UNIVERSIDADE DE SÃO PAULO
ESCOLA POLITÉCNICA

PATRICIA AKEMI SEKINI GUENKAWA

Detection and location of cavitation through artificial intelligence

Detecção e localização de cavitação via inteligência artificial

São Paulo
2024

PATRICIA AKEMI SEKINI GUENKAWA

Detection and location of cavitation through artificial intelligence

Detecção e localização de cavitação via inteligência artificial

Corrected Version

Dissertation presented to the Graduate Program in Electrical Engineering at the Escola Politécnica da Universidade de São Paulo in fulfillment of the requirements for the degree of Master of Science.

Concentration Area: Biomedical Engineering

Advisor: Prof. Dr. Sérgio Shiguemi Furuie

São Paulo
2024

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 11 de abril de 2024

Assinatura do autor: Patrícia Akemi Sekini Guenkawa

Documento assinado digitalmente

Assinatura do orientador: gov.br SERGIO SHIGUEMI FURUIE
Data: 11/04/2024 15:25:22-0300
Verifique em https://validar.iti.gov.br

Catalogação-na-publicação

Guenkawa, Patricia Akemi Sekini
Detecção e localização de cavitação via inteligência artificial / P. A. S. Guenkawa -- versão corr. -- São Paulo, 2024.
113 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Telecomunicações e Controle.

1.Sonotrombólise 2.Detecção de cavitação 3.Inteligência artificial 4.AlexNet I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Telecomunicações e Controle II.t.

PATRICIA AKEMI SEKINI GUENKAWA

Detection and location of cavitation through artificial intelligence

Detecção e localização de cavitação via inteligência artificial

Dissertation presented to the Graduate Program in Electrical Engineering at the Escola Politécnica da Universidade de São Paulo in fulfillment of the requirements for the degree of Master of Science.

Concentration Area: Biomedical Engineering

Date of approval: 02/19/2024

Dissertation Committee:

Prof. Dr. Sérgio Shiguemi Furuie
Escola Politécnica da Universidade de São Paulo
President
Approved

Prof. Dr. Mônica Mitiko Soares Matsumoto
Instituto Tecnológico de Aeronáutica
External Member
Approved

Prof. Dr. John Andrew Sims
Universidade Federal do ABC
External Member
Approved

To my mother and father, Isabel and Jorge, for their remarkable support, and motivation.

Acknowledgments

First of all, I would like to especially thank my parents who always supported me to pursue a higher education despite the limited opportunities they faced during life. I fully recognize that my achievements are much more the product of their endeavors and sacrifices during my upbringing than solely the result of my own efforts.

To Prof. Dr. Sérgio Shiguemi Furuie, my sincere gratitude for your invaluable support, guidance, and mentorship in both academia and life. Your influence extends beyond the professional realm; you are an inspiration not only in your expertise but also as a remarkable human being.

To my sister, in every chapter of my life, your presence has been a source of strength, love, and unwavering support.

To my colleagues, also friends, from the Biomedical Engineering Laboratory who helped me in completing this journey by collaborating with insightful thoughts and encouragement.

I extend my sincere gratitude to the professors of Poli-USP, with whom I had the privilege of being a student throughout my undergraduate and graduate studies. Their imparted knowledge and invaluable advice have been instrumental in shaping the development of this work.

I also want to acknowledge the financial support from the Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (CNPq) (130511/2021-8) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001.

Many of life's failures are people who did not realize how close they were to success when they gave up.

—THOMAS A. EDISON

ABSTRACT

Guenkawa, P. A. S. **Detection and location of cavitation through artificial intelligence.** 2024. Dissertação (Mestrado) - Programa de Pós-Graduação em Engenharia Elétrica, Escola Politécnica, Universidade de São Paulo, São Paulo, 2024.

Myocardial infarction is one of the main causes of morbidity and mortality worldwide. Among the possible treatments for blood flow obstruction is an emerging technique named sonothrombolysis. To reach satisfactory results, the event allied to the technique (cavitation of microbubbles) needs to be controlled to avoid harm to the patient. Given that, this study aimed to detect and classify the phenomenon during sonothrombolysis therapy through artificial intelligence, where the region of interest is the heart. The signals were generated using the k-Wave toolbox available for MATLAB, where features of the acoustic medium can be set, including non-linearities, attenuations, and the matrix array topology. After the simulation of those signals, an automatic and uncomplicated classifier method was proposed, based on the Continuous Wavelet Transform tool and Convolutional Neural Network (CNN) approach. The method made use of a pre-trained CNN, called AlexNet, operating a database of 2,800 synthetic waves for training (70%), testing (15%), and validation (15%). The evaluation included both the detection using broad and narrowband detectors, the noise level applied, and the database size. For the case of narrowband receivers, the results of the study indicated that the technique achieved state-of-the-art values of around 95.7%, and 96.0% for accuracy and precision, respectively. The considerable degree of accuracy demonstrated that using artificial intelligence could be an approach to explore the detection of cavitation for therapies using ultrasound signals.

Keywords: Sonothrombolysis. Cavitation detection. Artificial intelligence. AlexNet.

RESUMO

Guenkawa, P. A. S. **Detecção e localização de cavitação via inteligência artificial.** 2024. Dissertação (Mestrado) - Programa de Pós-Graduação em Engenharia Elétrica, Escola Politécnica, Universidade de São Paulo, São Paulo, 2024.

O infarto do miocárdio é uma das principais causas de morbidade e mortalidade em todo o mundo. Dentre os possíveis tratamentos para obstrução do fluxo sanguíneo, uma técnica emergente é denominada sonotrombólise. Para alcançar resultados satisfatórios, o evento aliado à técnica (cavitação de microbolhas) precisa ser controlado para evitar danos ao paciente. Diante disso, este estudo teve como objetivo detectar e classificar o fenômeno durante a terapia de sonotrombólise por meio de inteligência artificial, onde a região de interesse é o coração. Os sinais foram gerados utilizando a ferramenta k-Wave disponível para Matlab, onde podem ser definidas características do meio acústico, incluindo não linearidades, atenuações e a topologia da matriz de transdutres. Após a simulação desses sinais, foi proposto um método classificador automático e descomplicado, baseado na ferramenta Transformada Wavelet Contínua e abordagem de Rede Neural Convolucional (CNN). O método utilizou uma CNN pré-treinada, chamada AlexNet, operando uma base de dados de 2.800 sinais para treinamento (70%), teste (15%) e validação (15%). As métricas de avaliação incluíram tanto a detecção em banda larga e estreita, o nível de ruído aplicado e o tamanho da base de dados. Para o caso dos receptores de banda estreita, os resultados do estudo indicaram que a técnica alcançou valores em torno de 95,7% e 96,0% para acurácia e precisão, respectivamente. O considerável grau de acurácia demonstrou que o uso de inteligência artificial pode ser uma abordagem para explorar a detecção de cavitação para terapias que fazem uso de ultrassom.

Palavras-chave: Sonotrombólise. Detecção de cavitação. Inteligência artificial. AlexNet.

LIST OF FIGURES

2.1 (a) Initiation, progression, and complication of human coronary atherosclerotic plaque. (b) Thrombus formation 24

2.2 Schematic image of acoustic (a) inertial and (b) stable cavitation process . 29

2.3 Sequence of Bjerknes Forces acting on UEAs. 32

2.4 (a) PCD setup for measuring ultrasound-induced MB cavitation. (b) Illustration of the focused ultrasound system, and image acquisition and registration by PCI and positron emission tomography/computed tomography (PET/CT) 33

2.5 Schematic of ACD, SSCD, and PCD approaches for characterizing cavitation 34

2.6 AI and its subsets Machine Learning (ML) and Deep Learning (DL) 35

2.7 Techniques of ML 36

3.1 A general CNN architecture 40

3.2 Example of the convolutional computation, where $1^{st}, 2^{nd}, 3^{rd}, 5^{th}$, and 16^{th} steps are represented 42

3.3 (a) 2D convolution. (b) 3D convolution 43

3.4 Example of Max Pooling operation with a kernel of 2×2 44

3.5 Example of Average Pooling operation with a kernel of 2×2 44

3.6 Fully connected feedforward network with one hidden layer and one output layer . . . 45

3.7 ReLU Function for $x \in [-2, +2]$ . . . 46

3.8 AlexNet architecture: ImageNet 2012 winning CNN model . . . 47

3.9 General idea of Transfer Learning . . . 48

4.1 Mapping the wavelet scalogram . . . 50

4.2 Some families of CW . . . 51

4.3 Example of a general step of wavelet transform . . . 52

5.1 Overview of the Sonothrombolysis project in the LEB . . . 54

5.2 Standard heart dimensions and distance between patient skin and heart. . . 55

5.3 Illustration of the project's piezoelectric transducer matrix array (two-dimensional) . . . 56

5.4 Normals at transducers' centers. . . . 57

5.5 (a) Matrix array (two-dimensional). (b) Selected focus. (c) Activation matrix 57

5.6 Waveforms delayed, consequently, they sum constructively . . . 59

5.7 (a) Matrix array (two-dimensional). (b) Example of activation matrix. (c) Active subgroup focusing on the target (illustrated by the yellow point). (d) Example of 25 signals from a 5×5 subgroup with their respective delays 60

5.8 Excitations waveforms in elements 55, 46, and 37 from Figure 5.7, showing the comparative delay between them . . . 62

5.9 Schematic representation of acoustic pressure spectrums: frequency components . . . 64

5.10 Flowchart with the general steps of the process . . . 67

5.11 Reconstructed signals after delay and sum (D&S) 68

5.12 Signals (A1), (B1), (C1), (D1), and respective scalograms for each signal . 70

6.1 Confusion Matrix for narrowband receiver case - RN rate (840 waves for testing and validation, 210 for each target class - 30% of the database) . . 77

6.2 Confusion Matrix for broadband receiver case - RN rate (120 waves for testing and validation, 30 for each target class - 30% of the database). . . 78

A.1 Similarity between biological and artificial neural networks. (a) Representation of a biological neuron cell. (b) an artificial neuron model 96

A.2 Artificial neuron and the structure of the feed-forward artificial neural network 97

A.3 Layers representation: input level (red), hidden level (blue), and output level (green) 98

A.4 General flow graph of a neural network with 3 layers 99

A.5 General flow graph of a neural network with $(l+1)$ layers 100

A.6 Sigmoid function for $x \in [-4, +4]$ 102

A.7 Hyperbolic Tangent Function for $x \in [-2, +2]$ 103

A.8 ReLU Function for $x \in [-4, +4]$ 104

A.9 Softmax function for $x \in [-10, +10]$ 105

A.10 Gradient Descent example 110

A.11 Red arrow illustrates the direction of the negative gradient descent at iteration $t = 1$ 112

LIST OF TABLES

5.1 Features of the initial chosen transducer, as informed by the manufacturer 56

5.2 Simulation features 58

5.3 Types of signals received by the receptor elements and frequency aspects . 66

5.4 Dataset Summary 71

5.5 Dataset Distribution 72

5.6 List of hyperparameter values for the network model employed 72

6.1 Performance for the different sources, comparing narrow and broadband receivers, and distinct levels of noise (2%, 5%, and RN) 76

ACRONYMS

ACD	Active Cavitation Detection
ARF	Acoustic Radiation Force
AMI	Acute Myocardial Infarction
APC	American Piezo Ceramics
AI	Artificial Intelligence
BW	Bandwidth
CCECE	Canadian Conference on Electrical and Computer Engineering
CVD	Cardiovascular Disease
CT	Computed Tomography
CINC	Computing in Cardiology
CBEB	Congresso Brasileiro de Engenharia Biomédica
CWT	Continuous Wavelet Transform
CNN	Convolutional Neural Network
D&S	Delay-and-Sum
DL	Deep Learning
DUS	Diagnostic Ultrasound
ECG	Electrocardiogram
FC	Fully Connected
HIFU	High-intensity Focused Ultrasound
IC	Inertial Cavitation
IEC	International Electrotechnical Commission
ML	Machine Learning
MIx	Mechanical Index
MB	Microbubble
MI	Myocardial infarction
NSTEMI	Non-ST-elevation Myocardial Infarction
PCD	Passive Cavitation Detection
PCI	Percutaneous Coronary Intervention

PET	Positron Emission Tomography
RN	Random Noise
RP	Rayleigh-Plesset
ROI	Region of Interest
ReLU	Rectified Linear Unit
SC	Stable Cavitation
SGD	Stochastic Gradient Descent
SGDM	Stochastic Gradient Descent with Momentum
SSCD	Self-Sensing Cavitation Detection
STFT	Short-time Fourier Transform
STEMI	ST-elevation myocardial infarction
TL	Transfer Learning
UEA	Ultrasound Enhancing Agent
VGG	Visual Geometry Group
WHO	World Health Organization
WT	Wavelet Transform

SUMMARY

1 INTRODUCTION **19**

1.1 Motivation 19

1.2 Objective and Scope 21

1.3 Contributions 21

1.4 Organization of this work 21

2 LITERATURE REVIEW **23**

2.1 Medical Background 23

2.1.1 Thrombus Formation 23

2.1.2 Myocardial Infarction 25

2.2 Frontline Therapies 25

2.3 Sonothrombolysis 26

2.3.1 Exogenous and endogenous nuclei 27

2.3.2 Cavitational Interaction 28

2.3.2.1 Bubble Activity 30

2.3.3 Acoustic Radiation Force 31

2.4 Mechanisms of Detection 33

2.5 Artificial Intelligence and Cavitation Prediction 35

3 CONVOLUTIONAL NEURAL NETWORK 39

3.1 Introduction 39

3.2 Convolution Layer 40

3.3 Pooling Layer 43

3.3.1 Maximum Pooling 43

3.3.2 Average Pooling 44

3.4 Fully Connected Layer 44

3.5 Activation function 45

3.6 Image Classification Model 46

3.6.1 AlexNet 47

3.6.2 Transfer Learning 48

4 WAVELET TRANSFORM 49

5 METHODOLOGY 53

5.1 Biomedical Engineering Laboratory (LEB) 53

5.2 Physical Experiments 54

5.3 Numerical Experiments 56

5.4 Frequency and Cavitation classification 63

5.5 Detection: receivers' features 66

5.6 General Classification Process 66

5.6.1 Signals' generation 67

5.6.2 CWT application and RGB images 68

5.6.3 Dataset 70

5.6.4 Model Implementation 71

6 ANALYSES AND RESULTS **75**

6.1 Metrics 75

6.2 Results Analysis 76

7 DISCUSSION **79**

8 CONCLUSION **81**

REFERENCES **82**

A DEEP LEARNING **95**

1 INTRODUCTION

In this chapter, we present the motivation for this work. Also, we briefly state some choices made, the objective, and contributions. Finally, we present the organization of this project.

1.1 MOTIVATION

Cardiovascular disease (CVD) is a generic label for conditions affecting the heart or blood vessels. Coronary heart disease, cerebrovascular disease, rheumatic heart disease, deep vein thrombosis, and pulmonary embolism are part of the group of disorders encompassed by the condition. CVDs are the leading cause of death globally, taking an estimated 17.9 million lives each year. Of these deaths, 85% were due to myocardial infarction (heart attack) and stroke (WHO 2021).

Approximately 1.5 million cases of myocardial infarction (MI) occur annually in the United States; the yearly incidence rate is close to 600 cases per 100,000 people (Zafari and Abdou 2019), and MI is a major cause of morbidity and mortality worldwide. It is estimated that more than 3 million people each year have an acute ST-elevation myocardial infarc-

tion (STEMI) and more than 4 million have a non-ST-elevation myocardial infarction (NSTEMI) (White and Chew 2008).

The commonest cause of MI is the partial or complete epicardial coronary artery occlusion from erosion or rupture of vulnerable plaques. Those arterial plaques are composed of materials that enter the artery wall from the bloodstream, such as fat, cholesterol, calcium, waste products from cells, or a clotting agent called fibrin. The obstructions are responsible for around 70% of fatal events and the epicardial and microvascular renovation flow in acute STEMI can be done by surgical or non-surgical treatment, where catheter-based reperfusion and fibrinolysis are examples of the methods respectively.

The obstruction of a coronary artery or any of its large branches has long been regarded as a serious accident and the treatment approach has evolved over the years. MI was considered to be a "wound" of the heart in the early $20th$ century. In 1912, (Herrick 1912) described the importance of absolute rest in bed for several days as a resource for recovery, later authors recommended morphine for pain. Currently, myocardial reperfusion has been improved progressively: the addition of aspirin as an agent to the fibrinolytic drug (Sabatine et al. 2005); the exploration of tissue plasminogen activators to the thrombi lysing (Cannon et al. 1994); the addition of stents (Zhu et al. 2001), and more recently the use of diagnostic ultrasound (DUS) inducing cavitation as a method of dissolving intravascular and microvascular thrombi in acute myocardial infarction (AMI) (Xie et al. 2013).

Ultrasonic energy has been assessed extensively as a method to promote thrombolysis (Trübestein et al. 1977; Tachibana 1992), and when the amount of energy is sufficient along with the use of microbubbles (MBs), the activity can restore the blood flow of vessels, and such procedure is called sonothrombolysis.

Sonothrombolysis is a novel therapy that has noninvasive and non-ionizing features (Medel et al. 2009). By the application of acoustic emissions directed to the thrombus location, the intention is to mechanically break up the occlusive blood clot through the cavitation of MBs. A very important caution to be avoided is the trigger of bubble collapse in undesired spots, which could result in damage to healthy tissue. Therefore, the detection of the type of cavitation that is occurring and its location is crucial for methods that combine acoustic waves and MBs, and consistent sonification is necessary for the method's safety and efficacy.

1.2 OBJECTIVE AND SCOPE

The primary purpose of this project was to develop a feedback mechanism for cavitation detection based on an Artificial Intelligence algorithm. The guidance allows the classification of cavitation phenomena through temporal signals, with the database being obtained by simulating the use of ultrasonic sensors, and the piezoelectric crystals belonging to the equipment intended to cause sonothrombolysis. The hypothesis is that the detection is feasible by the development of a Deep Learning model.

1.3 CONTRIBUTIONS

We consider that this work affirmed the hypothesis that Artificial Intelligence could detect the cavitation phenomenon in Sonothrombolysis therapy. As contributions regarding this question, we can cite:

- The development of a classification method based on a feed-forward artificial neural network.
- We evaluated the Continuous Wavelet Transform as a tool to extract features of the analyzed signals.
- We verified the possibility of applying transfer learning to a medical problem.
- We evaluated our method's performance by the inspection of accuracy, recall, precision, and F1-score metrics.

