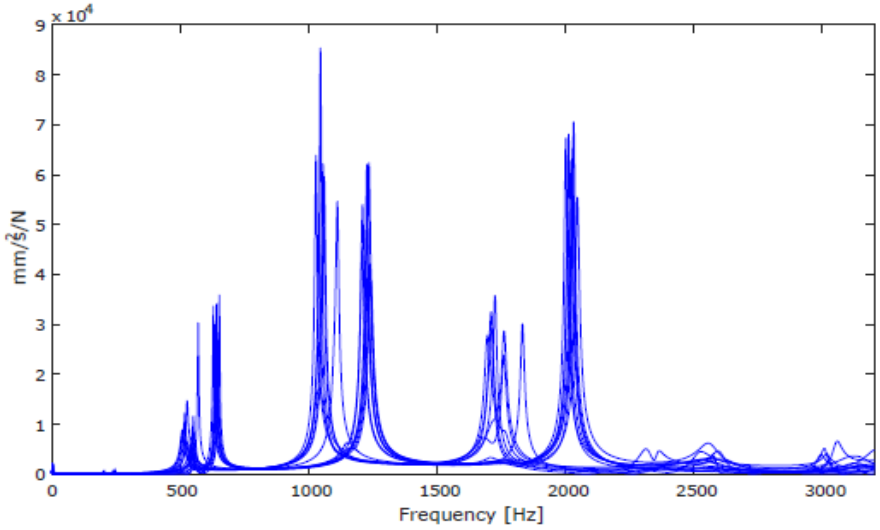
Source: (VÖLTZ, 2019).

Figure B.4 – Frequency response functions for all D1 samples in H_{11} position.



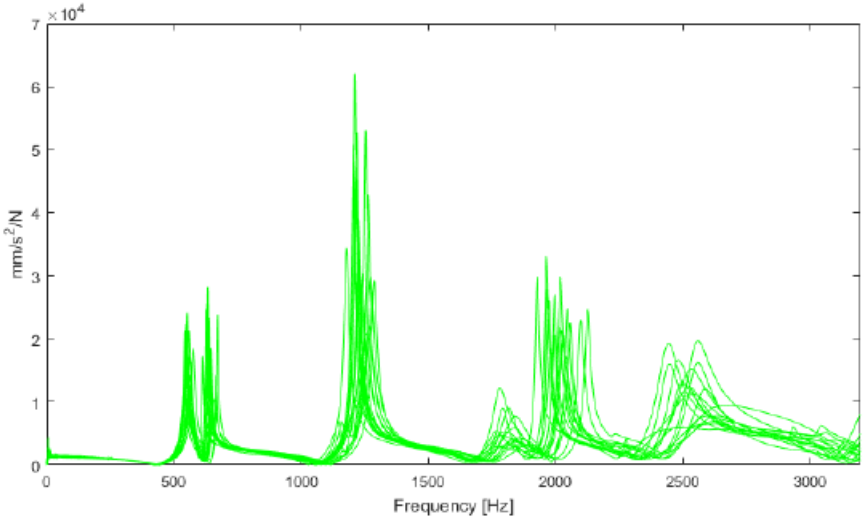
Source: (VÖLTZ, 2019).

Figure B.5 – Frequency response functions for all D1 samples in H_{21} position.



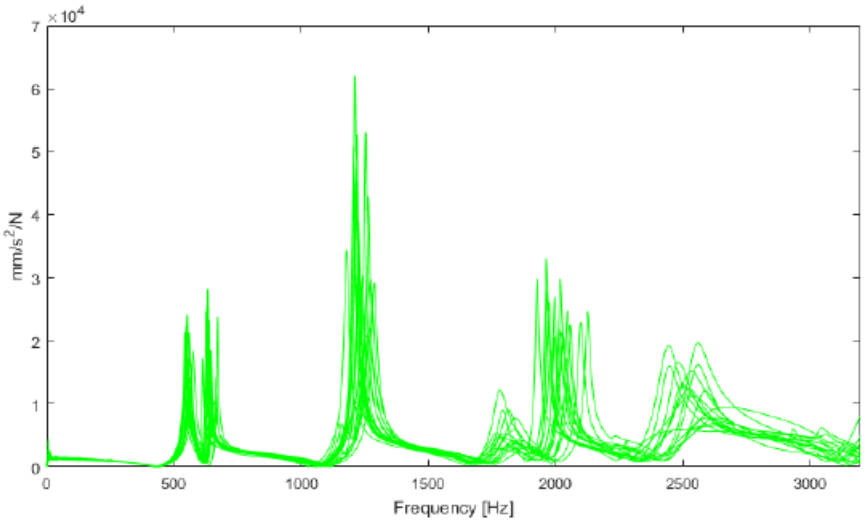
Source: (VÖLTZ, 2019).