1.4 ORGANIZATION OF THIS WORK

This work is organized as follows:

- Chapter 2 reviews the medical background for the condition that the therapy aims to treat, provides details for the Sonothrombolysis method, and presents an introduction to the detection technique we develop in this work;
- Chapter 3 will cover the theory behind Convolutional Neural Networks used in this work;
- Chapter 4 will cover the most relevant topics behind the Continuous Wavelet Transform tool;

- Chapter 5 will cover the methodology and simulations regarding the selected approach;
- Chapter 6 will cover the analysis and results obtained through the method applied, verifying its effectiveness; and
- Finally, Chapter 7 will share some conclusions and suggestions for future work.

2

LITERATURE REVIEW

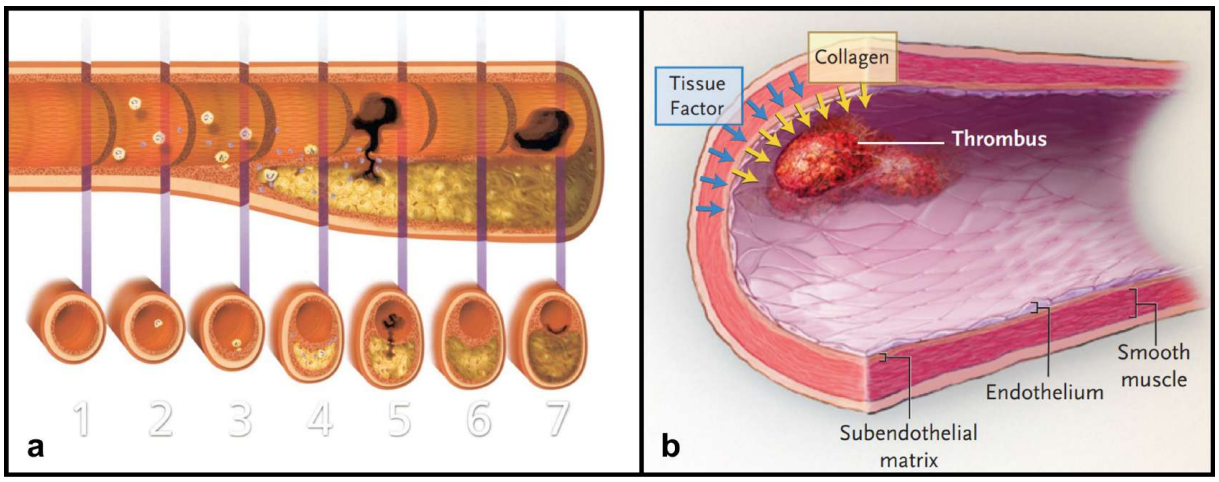
In this chapter, we highlight important concepts regarding the medical background of MI, then we detail the cavitation phenomenon, and finally, we briefly state some detection mechanisms for cavitation.

2.1 MEDICAL BACKGROUND

2.1.1 THROMBUS FORMATION

Atherosclerosis is the main contributor to CVD (Libby et al. 2019). The condition denomination derives from the Greek word for "gruel" or "porridge", which is related to the appearance of the typical atherosclerotic plaque (or atheroma). Those arterial plaques are composed of materials that enter the artery wall from the bloodstream, such as fat, cholesterol, calcium, waste products from cells, or a clotting agent called fibrin. The lipid material can become more fibrous with time, hampering the blood flow and leading to tissue ischemia.

Figure 2.1 – (a) Initiation, progression, and complication of human coronary atherosclerotic plaque. (b) Thrombus formation



Source: Libby 2001, Bruce Furie and Barbara Furie 2008.

Figure 2.1a (Libby 2001) illustrates a timeline of human atherogenesis from a normal artery (1) to more severe stages of atheromas (5, 6, 7). Stages (2) and (3) show the evolution to the fibrofatty phase. There is the onset of the lesion due to risk factors such as hyperlipoproteinemia, triggering inflammatory leukocytes to act on the spot. (4) characterizes the procoagulant feature. If the fibrous cap ruptures at the point of weakening, there is thrombosis on nonocclusive atherosclerotic plaque (5). When the thrombus is reabsorbed, it can lead to a healing process (6) and the fibrofatty lesion can become a calcified plaque. (7) describes the emergence of occlusive thrombi by superficial erosion of the endothelial layer.

Figure 2.1b (Bruce Furie and Barbara Furie 2008) gives more details about the thrombus formation. As described, it is a dynamic process in which the lesions formed in the endothelial cells cause changes in the vascular wall. Thus, as a defense and regeneration mechanism, there is platelet aggregation at the injured site, possible thrombus formation as a response to prevent bleeding from injured blood vessels (Gregg and Goldschmidt-clermont 2003), and the healing response leads to increased collagen accumulation. Decreased blood flow due to the presence of an atherosclerotic plaque can also occasion thrombus formation in some cases in which shear, flow, and turbulence greatly influence the architecture of the clot.

In the circumstance of the blood clot (thrombus) breaking loose, it is carried by the bloodstream and it might block another blood vessel in the distal vasculature. The condition is called thromboembolism and the occlusions cause hypoxia in distal tissues, leading to

infarction and tissue death (Bader et al. 2016). The slowly progressing obstruction can affect several arterial beds, and depending on the place of the tissue death, the blockages are divided into two main categories: venous and arterial. Some examples of venous thromboembolism are deep vein thrombosis and pulmonary embolism (Disease Control and Prevention 2020). Arterial thromboembolism often occurs in the brain (ischemic stroke) and heart (coronary thrombosis), which could lead to heart tissue damage or an MI (Lyaker et al. 2013).

As mentioned earlier, this study is part of a bigger project that focuses on heart disease treatments. For this reason, it is interesting to discuss MI more deeply.

2.1.2 MYOCARDIAL INFARCTION

MI is defined in pathology as myocardial cell death due to prolonged ischemia (Thygesen et al. 2012). The clinical evaluation involves an electrocardiogram (ECG), biochemical testing, and invasive and noninvasive imaging. As shown in Figure 2.1, the presence of the thrombus affects the local circulation, which can cause total occlusion and lead to STEMI, NSTEMI, or unstable angina (Anderson and David Morrow 2017).

MI was considered to be a "wound" of the heart in the early $20th$ century and it was believed that coronary thrombosis was always immediately fatal (Morrow and Braunwald 2016). In 1947, Wood and colleagues started to describe the condition as "intermediate coronary syndrome" and in 1961, the study (Wood 1961) based on the analysis of 150 cases over 10 years, showed that most deaths were due to the development of cardiac infarction related to coronary thrombosis. By that time, the true incidence of coronary insufficiency was difficult to determine.

Since then, our understanding of its pathogenesis has changed with the advance of new studies and data, leading to a vast improvement in less invasive, efficient, and innovative treatment options. Some of the therapies are reviewed in the next section.

2.2 FRONTLINE THERAPIES

When a patient is diagnosed with the pathology, work is required on the recanalization of the occluded coronary artery as soon as possible, which can be achieved mainly by one of the two widely accepted reperfusion strategies: the use of thrombolytic agents and percutaneous coronary intervention (PCI) (Bagai et al. 2014).

PCI, also known as coronary angioplasty, is a non-surgical procedure that requires cardiac catheterization. A catheter tube is inserted to open the coronary arteries that are narrowed or blocked by an atherosclerotic plaque. It is mostly a safe procedure, and serious complications are not common, but they include bleeding, blood vessel damage, arrhythmias, and damaged arteries. Furthermore, the patient needs to be cautious in the following months, since restenosis might occur, causing the artery to become narrow or blocked again (NHLBI 2021).

Fibrinolytic drugs have the role of stimulating the dissolution of blood clots. Heparin, aspirin, dipyridamole, or the combined application of thrombolytics can be added to therapy to help prevent the recurrence of occlusive thrombus, but it does not show the best results in the management of chronic occlusions. Even though clinical issues do not occur often, the risks of intravenous thrombolytic therapy are harmful and implicate systemic hemorrhage, immunologic complications, hypotension, and myocardial rupture (Califf et al. 1992).

A meta-analysis of 22 randomized trials of thrombolysis in 50,246 patients revealed that the group identified with early treatment within the first hour of symptom onset had the greatest reduction in mortality (Boersma et al. 1996), which demonstrates that these interventions are most effective within a limited timeframe. Even though those techniques show mostly safe outcomes, a major issue with current therapies is the phenomenon of "no-reflow" (Saint Victor et al. 2014). This happens when the recanalization of the coronary artery occurs improperly, consequently, blood flow to the ischemic tissue may still be blocked (Rezkalla and Kloner 2002). Some other treatment methods in current studies can be cited, such as embryonic stem cells, induced pluripotent stem cells, and bone marrow cells. These new strategies focus on the regeneration of cardiomyocyte discovery (Peng and X. Wu 2017). Among the mechanisms of unclogging, the main goal is to improve the safety and efficacy of the treatments.

Since this project depends on an emerging treatment technique with the use of therapeutic ultrasound, the method is detailed in the next section.

2.3 SONOTHROMBOLYSIS

Sonothrombolysis is described as ultrasound-assisted clot lysis conducted by the administration of acoustically active MBs and fibrinolysis, and potentially being a clinical emerging application for vessel occlusion therapy. The method takes advantage of the mechanical bioeffects of ultrasound, supporting the diffusion of thrombolytic drugs to mechanically

break up the thrombus through the cavitation of MBs.

The use of ultrasound technology is powerful and the optimization of its parameters is fundamental to achieving a target effect. This ability was explored in this study by simulating different choices of frequency, amplitude, and mode of vibration. The various mechanisms of thrombolytic enhancement can be combined to meet a specific clinical requirement, inducing stable cavitation, inertial cavitation, micro-streaming, and acoustic radiation force, all aiming to reestablish the blood flow. Among those tools, we first discuss the use of MBs.

2.3.1 EXOGENOUS AND ENDOGENOUS NUCLEI

MBs are broadly used in diagnostic and therapeutic medical applications, for instance, acting as echo-enhancers and improving the quality of DUS images as contrast agents, since blood is a poor ultrasound scatterer. For having a similar size as red blood cells, the MBs can freely circulate within blood vessels and achieve the intended location (Shpak et al. 2016).

Furthermore, there are two types of MBs, and the threshold for cavitation activity depends highly on the available nuclei. The endogenous bubbles are present in our body, that is, gas dissolved in our blood, and they are characterized as nano-scale "gas pockets" very stabilized with very high surface tension, around 30 MPa (Xu 2020). They can also be intentionally produced in the body for clinical applications such as lithotripsy (treatment for kidney stones), and another example is a well-known procedure named histotripsy, which uses ultrasound to mechanically liquefy the target tissue (Maxwell et al. 2009).

On the other hand, the insertion of exogenous nuclei allows a wider range of applications, such as drug delivery (T.-Y. Wang et al. 2013) and gene therapy (J. Wu and R.-K. Li 2016). MBs are commercially available as ultrasound enhancing agents (UEAs), frequently used as ultrasound contrast agents (H. Lee et al. 2017), as we already cited. The main benefit of the MBs administered intravenously is the control of the type and location of cavitation activity, and the use of UEAs greatly enhances the biological effects of sonothrombolysis since a higher cavitation phenomenon is allowed to occur (Shen et al. 2009). The presence of either endogenous or exogenous MBs can have multiple biological and thermal effects.

The injected MBs are composed of a shell of a biocompatible material such as a protein, lipid, or polymer, and they may be filled by air, perfluorocarbon, or sulfur hexafluoride inert gas (Quaia 2005). The commercially available MBs normally range in size from 1

to 4.5 µm (Brown et al. 2011), and that is because they need to be capable of passing through the circulation after intravenous injection, i.e., a diameter smaller than 8–10 µm (Quaia 2005).

This study, which is part of the sonothrombolysis project, makes use of exogenous MBs. The cavitation phenomenon, which involves the interaction with the MBs, is detailed in the next section.

2.3.2 CAVITATIONAL INTERACTION

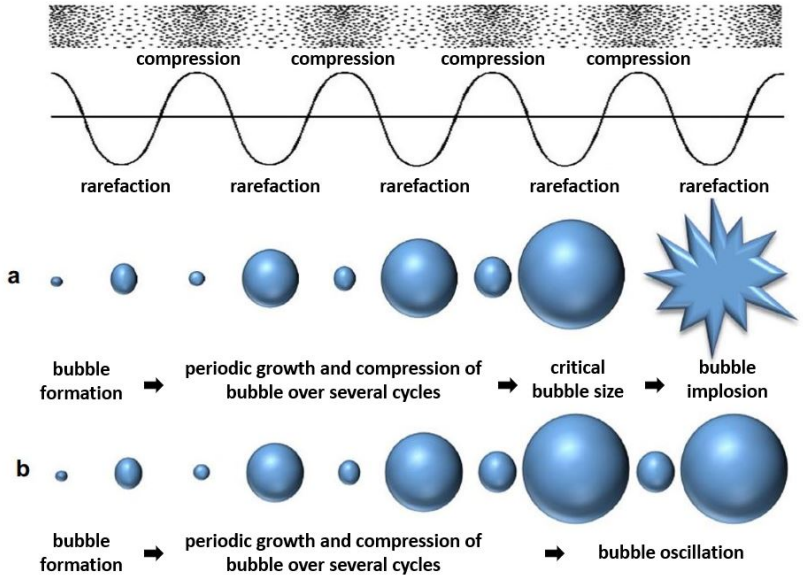
Acoustic waves are three-dimensional fluctuations in the pressure field. Ultrasound waves are defined as a set of acoustic waves, with frequencies higher than the limit of audible by humans, that is, higher than 20 kHz.

The interaction of acoustic waves and tissue can generally be categorized as thermal and mechanical effects. The absorption of ultrasound energy causes the elevation of temperature of the tissue and surroundings, and in some methods, that is exactly the application purpose. For instance, we have the High-intensity focused ultrasound (HIFU), which a common clinical application is the thermal ablation with consequent tissue necrosis (Dubinsky et al. 2008). When absorption and scattering mechanisms are directed towards the target, the resultant force is acoustic radiation (Bader et al. 2016) and it describes the primary mechanical effect.

The secondary mechanical effect is acoustic cavitation, which refers to both the formation and oscillation of bubbles due to acoustic pressure. The precise mechanism of sonothrombolysis is still not fully understood, but the previous hypotheses indicate the ability of MBs to potentiate ultrasound-induced thrombolysis, leading to destabilization and subsequent thrombus dissolution (Porter and Xie 2001; Schleicher et al. 2016).

Cavitation activity is mainly categorized into two types: stable and inertial cavitation. This physical phenomenon is induced when an ultrasound wave is applied to a liquid, so the medium goes through alternating cycles of expansion (rarefaction) and compression, as we can see in Figure 2.2.

Figure 2.2 – Schematic image of acoustic (a) inertial and (b) stable cavitation process



Source: Izadifar et al. 2018.

When sonicated, MBs are subjected to the interchange act of expansion and compression, which pulls molecules apart and pushes them together, respectively. The amplitude is maximal when the MB is sonificated at the resonant frequency, and that depends on the size of the bubble (resonant size decreases with frequency) (Ellens and Hynynen 2015). As we can see in Figures 2.2a and 2.2b, the growth process of the MBs is the same in both cases until it hits a critical bubble size. Before that moment, at moderate ultrasound intensities, the oscillation is stable over many cycles, where diffusion is proportional to the surface of the bubble, and the maximum size rarely exceeds twice the equilibrium radius (Dalecki 2004). Such interaction is known as stable cavitation, where the bubble motion stimulates fluid mixing through microstreaming, and the process is illustrated in Figure 2.2b.

At sufficiently high acoustic intensities, the bubble implosion happens because the oscillations become highly nonlinear, leading to a violent collapse (Figure 2.2a). As a result, the surrounding tissue can receive microjets, shock waves, and a punctual increase of temperature of thousands of degrees Kelvin (Dalecki 2004), characterizing inertial cavitation. Many methods explore those bioeffects from inertial cavitation, such as Histotripsy (Maxwell et al. 2009) or HIFU (Sun-young et al. 2013; Crouzet et al. 2013), but it is important to notice the deleterious effects as well, that is, the thermal coagulation for example, which is protein denaturation and irreversible cell damage on the tissue (Jolesz 2009). Another possible problem is the trigger of inertial cavitation in undesired locations,

resulting in damage to healthy tissue. So, the detection of the type of cavitation that is occurring and the location, become crucial for methods that combine acoustic waves and MBs, and consistent sonification is necessary for safe and efficient sonothrombolysis, which leads again to the importance of the study.

But before we approach the cavitation detection techniques, it is pertinent to understand the basics of bubble dynamics since it is part of the cavitation phenomenon.

2.3.2.1 BUBBLE ACTIVITY

A bubble in liquids is a spherical volume of gas. As we discussed before, the oscillations in size can be related to external forces, and cavitation can be categorized into stable and inertial. The well-known equation that describes the bubble radial oscillations in terms of hydrodynamics is the Rayleigh-Plesset (RP) equation

$$\ddot{R}R + \frac{3}{2}\dot{R}^2 = \frac{\Delta P}{\rho} \tag{2.1}$$

where R, $\dot{R}$, and $\ddot{R}$ are the radius, the velocity, and the acceleration of the bubble wall, respectively. ρ is the liquid mass density, and $\Delta P = P_L(R) - P_\infty$ is the pressure difference between the liquid at the bubble wall and the infinitely far external pressure. Equation (2.1) was originally derived by (Rayleigh 1917) and refined by (Plesset 1949), it assumes spherical symmetry of the bubble, and the liquid around it is considered incompressible.

A generalized RP equation (2.2) was obtained some years later with the contribution of colleagues. The surface tension constant of the bubble, σ, was introduced by (Noltingk and Neppiras 1950), the coefficient of the viscosity of the liquid, μ, was added by (Poritsky and Horvay 1952), R_0 is the equilibrium radius, and γ is the polytropic constant. The derivation of this equation and the main assumptions adopted can be found in bubble dynamics textbooks (e.g. Brennen 1995).

$$\rho\left(\ddot{R}R + \frac{3}{2}\dot{R}^2\right) = \left(P_0 + \frac{2\sigma}{R_0}\right)\left(\frac{R_0}{R}\right)^{3\gamma} - P_0 - \frac{2\sigma}{R} - 4\mu\frac{\dot{R}}{R} - P(t) \tag{2.2}$$

P_∞ can be considered the sum of the acoustic forcing $P(t)$ and the ambient pressure P_0 (equation (2.3)), and since the bubble is assumed to be significantly smaller than the acoustic wavelength, the acoustic pressure is treated as uniform.

$$P_\infty = P(t) + P_0 \tag{2.3}$$

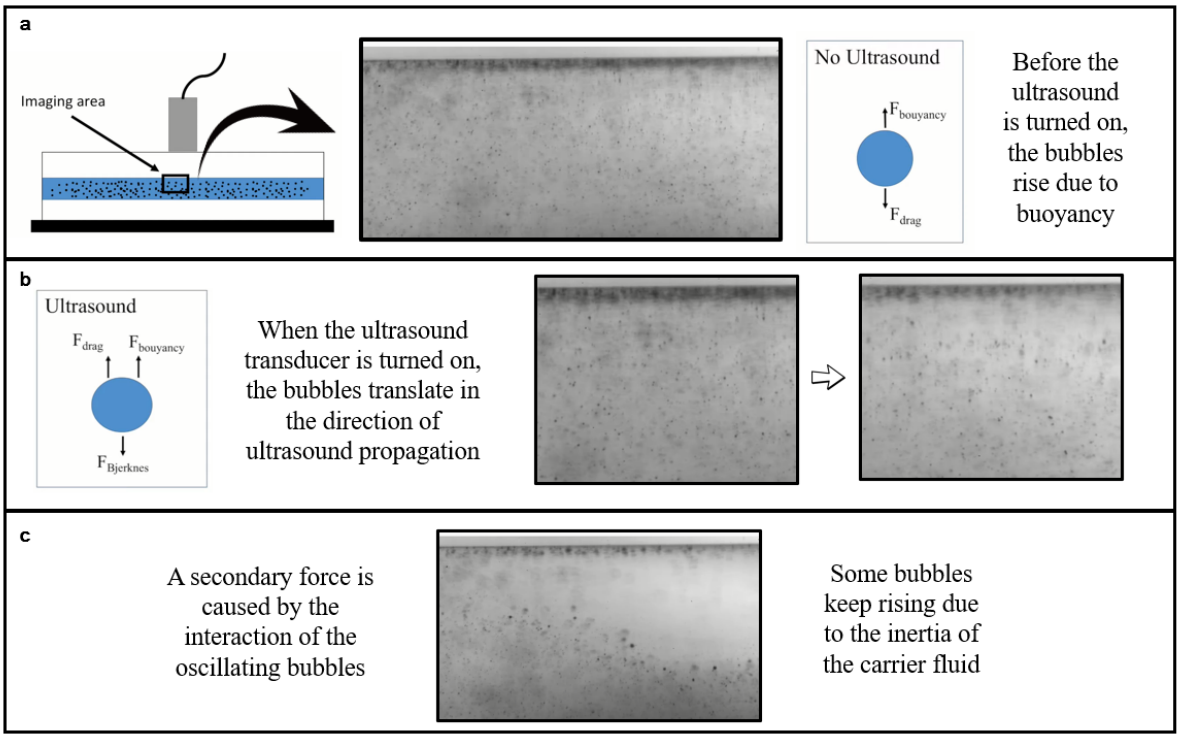
Experiments (Putterman et al. 2001) verified that the RP equation describes the dynamics of bubbles well from a macroscopic point of view, however, it is still unclear if the equation is applicable to the dynamic of nanobubbles or MBs. The main reason for that is the fragility of the bubbles of a small size, complicating the experimental investigation. Besides that, the RP equation only models the bubble's growth, disregarding the moment of collapse present in the inertial cavitation. Further, the limitation also exists because the analysis is restricted to an isolated bubble.

2.3.3 ACOUSTIC RADIATION FORCE

We discussed briefly in the previous section the equations (2.1 - 2.3) that describe the bubble radial oscillations. Now we analyze the effect of the ultrasound waves in those bubbles, and their behavior when an acoustic radiation force (ARF) is applied.

To understand the concept, (Clark and Aliseda 2016) recorded the fluid dynamics of UEAs, exhibiting their response under the influence of an acoustic field. The experimental setup consists of a cylindrical phantom, composed of water and glycerin and encircled by a silicon material, along with an ultrasound transducer (the device is off initially). UEAs (1 – 10 µm) are injected into the fluid mixture, and they rise due to the buoyancy force (Figure 2.3a).

Figure 2.3 – Sequence of Bjerknes Forces acting on UEAs: (a) Transducer is turned off and bubbles rising due to buoyancy. (b) The transducer is turned on and the Primary Bjerknes force is acting. (c) Interaction of the oscillating bubbles provokes a secondary Bjerknes force



Source: Figure assembled from Gallery of Fluid Motion video (Clark and Aliseda 2016).

When the ultrasound transducer is turned on, it causes the translation of the MBs in the direction of the ultrasound propagation, and that happens as a consequence of the primary Bjerknes force (Figure 2.3b). Furthermore, the interaction of the oscillating bubbles causes a secondary force (secondary Bjerknes force), so the bubbles clump together (Figure 2.3c). The forces were first described by Vilhelm Bjerknes in 1906 (Bjerknes 1906), and the contact of two bubbles is described by

$$F_{12} = -V_2 \Delta p_1 \tag{2.4}$$

where bubble 2, with volume V_2, experiences a force F_{12} as a result of the pressure emitted by bubble 1 (Δp_1).

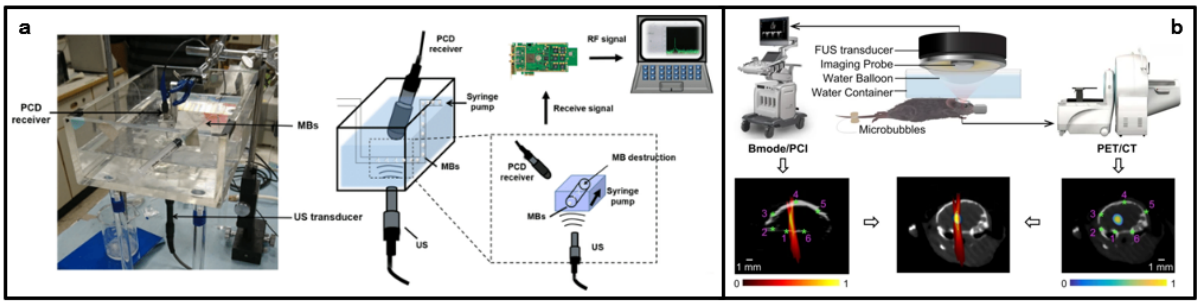
However, both the bubble oscillation equation and the forces between two MBs, do not describe the vaporization phenomenon well, since they model a limited interaction between the bubbles (one and two bubbles, respectively). The cavitation modeling is a challenge mostly due to the difficult numerical implementation and complexity of the process. That said, a preferable approach might be to simulate transmit and receive signals, rather than modeling the behavior of MBs. This is the approach used in our work.

2.4 MECHANISMS OF DETECTION

The optimization of sonothrombolysis treatments can be achieved by three golden rules (Apfel 1981): identify the sound field, know the liquid, and detect the cavitation activity. The first aspect will be seen in the Methodology section; the description of the liquid involves the cavitation nuclei we are dealing with, which we addressed in Section 2.3.1; and in this section, we will explore the most common mechanisms of monitoring cavitation activity employed nowadays for ultrasound applications.

So, to investigate Apfel's final rule, we start examining the passive cavitation detection (PCD) technique. As we can verify in Figure 2.4a, cavitation acoustic emissions are passively received by a transducer, and the signal is inspected by verifying its frequency spectrum, detecting whether or not cavitation is occurring and the type of cavitation detected, if any.

Figure 2.4 – (a) PCD setup for measuring ultrasound-induced MB cavitation. (b) Illustration of the focused ultrasound system, and image acquisition and registration by PCI and positron emission tomography/computed tomography (PET/CT)



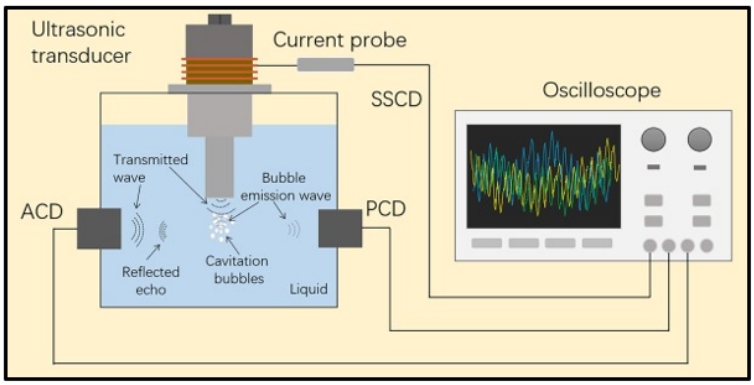
Source: Liao et al. 2020, Y. Yang et al. 2019.

A second method, and a more recent one, is passive cavitation imaging (PCI). The limited information received by PCD is overcome by the implementation of a monitoring system, which enables verifying the location and type of cavitation activity more easily. Figure

2.4b brings an example of the setup, and even with the real-time follow-up imaging, it is notable that the treatment process is not portable or practical when we envision using it outside of a hospital setting.

A third acoustic approach to characterize the phenomenon is active cavitation detection (ACD). Unlike PCD and PCI, in the ACD technique, a transducer sends an ultrasonic wave to the cavitation field, and the ultrasonic reflection from the cavitation bubbles is captured by the same transducer or a specific one for the function. Moreover, we have self-sensing cavitation detection (SSCD) as well. In this method, the different frequency components, which are acoustic indicators for cavitation, can also be analyzed from the electrical signals of the ultrasound transducer. In other words, the spectrograms of the transducers current signal for different current amplitudes are used to classify the cavitation (Saalbach et al. 2018). Figure 2.5 illustrates the ACD, SSCD, and PCD methods.

Figure 2.5 – Schematic of ACD, SSCD, and PCD approaches for characterizing cavitation



Source: P. Wu, X. Wang, et al. 2021.

Those are some of the most common methods applied nowadays as a feedback mechanism for cavitation activity detection, and each method is more appropriate for each experimental scenario, having its own limitations. Single-element PCDs for example are restricted by the trade-off between spatial sensitivity and specificity (Haworth, Salgaonkar, et al. 2015), meanwhile, due to focusing and size constraints, the ACD system is difficult to be applied in industrial practices (P. Wu, X. Wang, et al. 2021). Considering those detection techniques, the chosen approach to analyze the acoustic emissions is based on the PCD physical setup, and more details for the simulation arrangement are described in Section 5. In the next section, we briefly summarize artificial intelligence methods, which is the proposed detection strategy in the proposed study.

2.5 ARTIFICIAL INTELLIGENCE AND CAVITATION PREDICTION

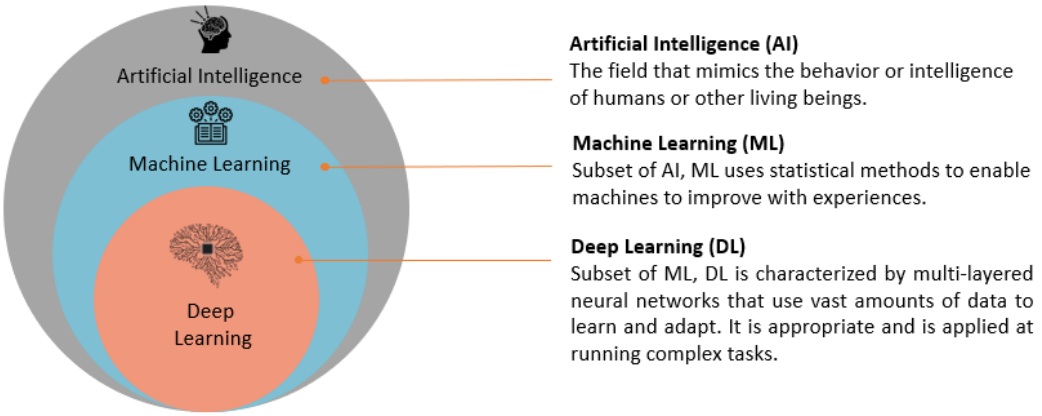
John McCarthy, recognized as the father of Artificial Intelligence (AI), defines it as

> The science and engineering of making intelligent machines, especially intelligent computer programs (McCarthy 2022).

In other words, AI's main goal is to create systems that understand, learn, and behave like humans. It has been used to develop and advance a wide range of fields, some applications include translation between languages, speech recognition, and visual perception. AI is also a tool that is extensively applied in medical sciences: diagnosing, remotely treating patients, and cancer detection based on CT scans are examples of its significant impact in healthcare (Basu et al. 2020).

Another application, still in the medical context, is the detection of anomalies during signal processing, which plays an important role in patient monitoring, indicating to be essential for treatments in general, including CVDs. The concern of this work is to analyze the acoustic emissions produced by the MBs when exposed to an acoustic field and explore AI for cavitation detection. The investigation to identify the phenomenon during the bubble activity is not recent in hydraulic systems, where the solutions based on Machine Learning (ML) are numerous (Dutta, Vishnu, et al. 2018; B.-S. Yang et al. 2005; Dutta, Subramaniam, et al. 2020). For our scenario, we want to detect cavitation in a healthcare application.

Figure 2.6 – AI and its subsets Machine Learning (ML) and Deep Learning (DL)

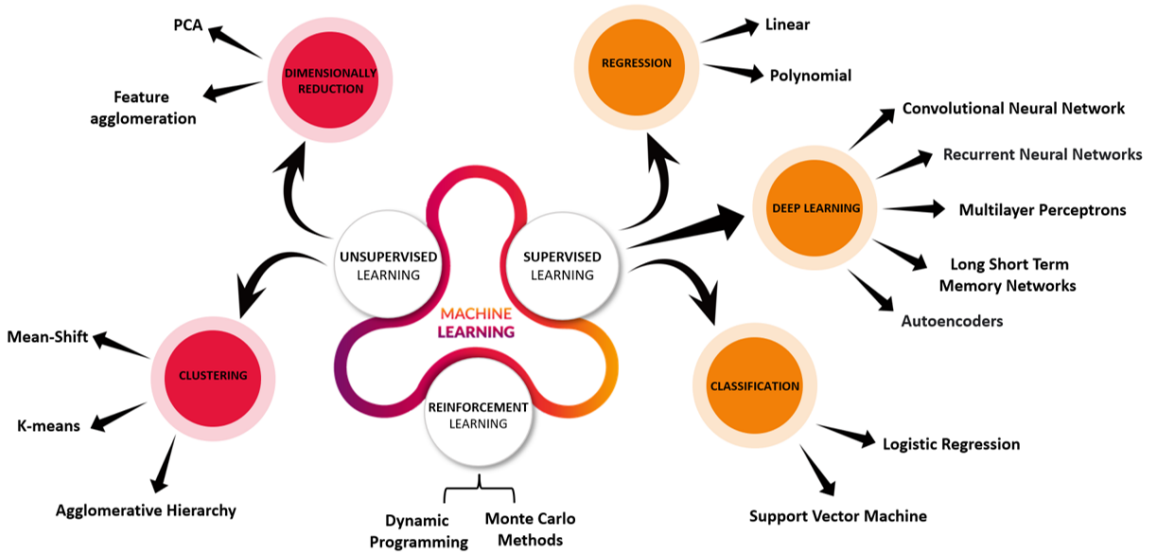


Source: Figure based on (Vrana and Singh 2020).

The analysis and detection of cavitation by AI is part of the feedback mechanism we aimed to achieve and seems to be a useful tool with the capability of processing raw data into clinical decision-making for effective detection. To do so, the range of possibilities that AI provides to the study topic starts with the AI subsets (Figure 2.6). Notably, the complexity of the proposed problems demands same-level solutions, and in each of the subsets, there are a large number of approaches to the problem.

The exploring options are exemplified in Figure 2.7. The ML subdivision is commonly branched into three fields: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning has the goal of finding the connection between the input data and the output target in a way to effectively reach the correct output. It is important to highlight that the question, of whether the result is right, is determined by the training data, depending on our assumptions and labeling. So issues such as incorrect data description or noise would reduce the effectiveness of models in the study. On the other hand, the idea of unsupervised learning is to make use of data without explicitly provided labels, and this attribute leads to the difficulty of discovering a way to compare the performance between the methods applied if more than one strategy is presented. Meanwhile, the reinforcement learning method is based on rewarding desired behaviors and/or punishing undesired actions. This means, the concept is to lead the action model to maximum cumulative reward feedback, where the model learns from its mistakes.

Figure 2.7 – Techniques of ML



Source: Figure adapted from (V. Sharma 2018).

Figure 2.7 shows just a few of the methods, and there are many others. We present more details in the Methodology section, but in advance, the proposed study explores a DL approach, categorized as a supervised learning. DL has become a reference in different data analyses because of its potential to recognize patterns. In healthcare, its applications cover fields such as medical imaging, drug delivery, and disease diagnosis, and extend into numerous other implementations (D. K. Sharma et al. 2022). We focused on DL methods because (i) this subset of AI eliminates some of the data pre-processing that is typically necessary for ML methods; (ii) it has shown substantial success in the different fields already cited; and (iii) it is appropriate for complex tasks. Even though there is a lack of literature on cavitation detection for biological signals, the phenomenon's analysis in hydraulic systems is well described, and it shows excellent results when DL approaches are employed (Sha et al. 2022; Tiwari et al. 2021). Therefore, we hypothesized that it could yield significant accuracy in the present study as well.

Inside the DL subset, the feed-forward neural network selected was the Convolutional Neural Network (CNN). CNNs are well-known for their effective image recognition (Long et al. 2015), text recognition (Vaillant et al. 1993), and other applications. Since CNNs focus on different kinds of features at each level (pixel level in the first layer, for instance), the network structure is highly invariant to translation, scaling, or other forms of deformation (Uijlings et al. 2013). Besides that, explicit feature extraction is avoided and learning is implicitly performed from the training data, enhancing the architecture performance and minimizing computational expenses (Zhang et al. 2018). Additionally, due to the consistent neuron weights across the same feature map, the network can undergo parallel learning (S. Lee et al. 2017). The distinctive local weight-sharing structure reduces the training parameters of the network, which results in a simpler structure and more versatile neural network. Another adaptability feature in CNNs is the use of linear filters. Linear filters tend to perform well with small window widths, and in image processing, the input image and topology of the network can be better matched (Taye 2023).