Figure B.6 – Frequency response functions for all D2 samples in H_{11} position.



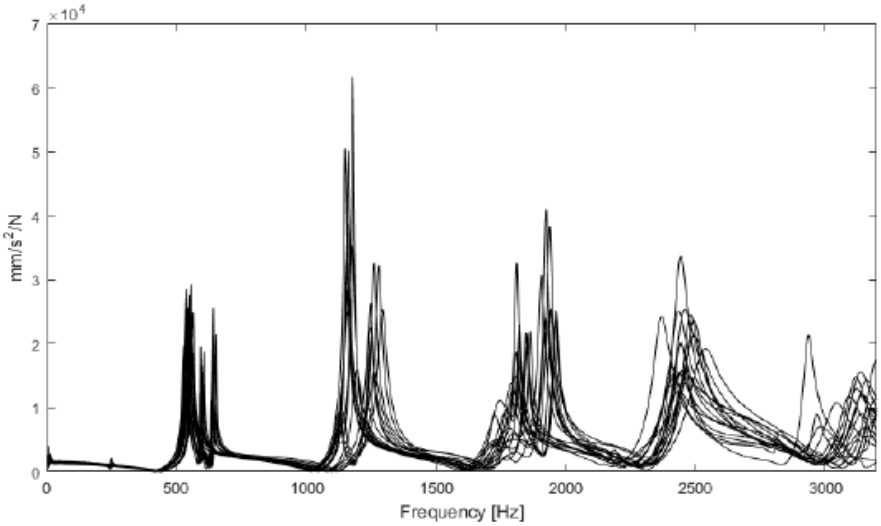
Source: (VÖLTZ, 2019).

Figure B.7 – Frequency response functions for all D2 samples in H_{21} position.



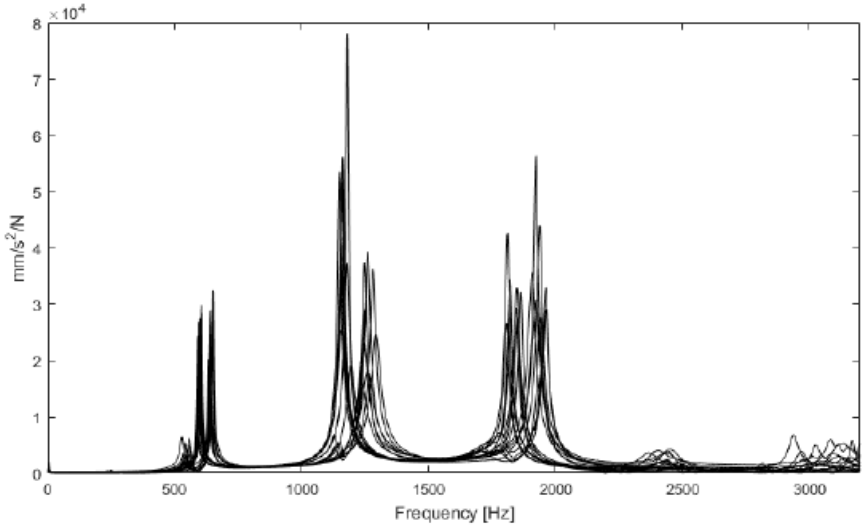
Source: (VÖLTZ, 2019).

Figure B.8 – Frequency response functions for all D3 samples in H_{11} position.



Source: (VÖLTZ, 2019).

Figure B.9 – Frequency response functions for all D3 samples in H_{21} position.



Source: (VÖLTZ, 2019).