Considering this context and these benefits, our research explored the supervised learning approach, more specifically a CNN proposal. Taking into account that CNNs are specialized in tasks involving image classification into different predefined classes, this work methodology involved the conversion of the echo signals to images to detect the cavitation phenomenon, and the main features of CNNs are presented in the following section.

3

CONVOLUTIONAL NEURAL NETWORK

In this chapter, we describe the convolution neural network class. Further details about the DL subset are presented in the Appendix A.

3.1 INTRODUCTION

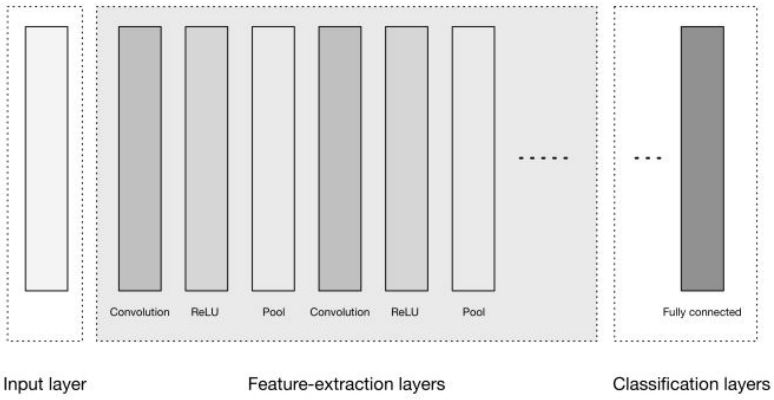
Haykin defines a convolutional network as

> A convolutional network is a multilayer perceptron designed specifically to recognize two-dimensional shapes with a high degree of invariance to translation, scaling, skewing, and other forms of distortion (Haykin 2009).

A typical CNN is compounded by convolutional layers, pooling layers, and a fully connected layer, as we can verify in Figure 3.1, and all weights in every layer of a convolutional network are learned through training.

In the Feature-extraction group, each neuron takes its synaptic inputs from a local receptive field in the previous layer, thereby forcing it to extract local features (Haykin 2009).

Figure 3.1 – A general CNN architecture



Source: Patterson and Gibson 2017.

The following sections provide an introduction to the constituent layers of CNNs, considering the calculations involved.

3.2 CONVOLUTION LAYER

A convolution is a mathematical operation that merges two sets of information (Patterson and Gibson 2017). The convolution of two functions, f and g, produces a third one $(f * g)$ defined as the integral of the product of these functions after one is reflected and shifted. It expresses how the shape of one function is modified by the other, and the symbol $*$ denotes the convolution operator. Equation (3.1) brings the convolution of the two functions.

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau)\, d\tau \tag{3.1}$$

In convolutional network terminology, for equation (3.1), the first argument (function f) is often referred to as the input. The second argument (function g) is attributed as the kernel, and the output can be indicated as the feature map.

For the case of digital information, time is discretized and the data is provided at regular intervals, i.e., t can have only integer values. Therefore, we can define the discrete

convolution as equation (3.2).

$$(f * g)(t) = \sum_{\tau=-\infty}^{\infty} f(\tau)g(t-\tau) \tag{3.2}$$

In machine learning applications, both input and kernel are usually multidimensional arrays of data. However, since each element of them must be explicitly stored separately, and these functions are usually assumed to be zero everywhere but the finite set of points for which we store the values, consequently, the implementation can be summarized as a summation over a finite number of array elements (Goodfellow et al. 2016).

Thus, given a two-dimensional image, where I (input) is convolved by a kernel K (also called *convolutional filter*), equation (3.3) can be used.

$$S(i,j) = (I * K)(i,j) = \sum_{m}\sum_{n} I(m,n)K(i-m, j-n) \tag{3.3}$$

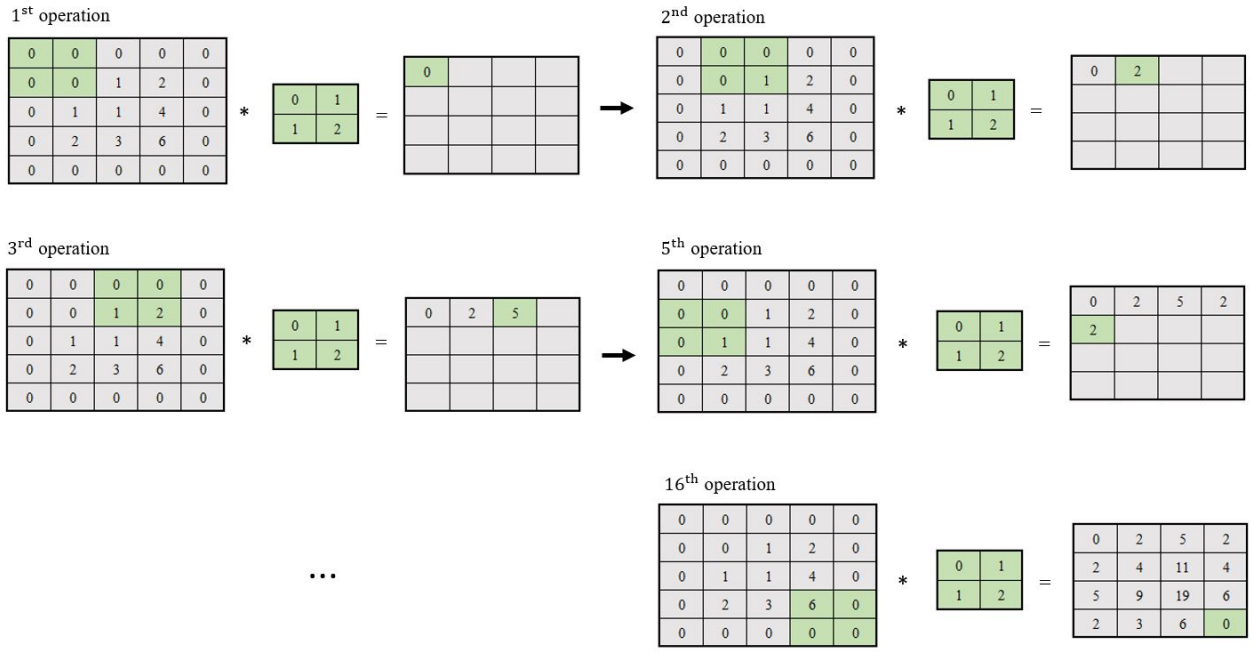
and since convolution is commutative, the equivalent equation can be written

$$S(i,j) = (K * I)(i,j) = \sum_{m}\sum_{n} I(i-m, j-n)K(m,n) \tag{3.4}$$

The convolution operator is also called the striding method since the kernel sweeps the input until the first dimension is reached, and stride is the number of pixels that shift over the input matrix. Through the same stride parameter, the kernel slides across the second dimension. For i and j where there are undefined values of $I(i-m, j-n)$, there are two options for adjustment: the padding operation (zero-padding) or discard the part of the input where the filter does not fit (valid padding).

For didactic purposes, Figure 3.2 presents the process for a two-dimensional case. In this example, the matrix $[0, 1; 1, 2]$ is our 2×2 kernel, and the highlighted area in the input matrix represents where the current operation occurs. The input is a 3×3 size image, padding of value 1, and stride of 1 as well. The operation is often referred to as convolution. To be more accurate, this procedure is called cross-correlation and not convolution. In the AI field, both operations are represented using $*$ and referred to interchangeably for convenience (Venkatesan and B. Li 2017).

Figure 3.2 – Example of the convolutional computation, where $1^{st}, 2^{nd}, 3^{rd}, 5^{th}$, and 16^{th} steps are represented



Source: Own work.

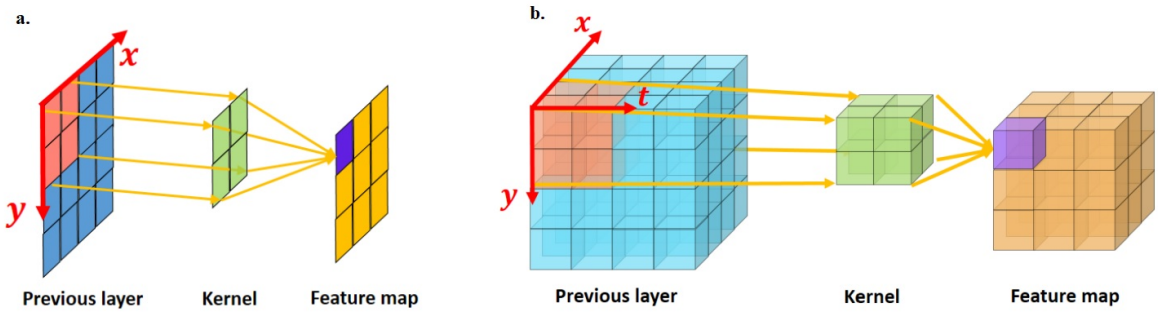
As we can notice, the output does not have the same size as the input. To calculate the convolution output size, we apply equations (3.5) and (3.6), where I refers to the input, f to the filter, and the subscripts w and h are for width and height, respectively. p and s are the padding and stride values (F.-F. Li et al. 2023).

$$\text{output width} = \frac{I_w - f_w + 2p}{s_w} + 1 \tag{3.5}$$

$$\text{output height} = \frac{I_h - f_h + 2p}{s_h} + 1 \tag{3.6}$$

Figure 3.3 illustrates the difference between 2D and 3D cases, where $p = 0$ and $s = 1$. In 3D convolution, the filter depth is smaller than the input layer depth. Consequently, the 3D kernel can move in all 3 directions, and besides the width and height, the third dimension is the channel of the image, and the output numbers are arranged in a 3D space as well.

Figure 3.3 – (a) 2D convolution. (b) 3D convolution



Source: Huang et al. 2015.

3.3 POOLING LAYER

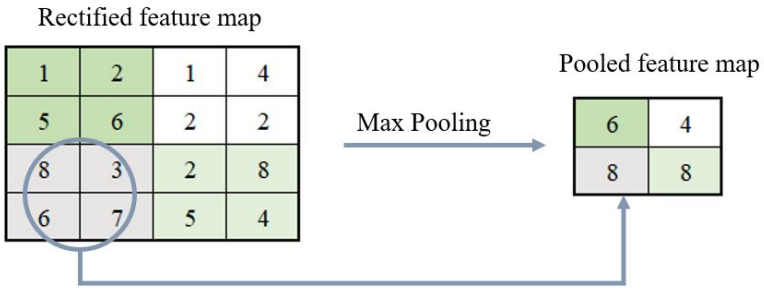
The Pooling layer is usually employed immediately after the convolutional layer, and it is used to perform downsampling from one layer to another, which reduces the number of feature-map coefficients. The operation helps to make the representation approximately invariant to small translations of the input, so if we apply a small amount of shift, the values of most of the pooled outputs do not change (Goodfellow et al. 2016).

Two common functions used in the pooling operation are Average Pooling and Max Pooling, and we describe both ahead.

3.3.1 MAXIMUM POOLING

Maximum Pooling, more usually called Max Pooling, calculates the maximum value for each patch of the feature map. It extracts stronger features and has been found to work better than average pooling for tasks such as image classification (Chollet 2017). Figure (3.4) presents an example of the operation: in the four squares of the gray rectified feature map, we take the maximum value among $(8, 3, 6, 7) = 8$, and place this in the gray cell of the pooled feature map.

Figure 3.4 – Example of Max Pooling operation with a kernel of 2×2

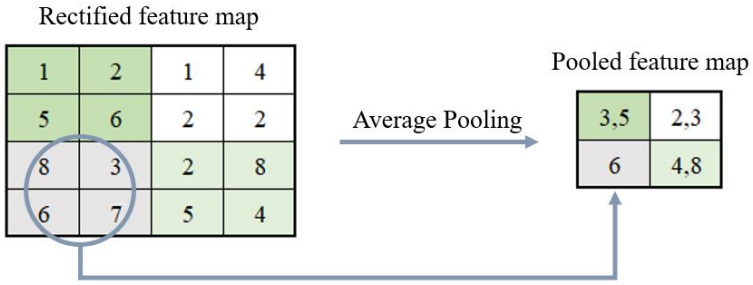


Source: Own work.

3.3.2 AVERAGE POOLING

The average pooling calculates the average value for each patch on the feature map. Figure (3.5) shows an example of the process: in the four squares of the gray rectified feature map, we take the average value among $(8, 3, 6, 7) = 6$, and place this in the gray cell of the pooled feature map.

Figure 3.5 – Example of Average Pooling operation with a kernel of 2×2



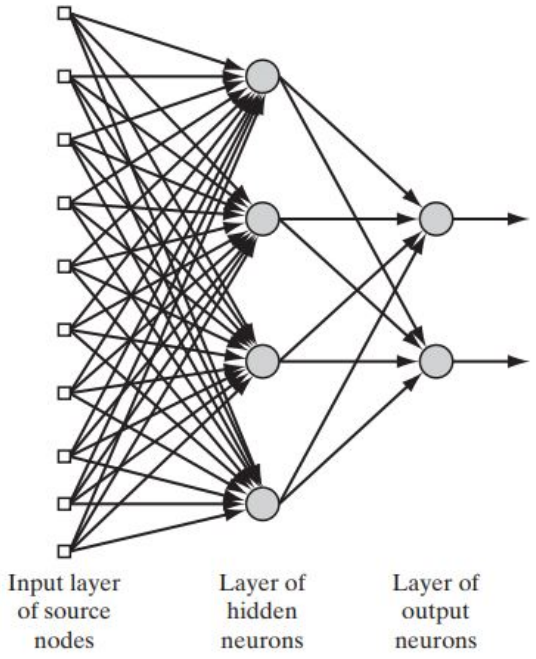
Source: Own work.

3.4 FULLY CONNECTED LAYER

Fully connected layers (FCs), also called densely connected, perform the dot product between current inputs and weights in each layer. Thus, all possible connections layer-to-layer are present, resulting in the influence of every input vector element in every component of the output vector, and Figure 3.6 illustrates the layout of an FC feedforward neural network.

As we saw in Figure 3.1, the FCs are the last structure of CNNs and their function is classification. It is important to understand that during the computational processing in the feature-extraction layers, the data are three-dimensional matrices while the FC has a unidimensional characteristic. Thus, to connect the convolutional and FC layers, it is necessary to convert the three-dimensional into one-dimensional matrices.

Figure 3.6 – Fully connected feedforward network with one hidden layer and one output layer



Source: Haykin 2009.

3.5 ACTIVATION FUNCTION

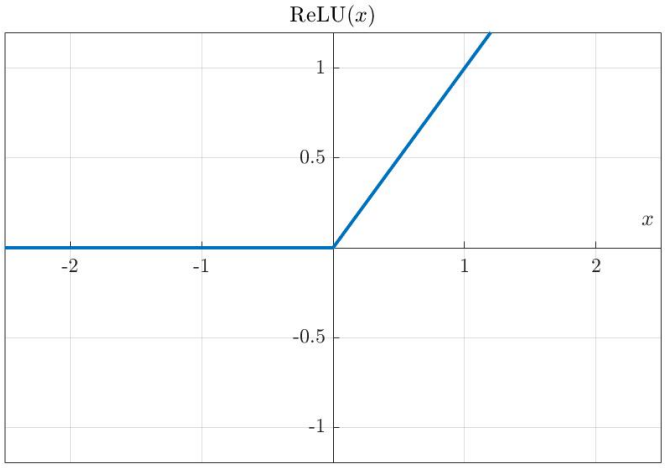
The activation functions are responsible for transforming the summed weighted inputs from the neuron and addressing them to the next layer. They implement the nonlinear component to the model, permitting the performance of more complex tasks. Among the most common activation functions, the rectified linear unit (ReLU) is used to learn complex structures from the data and it is computationally efficient.

ReLU is defined by equation 3.7, and its greatest advantage is the non-saturation of the gradient, which accelerates the convergence of the algorithm optimization.

$$\text{ReLU}(x) = max(0, x) \tag{3.7}$$

Figure 3.7 represents the function's plot. We can observe that the activation is a simple threshold $x = 0$, making ReLU very efficient (Krizhevsky et al. 2012).

Figure 3.7 – ReLU Function for $x \in [-2, +2]$



Source: Own work.

Having briefly stated the main features of a CNN, we can now discuss possible approaches to be explored in our problem.

3.6 IMAGE CLASSIFICATION MODEL

Among popular models for image recognition, each with its own strengths and weaknesses, some examples of deep CNN architectures are LeNet, AlexNet, Visual Geometry Group (VGGNet), and Residual Neural Network (ResNet).

One of the earliest CNNs, LeNet (Lecun et al. 1998) was designed for handwritten digit recognition. Its architecture consists of several convolutional and pooling layers, followed by fully connected layers. VGGNet (Simonyan and Zisserman 2014) is characterized by using small filters (3×3), and two examples of its variants include VGG-16 and VGG-19, which have 16 and 19 layers, respectively. ResNet (He et al. 2015) has several variants as well, ResNet-50 and ResNet-150 have 50 and 150 layers, respectively.

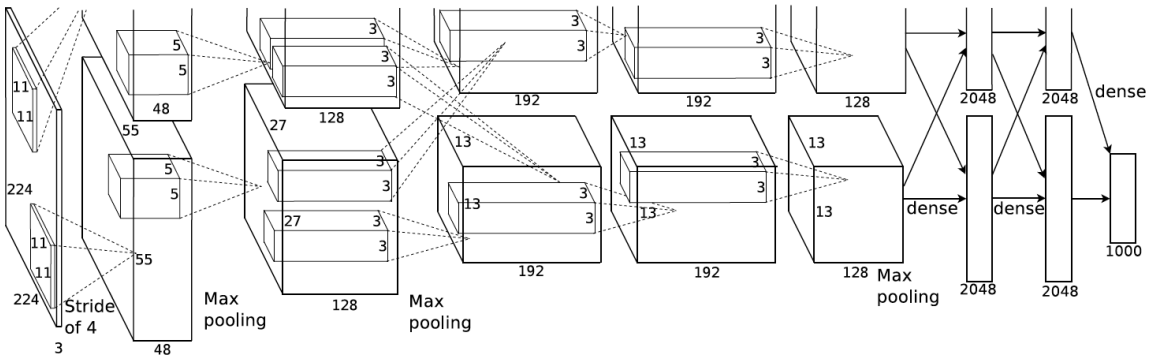
When compared to these other models, AlexNet has less depth, holding 8 layers. However, the selection of a complex architecture is not always necessary and the computational cost needs to be evaluated as well. Considering the resource availability, and the authors' experience with the chosen CNN, this study chose the AlexNet architecture, and details about this CNN are described below.

3.6.1 ALEXNET

The AlexNet network (Krizhevsky et al. 2012) was introduced in 2012 in the ImageNet Large Scale Visual Recognition Challenge competition, resulting in the victory by a significant margin. The event is an annual contest for large-scale object detection and image classification algorithms, and in contrast to other respondents who used standard classification characteristics and training techniques, the winners used neural networks, particularly CNNs. AlexNet is a CNN and is a leading architecture for any object-detection task, it is composed of 3 fully connected layers and 5 convolutional layers. The first layer is used to input a filtered image with a dimension of $227 \times 227 \times 3$ respectively for width, height, and color channel (red, green, blue). Its architecture can be seen in Figure 3.8.

The last fully connected layer connects 1,000 neurons and the rest of the layers work as feature extractors. To give an insight into the robustness of the model, its architecture includes 650,000 neurons and 60 million parameters, being trained on approximately 1.2 million training images and tested on 150,000 test images from ImageNet datasets (large-scale hierarchical image database).

Figure 3.8 – AlexNet architecture: ImageNet 2012 winning CNN model



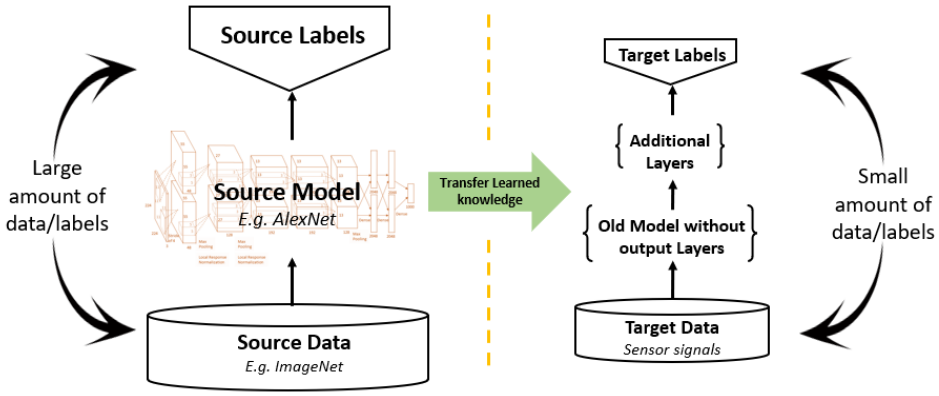
Source: Krizhevsky et al. 2012.

Taking advantage that the network has learned rich feature representations for a wide range of images, we employed the AlexNet model as a transfer learning framework with pre-trained ImageNet weights. More details about the knowledge transference are presented following.

3.6.2 TRANSFER LEARNING

An interesting approach enabled by ML is the use of Transfer learning (TL). TL is a tool that makes use of the knowledge gained while solving one problem and applies it to a different problem in a similar application. So, to implement this idea, we make use of the AlexNet CNN model, thus we take a pre-trained network and use it as a starting point to learn a new task.

Figure 3.9 – General idea of Transfer Learning



Source: Figure adapted from Natrajan et al. (2018).

So, the idea is to reuse a pre-trained model (AlexNet) with a new problem and remove the need to create a CNN model from scratch. Figure 3.9 illustrates the basic idea of TL for AlexNet. The last three layers of AlexNet are replaced (Fully Connected, SoftMax, and Classification) considering our problem attributes so the model can learn specific features from our dataset.

AlexNet input starts with $227 \times 227 \times 3$ images, and since we are inspecting signals, before feeding the network, we transform and decompose those waves in the sized inputs. To understand the mathematical tool applied for feature extraction, in the next Section, the Wavelet Transform decomposition is detailed.

4

WAVELET TRANSFORM

An important aspect when analyzing biomedical signals is that the information of interest is frequently a combination of phenomena. Classical signal processing tools, such as the Fourier Transform, are not suited for non-stationary signals, thus, an approach that performs multi-scale analysis is necessary to manage our inputs. Processing is faster and simpler in a sparse representation where few coefficients reveal the information we are looking for (Mallat 2008), and a well-structured mathematical approach is fundamental to analyzing those input signals efficiently.

An interesting technique is the Wavelet Transform (WT), which has shown good results for physiological signals studies, for example, electrocardiogram (ECG) cycles investigations (T. Wang et al. 2021; Mashrur et al. 2019). WTs can be classified into two broad classes: the Continuous Wavelet Transform (CWT) and the Discrete Wavelet Transform (DWT). DWT is useful for compressing and denoising signals and images while maintaining important features since its scales are discretized more coarsely than with CWT. For our problem, we want to differentiate cavitation signals, and since those signals are structured of different frequency components and are non-stationary, the CWT could be a more appropriate tool for exploration. The signals are transformed into the time-

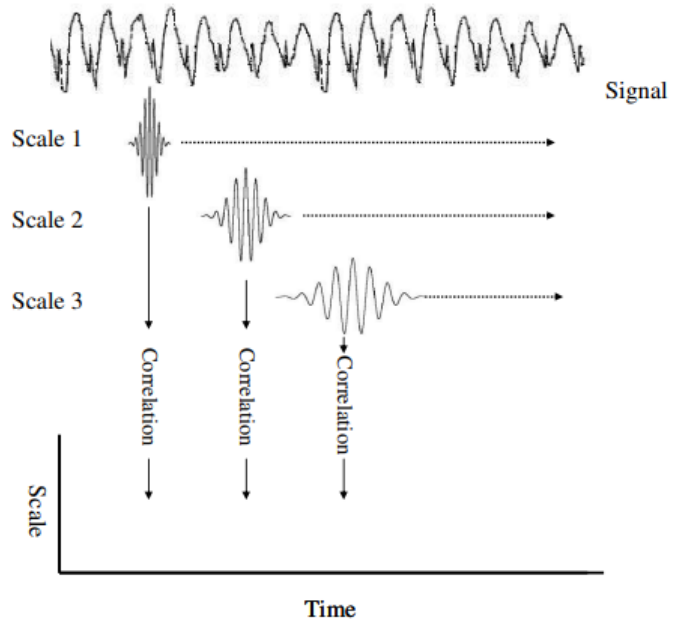
frequency domain, decomposing a signal into wavelets, which facilitates the extraction of features.

As the CWT window shrinks and expands, it maps the signal differences onto a time-frequency spectrogram, thus inheriting and developing the idea of location from the Short-time Fourier transform (STFT). The drawback of STFT is related to the dilemma of resolution, i.e., a narrow window (good time resolution) provides a poor frequency resolution, and a wide window (poor time resolution) results in a good frequency resolution. Such uncertainty of what frequency exists at what time intervals is overcome in the CWT analysis. The mathematical expression of the CWT for a signal $x(t)$ is defined as

$$C(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) \cdot \Psi \left(\frac{t-b}{a} \right) dt \tag{4.1}$$

where a and b are the scales and time shifts of a reference wavelet $\Psi(t)$, respectively, and by continuously varying those parameters, the CWT coefficients $C(a,b)$ are obtained.

Figure 4.1 – Mapping the wavelet scalogram



Source: Shoeb and Clifford 2006

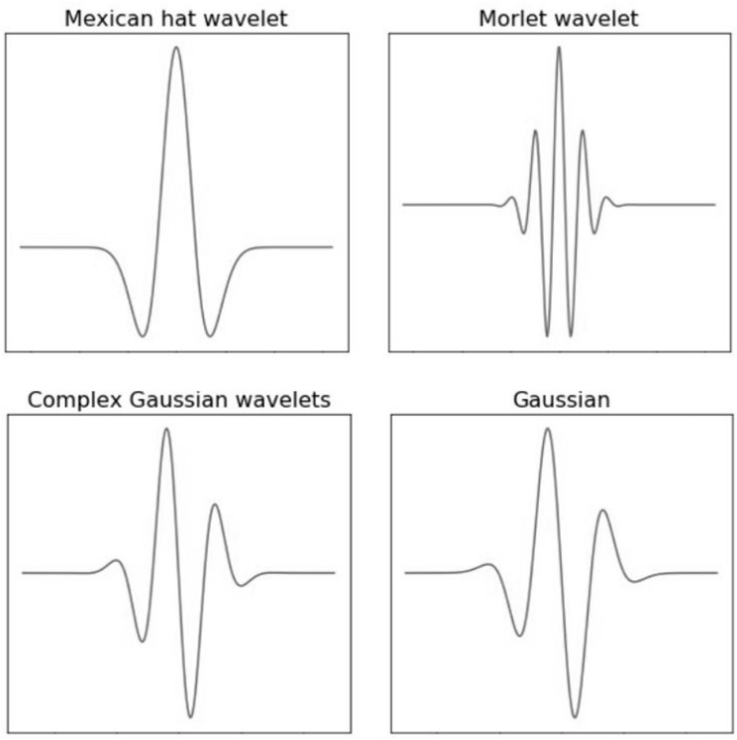
The results of the equation 4.1 are a set of coefficients, which represent the correlation with the wavelet $\Psi(t)$ and the signal $x(t)$. The correlation is localized in time, where the

time begins at $t = b$ and ends at $t = b + L$, and L is the $\Psi(t)$ duration. The plot of such correlation is called a scalogram, and the idea of "matching" the signal and the chosen wavelet for different scales, is illustrated in Figure 4.1.

For the $a < 1$ case, the wavelet is contracted, offering high temporal resolution, which fits well with short-time events such as spikes. On the other hand, values of $a > 1$ result in high spectral resolution since the wavelet is dilated, and is well-suited for long-term events such as baseline oscillations. The scales a from the CWT are used to create the wavelet bandpass filters and depending on the chosen parameters as input for the CWT, those values change. The number of levels of decomposition depends on the problem being inspected, the sampling frequency, and the range frequency of interest. Thus, a few parameters have to be tested so the finest value can be chosen.

Another important parameter is the choice of the wavelet family, also called wavelet mother. Some examples are: Gaussian Wave, Mexican Hat, and Morlet (Gabor), and the goal is to choose one with the most similar morphology to the signal, which follows better outcomes. In advance, the wavelet that yielded the best results for this study was the Morlet.

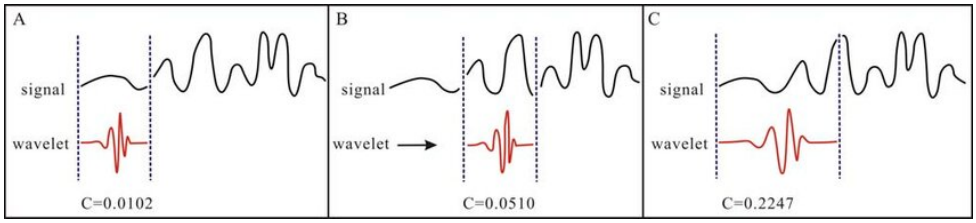
Figure 4.2 – Some families of CW



Source: Bajaj 2020

Figure 4.3 illustrates an example of the wavelet transform decomposition applying the Gaussian wavelet, and the coefficients calculation. Since the $C(a, b)$ coefficients reflect the correspondence between the selected wavelet function and the signal $x(t)$, the greater the $C(a, b)$, the higher the correlation. Figures 4.3A and 4.3B show the wavelet function moving rightwards along the time axis (parameter b), and the respective coefficient values. Figure 4.3C displays the condition in which the wavelet function is stretched (parameter a). The greater the value of $C(a, b)$, the more significant the similarity between the section of the signal inspected and the wavelet.

Figure 4.3 – Example of a general step of wavelet transform



Source: Liu and Jiang 2010

The same procedure is applied to our problem, and after the scanning, we obtain the coefficients produced at different scales by different sections of the signal. It is important to clarify that the coefficients acquired reveal patterns from the signals and are essential for the feature extraction of our neural network. Following, the signals' decomposition, the idea is to convert the obtained coefficients into images, and those figures will be the input for our neural network model.

More details of the CWT approach for the problem are presented in Section 5.6.2.

5

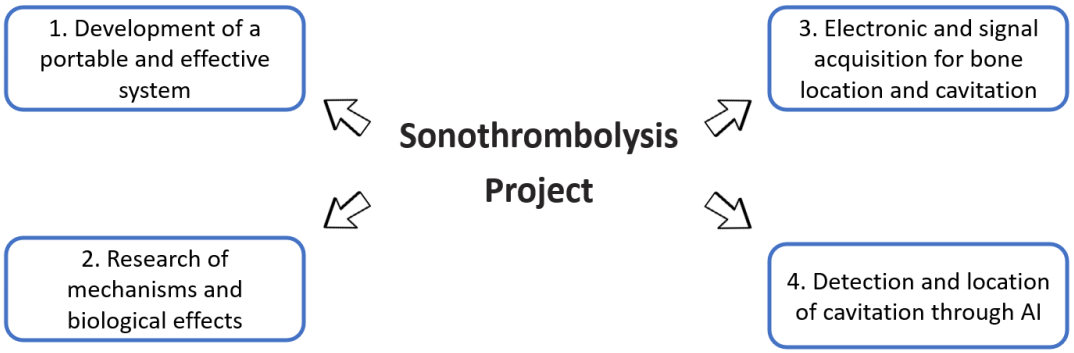
METHODOLOGY

In this section, we outline the selected approach to develop the feedback mechanism involving AI, and how we manipulated and inspected the acoustic emissions. But before we start to explore the signal analysis and the detection method, we provide an overview of the different studies being conducted in the research group this study is part of.

5.1 BIOMEDICAL ENGINEERING LABORATORY (LEB)

The overview of the Sonothrombolysis project being operated in the LEB is branched as in Figure 5.1. The first subsection is the development of a portable and effective system, and the goal is to assemble a prototype for the therapy (Silva and Furuie 2024). The second ramification study aims to evaluate the mechanisms and biological effects due to the Sonothrombolysis technique and verify the thresholds for the application (Souza and Furuie 2023). The third division explores the electronic and signal acquisition for bone location and cavitation, inspecting the heterogeneous scenario. And lastly, this work itself.

Figure 5.1 – Overview of the Sonothrombolysis project in the LEB



Source: Own work.

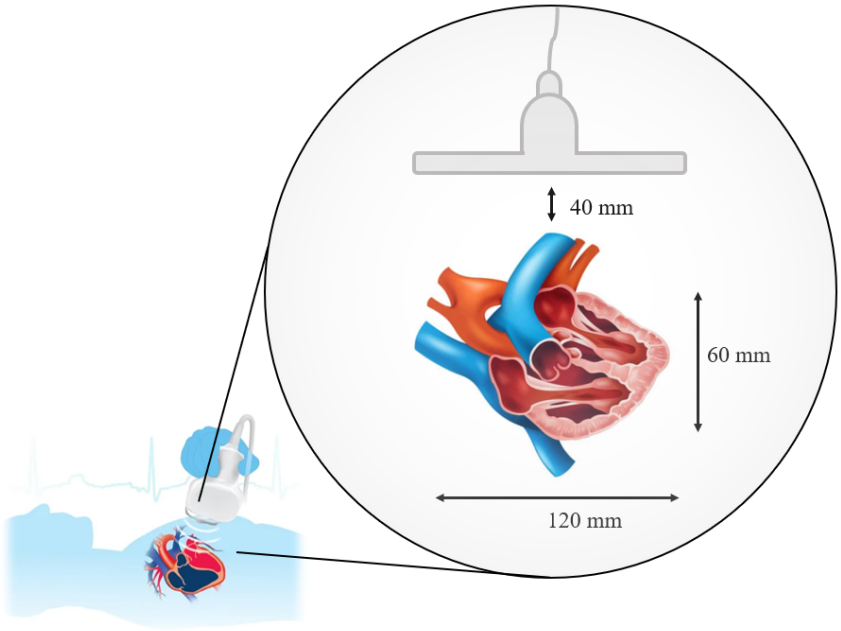
For this work, it is important to provide an overview of the physical experiments as well since the numerical simulations are directly linked to those conducted tests, in a way to try to mimic reality. In this sense, the details are provided ahead.

5.2 PHYSICAL EXPERIMENTS

The experimental procedure that is being developed by the laboratory group aims to explore and propose an equipment prototype for the recanalization of thrombotic occlusion, along with investigating ways to improve the efficiency of the sonothrombolysis technique. The idea is to make use of a 2-D array of transducers capable of producing beamforming ultrasound waves that can be guided to sonify a defined 3-D volume (heart and surroundings).

To understand the selected parameter settings exhibited below, first, we need to inspect the target volume, i.e., the heart dimensions. Mean values of the heart's length, width, and thickness in Grays Anatomy are 120 mm, 85 mm, and 60 mm, respectively (Gray 2005). These measurements can be seen in Figure 5.2, and are important when we consider the region of interest (ROI), which is the cardiac volume.

Figure 5.2 – Standard heart dimensions and distance between patient skin and heart.



Source: Figure assembled from (123RF 2022).

The average skin-to-heart parasternal distance found in adults is 32.1 ± 7.9 mm (Rahko 2008), so the deepest the ultrasound wave needs to penetrate is 100 mm since the thickness value adopted is 60 mm and the skin-to-heart distance can reach 40 mm. In other words, the attenuation factor is an important key to be analyzed, and the maximum value set in the project is 3 dB (Silva and Furuie 2024). Considering that depth and the limit attenuation, the maximum center frequency could be evaluated, and it was found that it should be lower than 428 kHz (Silva and Furuie 2024).

Out of possession of the cited aspects, the piezoelectric transducer from American Piezo Ceramics (APC), whose features are present in Table 5.1, was selected to perform the physical tests and simulations.

For the current experiments, 64 piezoelectric ceramics were assembled in the form of an 8×8 matrix (Figure 5.3), and an acoustic tank of approximate internal dimensions of 550.1 mm × 302.4 mm × 200.0 mm (length×width×height), designed by the Mechatronic Engineering Department of the University of São Paulo, was used in the tests. After the electric excitation of the elements, a needle-type hydrophone was used to measure and map the acoustic field.

It is important to acknowledge the features of the physical settings so the numerical

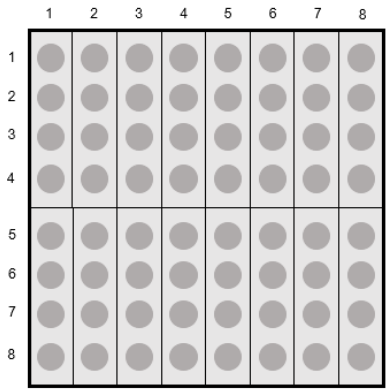
Table 5.1 – Features of the initial chosen transducer, as informed by the manufacturer

Transducer features	
Geometry	Circular plate
Diameter	10.15 mm
Thickness	6.4 mm
Center frequency	318.75 kHz
Brand	APC
Material	APC850
Model	408

Source: APC.

analyses draw nearer to substantial results.

Figure 5.3 – Illustration of the project's piezoelectric transducer matrix array (two-dimensional)

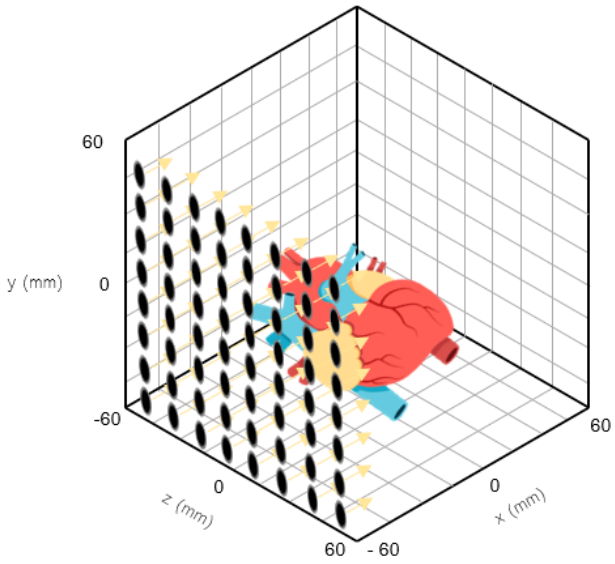


5.3 NUMERICAL EXPERIMENTS

In order to describe the simulations that are being performed, this section brings the parameters considered for the model.

First, we explore the geometry and coordinate system applied. As already mentioned, the ROI consists of the volume composed by the heart and surroundings, therefore, to ensure scanning the entire target, the volume of simulation selected is $120 \times 120 \times 120$ mm^3. The transducers' faces are in the yz-plane, evenly distributed, and located in $x_0 = -60$ mm.

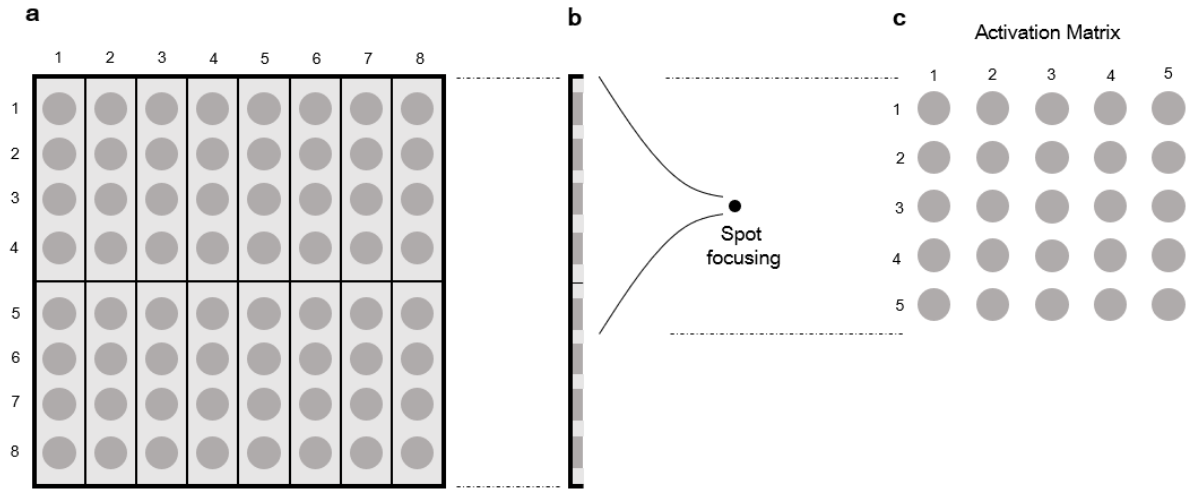
Figure 5.4 – Normals at transducers' centers.



Source: Figure based on author's advisor results.

To understand the geometry better, Figure 5.4 shows the transducer elements, ROI, and normals at transducers' centers.

Figure 5.5 – (a) Matrix array (two-dimensional). (b) Selected focus. (c) Activation matrix



Source: Figure based on (EWI 2010).

The geometry for the computer simulation follows the description mentioned in Figure 5.3 and Table 5.1. We assume initially that each element is capable of emitting and receiving ultrasound pulses, and the focus spot is achieved by exciting multiple elements simultaneously with a relative delay between them. The idea is to activate a submatrix

of 5×5 elements, Figure 5.5c for example, and focus on a target spot (Figure 5.5b).

Computer simulations were conducted using k-Wave (Treeby et al. 2020), an open-source acoustics toolbox for MATLAB. The basic modeling features are displayed in Table 5.2, some of them were evaluated by other projects of the group, and some during the simulation tests. The estimation for the number of grid points can be made by the function *checkFactors*, which will return appropriate values to improve computational performance. But more than that, to avoid the aliasing effect it was concluded that the voxel dimension should be smaller than $d = 2.415$ mm in the three directions (Silva and Furuie 2024), which led to the result of 128 pixels/axis.

Table 5.2 – Simulation features

k-Wave simulation features		
Computational grid	Grid nodes	128 pixels/axis
Acoustic Medium	Isentropic sound speed	1540 m/s
	Ambient mass density	1000 kg/m^3
	Nonlinearity parameter	Non-defined
	Absorption coefficient	0.7 dB/(MHz cm)
	Absorption exponent	1.1
Acoustic Source	Central frequency	250 kHz
	Number of cycles	7
	Source magnitude	100 kPa
	Elements' geometry	Circular plate
	Number of elements	64
	Elements' distribution	8 × 8
	Activation matrix	5 × 5
	Distance between centers	15.3 mm
	Distance between edges	5.6 mm
Sensor	Sensor mask	Default

Source: Own work, values assembled from the group's work.

The setting parameters considered for the acoustic medium follow the characteristics of the average soft tissue for humans (Culjat et al. 2010; Mast 2000), the simulation is assumed to be linear, the mean attenuation for the human body is applied (Szabo 2004), and the absorption exponent was selected during some tests that probably will be refined

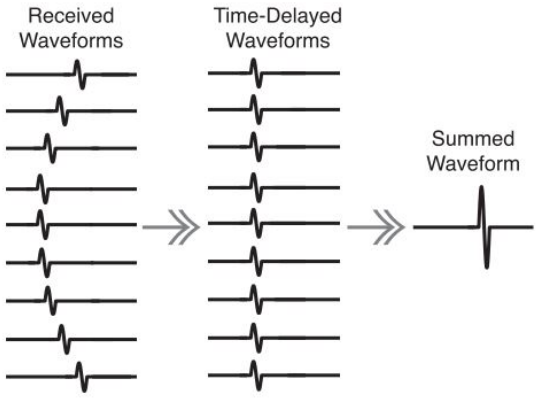
after some deeper studies related to the heat impacts on the tissue.

As we mentioned in the Physical Experiments section, the acoustic source features obey the attributes of the APC transducer and the geometry already discussed. Equation 5.1 brings the relation between the source magnitude and the desired Mechanical Index (MIx). Considering the central frequency of 318.75 kHz from the manufacturer manual (Table 5.1) and MIx = 0.5, we acquire the value of $P_r = 282$ kPa for the rarefactional acoustic pressure, and to avoid focal overlapping, 200 kPa was selected so we assure that the MIx does not exceed the desired estimation. However, this value of pressure was still so high that it triggered reflections, so the amplitude used in the simulations was 100 kPa.

$$\text{MIx} = \frac{P_r}{\sqrt{f_c}} \tag{5.1}$$

The resonant frequency obtained in the assembled transducer for the physical experiments was not 318.75 kHz as described in Table 5.1. As a result of the matching layer's insertion and steel welding, the center frequency obtained was near 250 kHz, and this frequency was applied in the numerical simulations. For the pulse duration, the number of cycles chosen was 7 after some tests.

Figure 5.6 – Waveforms delayed, consequently, they sum constructively

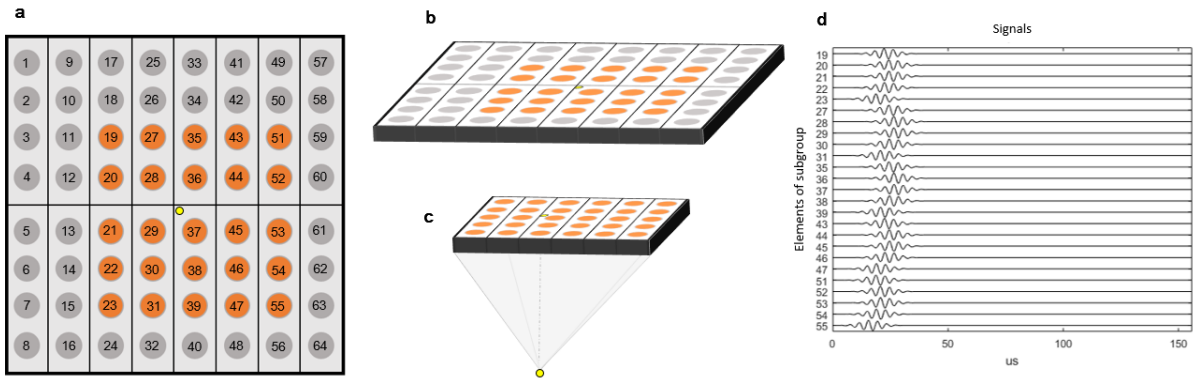


Source: Haworth, Bader, et al. 2016.

Due to the different distances along each source's centers and the focal location, it is required to calculate the delays for each element in order to energize the focusing spot simultaneously by all components of the activation matrix. The idea is to make the pulses arrive at the same time so the waveforms sum constructively, which provides a higher amplitude and more energy. Figure 5.6 brings the concept for the linear array.

Even though our case is a 2-D array, the approach is similar. To contextualize the procedure, firstly we can imagine a projected focus (yellow point) in the transducers' faces plane, as in Figure 5.7a. Subsequently, the closest element to the projection is detected, which is the central element of the activation matrix, allowing the nomination of the subgroup matrix. Figures 5.7b and 5.7c bring the notion of different distances between elements and focus spots.

Figure 5.7 – (a) Matrix array (two-dimensional). (b) Example of activation matrix. (c) Active subgroup focusing on the target (illustrated by the yellow point). (d) Example of 25 signals from a 5×5 subgroup with their respective delays



Source: Own work.

Figure 5.7d illustrates how the signal for each element of the activation matrix behaves after applying the appropriate delays. The calculation relies on relative time-of-flight information, and we assume the excitation of the furthest element as the initial time, in this case, element 55. The general delay for any emitter i is

$$d_i = t_{max}(\vec{r}_f) - \frac{\|\vec{r}_i - \vec{r}_f\|}{c} \tag{5.2}$$

where $\vec{r}_f$ is the location of the focused spot, and $\vec{r}_i$ is the position of the emitter element ($i = 1 : n$, being n the size of the active subgroup). c is the ultrasound propagation velocity, and since the internal structure of the ROI is unknown, we use a constant velocity (c = 1540 m/s). Besides that, $t_{max}(\vec{r}_f)$ is related to the furthest element, and it is obtained by

$$t_{max}(\vec{r}_f) = \max_i \frac{\|\vec{r}_i - \vec{r}_f\|}{c} \tag{5.3}$$

In short, the closest element to the focal point is the last to be excited, and observing Figure 5.7a again, the nearest component is number 37. So, to exemplify the calculations for Figure 5.7 case, we can imagine the focus in $\vec{r}_f = (10; 0; 0)$ position for instance. As we already mentioned, the transducers' faces are in the yz-plane, and located in $x_0 = -60$ mm (Figure 5.4), so following the elements' distribution, we have

$$\vec{r}_{55} = (-60; 38.25; 38.25)$$

$$\vec{r}_{37} = (-60; 7.65; 7.65)$$

in other words, the distance between the elements and the focus are

$$d(f, 55) = \|\vec{r}_{55} - \vec{r}_f\| = 88.47\text{mm}$$

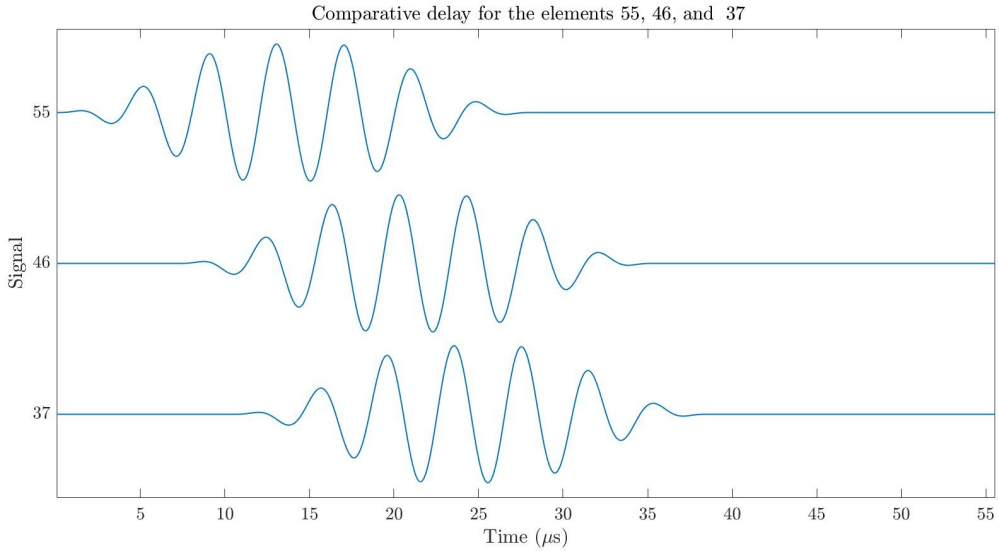
$$d(f, 37) = \|\vec{r}_{37} - \vec{r}_f\| = 70.83\text{mm}$$

which allows us to determine the delay we need to apply on element 37

$$d_{37} = \frac{[d(f, 55) - d(f, 37)]}{c} = 11.45\mu\text{s}$$

To illustrate the comparative delay, Figure 5.8 shows the excitation waveforms in elements 55 (the furthest from the focus), 37 (the closest to the focus), and 46, which is a component between the two. The previous analytical values obtained can be verified in the simulation below, and the other delays can be acquired the same way.

Figure 5.8 – Excitations waveforms in elements 55, 46, and 37 from Figure 5.7, showing the comparative delay between them. Parameters applied: Sampling frequency: 20 MHz, Central frequency: 250 kHz, number of cycles: 7, respective offsets



Source: Own work.

The instant that the initial stimulus (t_{hit}) achieves the focal point is determined by the shortest time-of-flight, in other words, it is associated with the element 55. We already know that $d(f, 55) = 88.47$ mm, therefore

$$t_{hit} = \frac{d(f, 55)}{c} = 57.4\mu s$$

After reaching the focused spot, the echo propagates towards the transducer, and now the elements behave as receivers. Depending on how those waves meet at the focused spot, we have some possibilities for the returned signals. For our analysis, there are two alternatives: the presence of MBs or not. In other words, the dynamic at the focus can be determined by the interaction of the waves with the MBs and if there is enough energy to instigate cavitation.

But in the same way, as when we generate the acoustic emissions, the time of travel is different to reach each element since the distances are different, and the received signals must be synchronized so the information can be combined. Besides that, it is important to mention that we are initially assuming that the transducers are capable of reading distinct ranges of frequencies.

The synchronization time for each receptor can be obtained by

$$t_{syn}(f,i) = t_{hit} + \frac{\|\vec{r}_i - \vec{r}_f\|}{c} \tag{5.4}$$

which allows us to match the received signals from the focal point, thus, it is possible to classify the type of source. The evaluation of the signal response is based on the delay-and-sum (D&S) beamforming algorithm, and the idea is to reconstruct the acoustic emissions received on each ultrasonic sensor and sum those emissions after applying the proper shifting (equation 5.4). The resulting signal can be estimated by

$$s_f(t) = \frac{1}{m}\sum_{j=1}^{m} s_j(t + t_j) \tag{5.5}$$
$$t \geq t_{syn}$$

where m is the number of active sensors, s_j is the signal received by each element, t_j is the time-of-flight for every receptor, and $j = 1 : m$, with $m > 1$. Besides, t_j can be obtained by

$$t_j = \frac{\|\vec{r}_j - \vec{r}_f\|}{c} \tag{5.6}$$

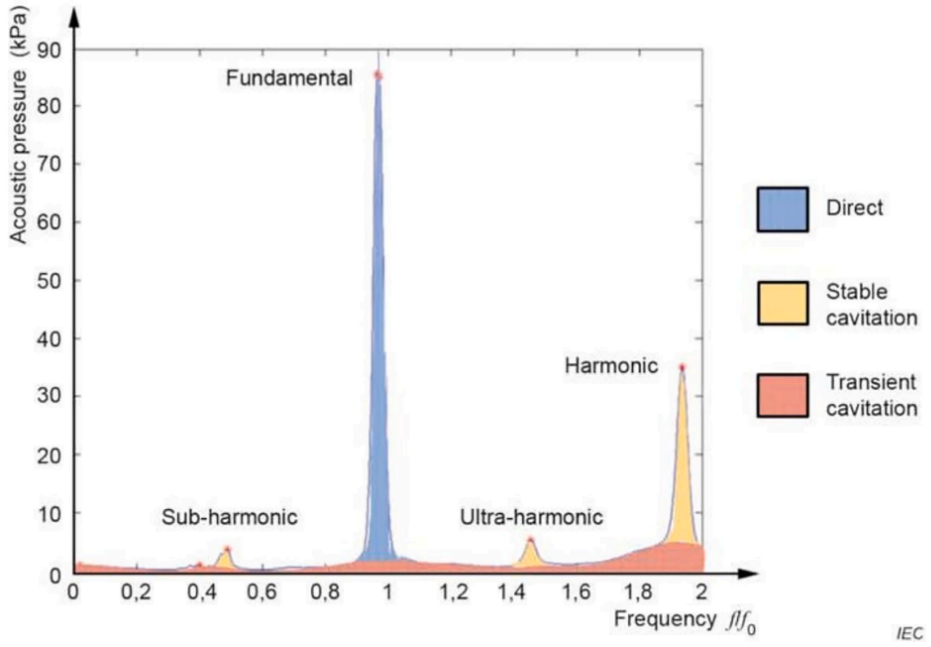
In addition to analyzing the signals in the time domain, the discussion in the frequency domain is important because it highlights the behavior characteristics of the cavitation phenomenon itself, both inertial and stable. For that, the next section brings the categorization based on the spectrum of received signals.

5.4 FREQUENCY AND CAVITATION CLASSIFICATION

As we discussed in Section 2.3.2, when MBs are under the power of acoustic pressure, it could lead to their collapse or just their size oscillation. The need for inertial cavitation occurrence to provide thrombolysis is still not quite understood, since many studies point out the effectiveness of stable cavitation to recanalize the micro-circulation (Shi et al. 2010; Hu et al. 2018). Of course, many aspects of ultrasound use are essential to be considered to test such capability, including intensity range, number of cycles, operating frequency, and so on. Therefore, properly understanding and modeling both cavitation behaviors are important to classify the signals we are analyzing.

Recently, the International Electrotechnical Commission (IEC 2019)[1] reported a technical specification to measure cavitation noise in ultrasonic baths and ultrasonic reactors. In the document, the acoustic pressure spectrum (Figure 5.9) is available schematically, being divided into three types of frequency components: in blue, the largest peak (called direct), is the fundamental frequency (f_0); smaller peaks are described as stable cavitation (shown in yellow); and the broadband emissions are labeled as transient cavitation, i.e., inertial cavitation (red).

Figure 5.9 – Schematic representation of acoustic pressure spectrums: frequency components



Source: IEC 2019.

The distinction of frequency components present in Figure 5.9 is important to illustrate the shape of the acoustic pressure spectrum we are aiming to obtain for classification purposes, even though there is some lack of alignment between the IEC's representation and studies in the literature. The first conflict is the absence of stable cavitation at the f_0, (P. Wu, Lin, et al. 2018) found that the bubbles' oscillation that characterizes the stable cavitation also vibrates at f_0, and they emit the fundamental acoustic signal, which is not observed in Figure 5.9. A second distinction is related to the sub-harmonic oscillation, (Song et al. 2016) discovered that the sub-harmonic component can be related to the collapse of cavitation clouds during HIFU, and IEC's diagram does not represent that.

[1] An updated version is available (IEC 2024).

To summarize, it is important to acknowledge that the understanding of the discernment of which frequency component relates to which type of signal is still not fully understood, therefore, certain simplifications and compromises are necessary to advance the studies.

Thus, the start of the problem formulation was determined by values in the literature. Based on Chen et al. (2015), the stable cavitation is related to the ultra-harmonic waves, hence, it operates at a frequency that is between integer multiples of the f_0. Equation 5.7 describes the frequency for the n_{th} ultra-harmonic

$$f_n = \frac{2n+1}{2} \cdot f_0 \qquad n = 1, 2, ... \tag{5.7}$$

In contrast, since inertial cavitation occurs when MBs collapse violently, the literature correlates it to the fundamental concept of Kelvin impulse (J. R. Blake 1988; J. Blake et al. 2015), describing the behavior of a jet impact in consequence of a collapsing bubble. From there, we represent the response of the source as an impulse, and the component frequencies are spread across the frequency spectrum. To investigate the signal of inertial cavitation detection, we distinguish such frequency from harmonic and ultra-harmonic waves, which leads us to

$$f_n = \frac{2n+1}{4} \cdot f_0 \qquad n = 1, 2, ... \tag{5.8}$$

Besides those types of sources, if there is a lack of MBs in the focal spot, we will receive just an echo if the medium conditions enable it. To put it simply, the greater the impedance difference between two structures, the greater the intensity of reflection, and we have echoes with higher amplitude. To summarize, we explore four forms of classification, and they are condensed in Table 5.3.

It is important to highlight that the addition of noise to the generated signals is fundamental to approximating the simulations to the real scenario. Figure 5.9 illustrates that the transient cavitation would be hardly distinguished from any noise fluctuation, demonstrating that the addition of noise is essential to emulate the real signals. The appropriate undesired random disturbances motivated by electronics, or the environment, are added to the signal processing through the incorporation of a noise level.

Table 5.3 – Types of signals received by the receptor elements and frequency aspects

Type of source	Frequency feature	Occurrence
Echo	Signal similar to the excitation	Impedance difference between two media
Inertial Cavitation (IC)	Broadband emissions	Presence of MBs and enough energy to provoke the MBs collapse
Stable Cavitation (SC)	Ultra-harmonics	Presence of MBs and not enough energy to provoke the MBs collapse
Combination	Occurrence of 2 or 3 phenomena above	The pixel region has sub-regions with different phenomena occurrence

Source: Own work.

The noise is random, independent, with normal distribution and zero mean, and it follows the dynamic range of the signals, being set values of 2%, 5%, and a random noise (RN) level between 2 and 5%.

5.5 DETECTION: RECEIVERS' FEATURES

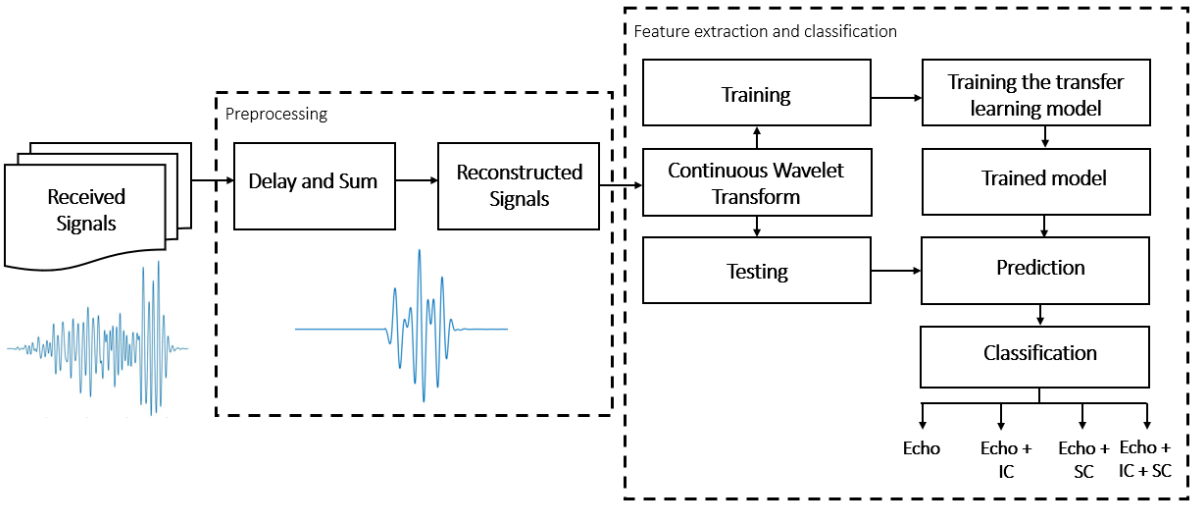
For detection purposes, the used ceramic piezoelectric receivers behave as bandpass filters with the central frequency corresponding to the resonant frequency and an imputed bandwidth (BW). Since the receivers attenuate some of the components of interest, ultra-harmonics for instance, the compensation for the signals is necessary for the detection. However, there are more expensive options using other types of piezoelectric transducers with broader BWs or even ceramics with different resonant frequencies.

Basically, the features presented in Table 5.2 are preserved with some adaptations for the reception attribute. For the narrowband receivers, the used frequency band is 100% at -6 dB (BW = 250 kHz), while for broadband receivers the frequency band is extensive, capturing and processing a wide spectrum of data. The examination of both scenarios is relevant considering that one of the inquiries presented in this work is if the same emission group of transducers could be applied for the detection operation.

5.6 GENERAL CLASSIFICATION PROCESS

To state the general steps of the process, figure 5.10 brings a flowchart of the procedure.

Figure 5.10 – Flowchart with the general steps of the process



Source: Own work.

5.6.1 SIGNALS' GENERATION

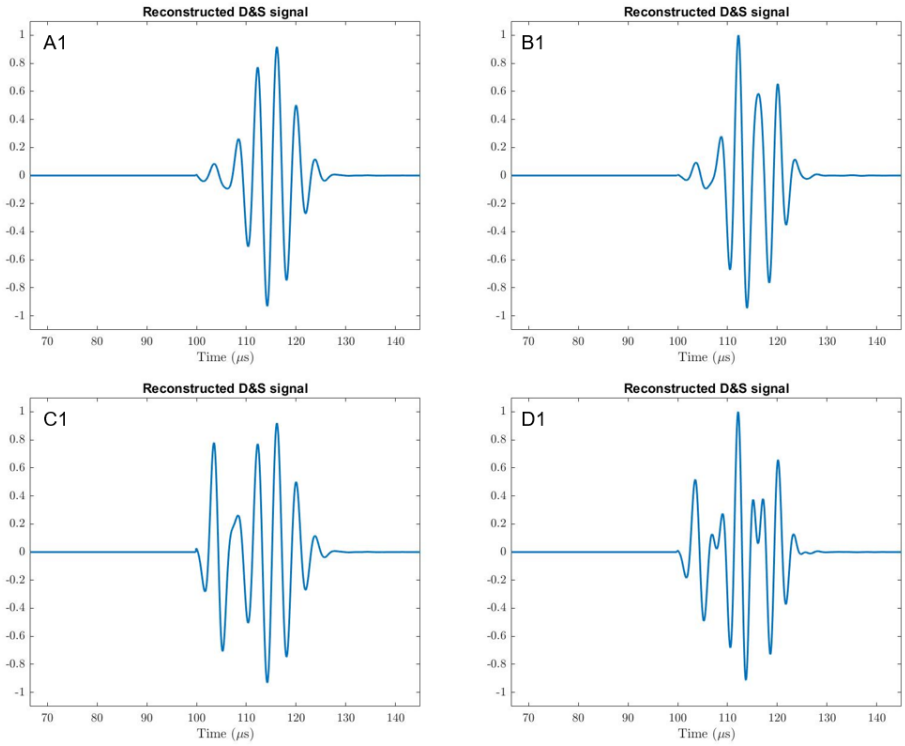
The first part of Figure 5.10 (signals' generation) was performed by Furuie et al. (2023). As we illustrated in Figure 5.7, considering a focal source, the temporal signal emitted is simulated as

$$\text{source}(t) = E_0 \times \text{echo}(t) + S_0 \times \text{SC}(t) + I_0 \times \text{IC}(t) + \text{noise}(t) \tag{5.9}$$

The amplitude factors E_0, S_0, and I_0 correspond to the echo, ultra-harmonic signal, and inertial signal, respectively. The assembling of a database with different signals is possible because we set several pairs (S_0, I_0) in which those coefficients are multipliers of the predefined 100 kPa amplitude for SC and IC, correspondingly. Thus, during the simulation of the signals, a couple of thresholds for the pairs were tested, and it was decided to follow values that had already shown consistent results (Furuie et al. 2023).

As shown in (Furuie et al. 2023), the criteria used to find the minimum S_0 and I_0 values were the search for estimations that yielded the maximum number of appropriate detections (true positives) and that respected the real scenario of occurrence of IC or SC. In short, it was established that the SC occurs for $S_0 \geqslant 0.035$ and the IC arises when $I_0 \geqslant 2.0$, while E_0 is always 1.

Figure 5.11 – Reconstructed signals after delay and sum (D&S). (A1) Example of an estimated temporal source signal with echo, and noise. (B1) Example of an estimated temporal source signal with echo, SC, and noise. (C1) Example of an estimated temporal source signal with echo, IC, and noise. (D1) Example of an estimated temporal source signal with echo, SC, IC, and noise



Source: Own work.

To exemplify the sources described in Table 5.3, Figure 5.11 illustrates samples of the 4 different sources previously indicated. After the signals' generation and processing, we enter the last box present in the flowchart, and that is discussed in the next section.

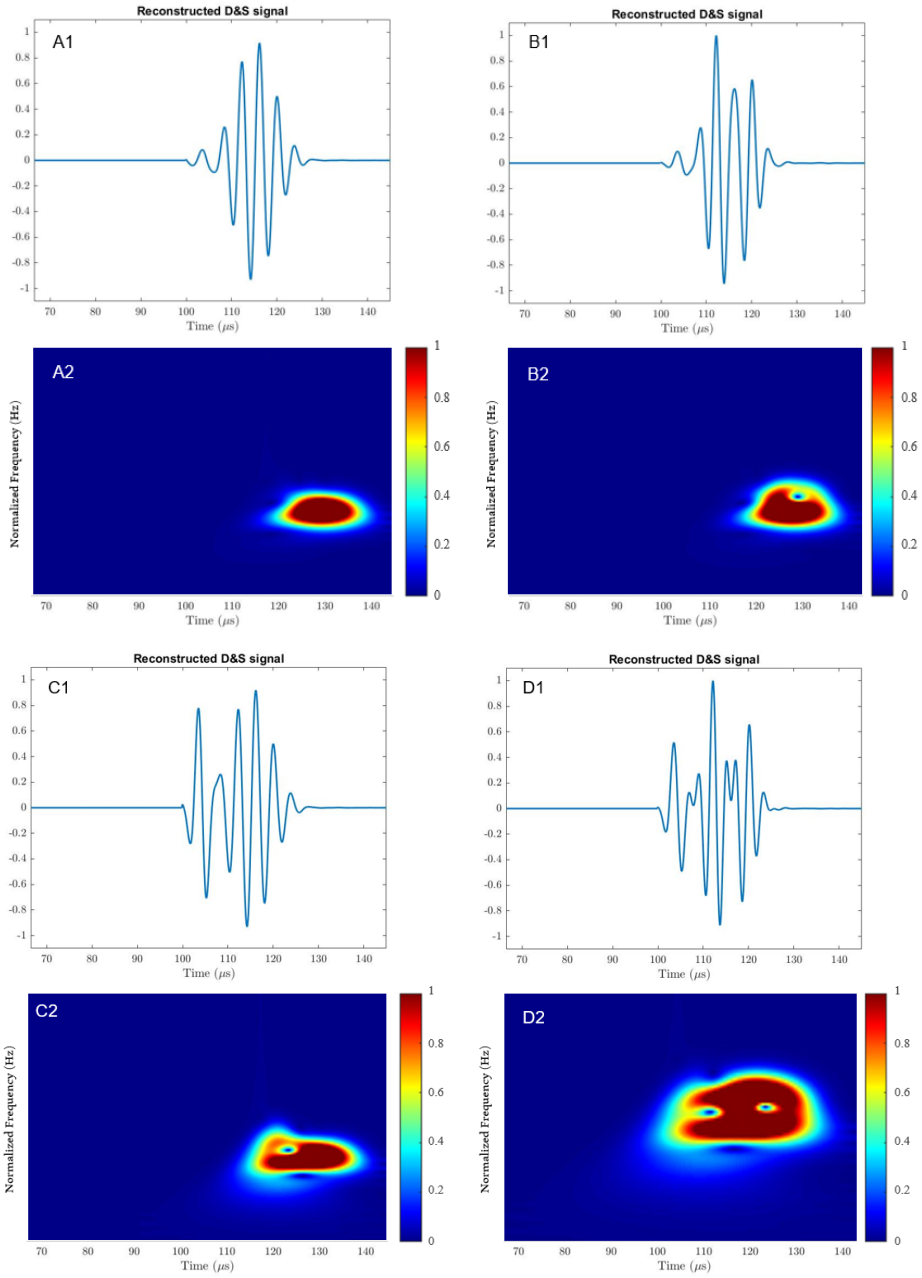
5.6.2 CWT APPLICATION AND RGB IMAGES

As shown in the procedure diagram (Figure 5.10), after creating the signals, we started the feature extraction step. For our problem, each signal comprises 2,257 temporal samples, and the data is converted into time-frequency spectrograms (CWT tool) in which no information overlaps, thus overfitting problems are avoided. The chosen wavelet mother $\Psi(t)$ was the Wavelet Morlet (Gabor) as a parameter of the CWT. As we mentioned in Section 4, the wavelet selection was based on the best correlation since the coefficients constitute the outcomes of a regression of the original signal performed on the wavelets.

After we obtain the absolute wavelet coefficient values from the signals through the CWT filterbank from MATLAB, we rescale those coefficients into the [0,1] interval. The application of the *rescale* function is necessary for the conversion of the coefficients into $8-bit$ unsigned integers so that we can transform the values to an RGB image using the selected colormap. Since our neural network is the AlexNet, the Jet colormap selected has 256 colors, i.e., compatible with AlexNet inputs. Figure 5.12 shows the time domain for signals with different features and the respective scalograms to exemplify the process.

As we mentioned in Section 4, the scales a from the CWT are used to create the wavelet bandpass filters, and depending on the chosen parameters as input for the CWT, those values change. They are discretized considering the number of wavelet filters per octave, which is 12 for this work, resulting in 141,000 coefficients for each signal. The larger the number of voices per octave, the finer the discretization of a, thus, the selection of the value considers the amount of computation required since it increases with its increment. The number of scales is calculated considering the energy spread of the wavelet in frequency and time (MATLAB 2016).

Figure 5.12 – Signals (A1), (B1), (C1), (D1), and respective scalograms for each signal: (A2) RGB image for signal (A1); (B2) RGB image for signal (B1); (C2) RGB image for signal (C1); (D2) RGB image for signal (D1)



Source: Own work.

5.6.3 DATASET

It is pertinent to clarify that there are four types of problems that we explored considering each with a database size of 2,800 waves. They are:

- Broadband receivers + RN level
- Narrowband receivers + RN level
- Narrowband receivers + 2% noise level
- Narrowband receivers + 5% noise level

Thus, the total number of synthetic signals inspected for the different scenarios is $4 \times 2,800 = 11,200$.

Table 5.4 shows a summary of the dataset features for our problem.

Table 5.4 – Dataset Summary

Attribute	Value
Total samples (each base)	2,800
Height × Width	227 × 227 px
Image size	≈ 3kb, 96 dpi
Total targets	4

Source: Own work.

The generation of each database took approximately 8 days. The measurements and simulations were done with Intel (R) Core (TM) i5-4460 CPU @ 3.20 GHz processor, and the system used is the NVIDIA GeForce RTX 970 graphical processing unit.

We randomly divided the RGB images into 3 new datasets following the proportion of 70% for training, 15% for validation, and 15% for testing purposes. The dataset placement for each case analyzed has a distribution according to Table 5.5.

5.6.4 MODEL IMPLEMENTATION

In this study, we took advantage of the transfer learning approach, where there is the reuse of a pre-trained model (AlexNet) on a new problem and no need to create a CNN model from scratch. To operate the model for our problem, the last three layers of AlexNet were replaced: Fully Connected, SoftMax, and Classification. Besides that, the hyperparameter selection is relevant since they are values that control the learning process of the model. Below, we have the definition of each one of them:

Table 5.5 – Dataset Distribution

Type	Quantity	Percentage
Training images	1,960	70%
Validation images	420	15%
Testing images	420	15%

Source: Own work.

- Batch size: the batch size defines the number of samples that will be propagated through the network at each iteration during the training.
- Learning rate (α): it determines the step size an algorithm updates or learns the values of a parameter estimate.
- Activation function: its main purpose consists of permitting the neural network to handle intricate tasks by performing non-linear computations.
- Epoch: it commands the number of complete passes through the training dataset.
- Optimizer: optimizers are algorithms or methods used to change the attributes of the neural network in order to reduce the losses. A common optimization technique is the Stochastic Gradient Descent (SGD), and when the runtime is comparatively high, momentum is suitable for quick learning, thus leading to faster converging. Then the optimizer is called Stochastic Gradient Descent with Momentum (SGDM).

Table 5.6 – List of hyperparameter values for the network model employed

Parameter	Selected value
Minibatch size	32
Activation	ReLU
Optimizer	SGDM
Learning rate	1^{-4}
Epochs	40
Loss function	Cross entropy

Source: Own work.

CNN is implemented using the DL Toolbox Model for AlexNet Network (MATLAB 2020). During the hyperparameter optimization, we tested different values of the attributes,

considering the computational cost as well, and the best results were obtained when the model was implemented with the values synthesized in Table 5.6. More details of the hyperparameters for a DL model and their influence in updating the network are presented in Appendix A.

6

ANALYSES AND RESULTS

6.1 METRICS

The evaluation of our method's performance is obtained by four metrics: accuracy, recall, precision, and F1-score are applied to analyze the results of transferred AlexNet. The equations are presented ahead:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{6.1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6.2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6.3}$$

where TP, TN, FP, and FN are the number of true positives, true negatives, false positives, and false negatives, respectively.

The accuracy has the advantage that it is very easily interpretable, but it is not robust if the data is unevenly distributed, or the existence of a particular type of error. To overcome possible issues, the F1-score is also inspected as a performance metric, and it is defined as

$$\text{F1-score} = 2\,\text{x}\,\frac{\text{Precision x Recall}}{\text{Precision + Recall}} \tag{6.4}$$

6.2 RESULTS ANALYSIS

Table 6.1 shows the results both considering narrow and broadband receivers. The broadband receivers yielded the best results (accuracy of 98.8%), which was expected among the cases proposed. However, the scenario applying narrowband receivers also produced acceptable detection indicators, with values of 96% for precision, 95.9% for F1-score, and 95.7% for accuracy for RN level. Those results suggest that it is possible to detect the phenomenon using the same set of ultrasound transducers or alternative broadband receivers.

Table 6.1 – Performance for the different sources, comparing narrow and broadband receivers, and distinct levels of noise (2%, 5%, and RN)

	Narrow Band			Broad Band
Measure	RN	2%	5%	RN
Accuracy	95.7 ± 1.0	97.1 ± 0.6	96.1 ± 0.4	98.8 ± 0.4
Precision	96.0 ± 1.1	97.3 ± 0.5	96.3 ± 0.9	98.7 ± 1.0
Recall	95.8 ± 1.1	97.3 ± 0.6	96.2 ± 0.6	98.8 ± 0.6
F1-score	95.9 ± 1.7	97.2 ± 0.7	96.2 ± 0.5	98.7 ± 0.5

Note: for each case, 5 simulations were conducted.

Source: Own work.

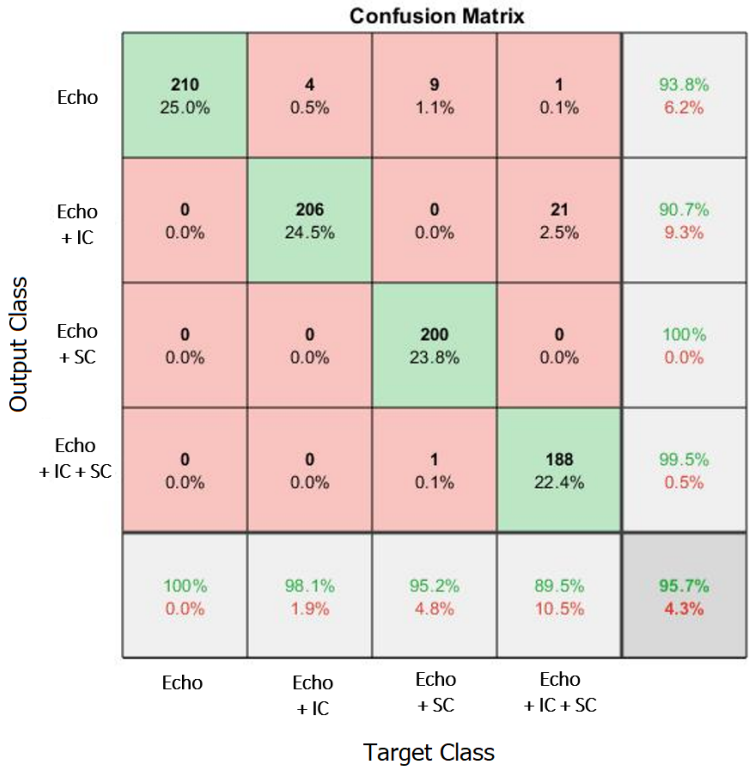
Furthermore, by comparing the narrowband results, it is possible to verify the influence of the noise level in the algorithm's evaluation. RN level seems to have a slight major impact on the disturbance in the classification, where the metrics exhibit less accurate results and are less consistent (higher variance).

A confusion matrix is also presented to assess the procedure's performance for the RN case (narrowband) (Figure 6.1). As we can observe, it referred to the 4 types of sources

from Table 5.3. The columns correspond to the target class (real classification), with the one on the far right of the matrix showing the percentages of the correctly (green) and incorrectly (red) predicted values. These metrics represent the precision and false discovery rate, respectively.

The rows meet the output obtained (method's classification), with the row at the bottom of the plot indicating the percentages that are correctly (green) and incorrectly (red) classified. The accurately categorized value is also called recall (or true positive rate), and the misclassified result is the false negative rate. The diagonal cells represent the correctly classified inspections and the off-diagonal values correspond to incorrectly classified observations. The last cell shows the approach's global accuracy acquired and it was 95.7%.

Figure 6.1 – Confusion Matrix for narrowband receiver case - RN rate (840 waves for testing and validation, 210 for each target class - 30% of the database)



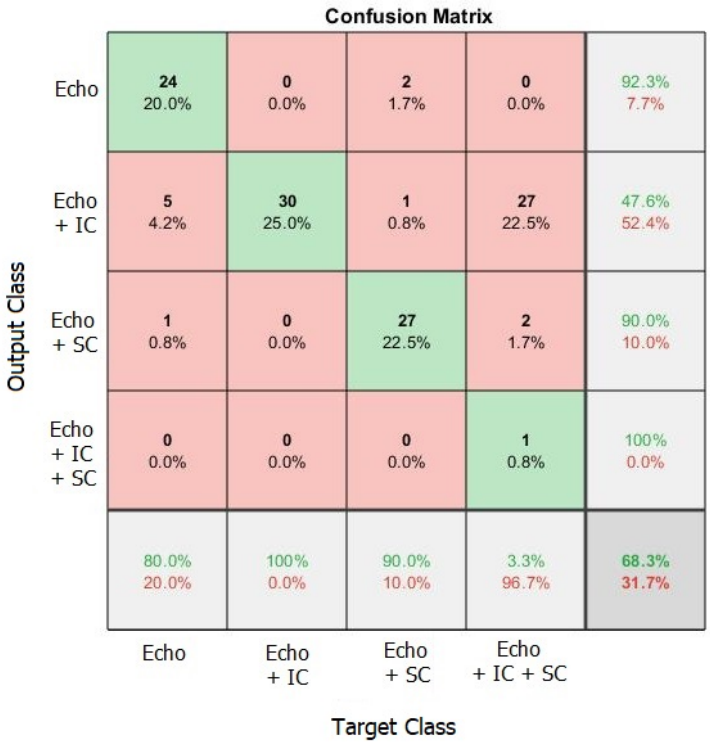
Source: Own work.

We can observe that the mislabeling concentration occurs between *Echo + IC* and *Echo + IC + SC* signals (cell(2,4)). Therefore, there are two important aspects to notice in this result: even confounding the two sources, the goal of identifying the IC was achieved,

and receiving such feedback, the medical professional would be able to avoid harm to the patient since this event is undesirable. However, the second point of attention is the failure to recognize the SC presence, which is the desired phenomenon for the sonothrombolysis technique.

Verifying the influence of the database size in the classification was also done. We tested a dataset with a smaller dimension as input of the CNN, and we attested a significant decrease in accuracy. For a database of 400 signals, for example, we obtained an accuracy of 68.3% for broadband receivers (Figure 6.2). It was noticed that the classification starts to produce appropriate results for databases with more than 1,000 samples.

Figure 6.2 – Confusion Matrix for broadband receiver case - RN rate (120 waves for testing and validation, 30 for each target class - 30% of the database).



Source: Own work.

For the case with fewer samples, the higher inaccuracy is clear. The *Echo + IC* class is mislabeled in almost 50% of the trials.

7

DISCUSSION

The experiments conducted across the present work revealed that

- The selection of the Mother Wavelet is important but not crucial to extracting the features of the signals.
  Even though the inspection and care for a wavelet that would show a better correlation with the signals was made, the application of a different wavelet did not disclose results with a considerable discrepancy. When the "Bump" wavelet was tested, for instance, the accuracy was still bearable ($\approx 93\%$).

- Analysis of different scenarios.
  Although it was not present in the results, scenarios with different thresholds for (S_0, I_0) were conducted, yielding acceptable results even though not as fine as the ones we showed for the cases analyzed.

- The project's limitations.
  Since AI models heavily rely on data for training and testing, a limitation that is often observed is the influence of the database size on classification. The results are directly affected by the number of samples available, which could be challenging if

the method is applied to experimental data signals. Furthermore, DL models are often considered black boxes because the decision-making processes of the system are not transparent or easily understandable. With millions of parameters organized in intricate architectures, understanding how input data transitions into output predictions is not trivial.

- Presentation of results.
  Through the development of the laboratory projects, we presented and published our results at prestigious conferences such as the Congresso Brasileiro de Engenharia Biomédica (CBEB) (Guenkawa, Furuie, and Caurin 2022), the Computing in Cardiology (CINC) Conference (Guenkawa and Furuie 2023), and the Canadian Conference on Electrical and Computer Engineering (IEEE CCECE) (Furuie et al. 2023). It is anticipated that this work will be submitted for publication in a Journal as well.

Given the results obtained in this project and the appositeness of the topic, we propose other research lines as follows

- Explore other Machine Learning algorithms that apply transfer learning.
  The implementation of other transfer learning approaches would enrich the work so the performance metrics could be compared. Some other architectures are LeNet (Lecun et al. 1998), VGGNet (Simonyan and Zisserman 2014), or ResNet (He et al. 2015).

- Insert experimental signals in the database.
  Since experiments are being conducted in the group's laboratory, the insertion of tested signals with the reevaluation of the algorithm would reinforce the suitability of the method implemented if good results were obtained.

- Explore Deep Learning algorithms often applied to medical problems.
  U-Net is a CNN that was developed initially for biomedical image segmentation and cell tracking (Ronneberger et al. 2015). But in recent years, its applications have expanded, meeting the signal processing field as well. Speech enhancement and ECG analysis are two examples of the applicabilities for signal analysis, and thus it might be a solution to be implemented for our problem.

8

CONCLUSION

Considering the relevance of the study when we highlight CVDs conditions, accounting for 17.9 million deaths each year globally (WHO 2021), and the potential and innovation involving the sonothrombolysis technique, it is important to emphasize the significance of favorable outcomes for studies related to the therapy. Besides that, since the clot lysis mechanism is still not fully understood, the detection of the type of cavitation that is occurring and its location is crucial for the methods safety and efficacy.

The main goal of this work was to evaluate the possibility of applying an AI approach to detect and classify the cavitation phenomenon during therapy. The tested technique was a DL model (AlexNet), carried with transfer learning, and the CWT tool to extract information from the signals. The choice considered the computational cost of the method, the presence of satisfactory results for similar applications, and the particular interest of the authors in CNN models.

The results revealed that the automatic and uncomplicated classifier method proposed reached an acceptable performance and in fact, it also showed the possibility of applying the same set of transducers as emitters and receivers, since the narrowband device

simulation yielded appropriate outcomes as well. For the case of narrowband receivers, the study's result indicated that the technique achieved state-of-the-art values of around 95.7%, and 96% for accuracy and precision, respectively. The considerable degree of accuracy demonstrated that using AI could be an approach to explore the detection of cavitation for therapies using ultrasound signals.

We also learned that the number of samples is essential to achieve a considerable degree of performance. We noticed that databases with a low quantity of samples (400 signals) did not offer tolerable results (68.3% of accuracy), while acceptable outcomes started to show for bases greater than 1,000 waves (accuracy $> 70\%$). Furthermore, AlexNet showed substantial learning potential for the metrics explored, especially considering the few epochs applied (40). The dispensable high number of epochs is due to the use of the transfer learning technique, revealing it to be a fast approach. Since the model is already trained, we adapt the algorithm to our problem by replacing the three last layers, so the learning process becomes fast and efficient. In short, the duration of each round for the training, testing, and validating varied approximately from 20 to 40 minutes.

Besides that, due to the fact the cavitation signals are structured of different frequency components and are non-stationary, the CWT tool was demonstrated to be a suitable feature extractor instrument for the ultrasound signals we inspected. Through the analysis in the time-frequency domain and the use of the calculated coefficients, we constructed the RGB images and fed our neural network. By direct observation, the images do not show an evident dissimilarity, demonstrating the great potential of AI approaches for characteristic recognition.

REFERENCES

123RF. *Stock photos, vectors, and royalty.* 2022. URL: https://www.123rf.com/.

Anderson, Jeffrey and David Morrow. "Acute Myocardial Infarction". In: *New England Journal of Medicine* 376 (May 2017), pp. 2053–2064. DOI: 10.1056/NEJMra1606915.

Apfel, Robert E. "Acoustic cavitation". In: *Methods in experimental physics* (1981), pp. 355–411.

Arbib, Michael. *The Handbook of Brain Theory and Neural Networks.* Mar. 2023.

Bader, Kenneth, Guillaume Bouchoux, and Christy Holland. "Sonothrombolysis". In: vol. 880. Jan. 2016, pp. 339–362. ISBN: 978-3-319-22535-7. DOI: 10.1007/978-3-319-22536-4_19.

Bagai, Akshay et al. "Reperfusion Strategies in Acute Coronary Syndromes". In: *Circulation research* 114 (June 2014), pp. 1918–28. DOI: 10.1161/CIRCRESAHA.114.302744.

Bajaj, Nikesh. "Wavelets for EEG Analysis". In: Nov. 2020. ISBN: 978-1-83881-947-7. DOI: 10.5772/intechopen.94398.

Basu, Kanadpriya et al. "Artificial Intelligence: How is It Changing Medical Sciences and Its Future?" In: *Indian Journal of Dermatology* 65 (May 2020), p. 365. DOI: 10.4103/ijd.IJD_421_20.

Bjerknes, Vilhelm. "Fields of force". In: *Columbia University Press* (1906).

Blake, John R. "The Kelvin Impulse: Application to cavitation bubble dynamics". In: *The Journal of the Australian Mathematical Society. Series B. Applied Mathematics* 30 (Oct. 1988), pp. 127–146. DOI: 10.1017/S0334270000006111.

Blake, John, David Leppinen, and Qianxi Wang. "Cavitation and bubble dynamics: The Kelvin impulse and its applications". In: *Interface Focus* 5 (Oct. 2015), p. 20150017. DOI: 10.1098/rsfs.2015.0017.

Boersma, Eric et al. “Early thrombolytic treatment in acute myocardial infarction: Reappraisal of the golden hour”. In: *Lancet* 348 (Oct. 1996), pp. 771–5. DOI: 10.1016/S0140-6736(96)02514-7.

Brennen, Christopher. “Cavitation and Bubble Dynamics”. In: *Cavitation and Bubble Dynamics* 44 (Jan. 1995). DOI: 10.1017/CBO9781107338760.

Brown, Aliza et al. “Microbubbles Improve Sonothrombolysis In Vitro and Decrease Hemorrhage In Vivo in a Rabbit Stroke Model”. In: *Investigative radiology* 46 (Mar. 2011), pp. 202–7. DOI: 10.1097/RLI.0b013e318200757a.

Califf, Robert et al. “Clinical risks of thrombolytic therapy”. In: *The American journal of cardiology* 69 (Feb. 1992), 12A–20A. DOI: 10.1016/0002-9149(92)91168-4.

Cannon, Christopher, Carolyn McCabe, and Eugene Braunwald. “The Thrombolysis in Myocardial Infarction (TIMI) Trial”. In: Jan. 1994, pp. 53–67. ISBN: 978-1-4613-6122-0. DOI: 10.1007/978-1-4615-2618-6_5.

Cauchy, Augustin-Louis. “Méthode génerale pour la résolution des dystèmes d’équations simultanées”. In: *C. R. Acad. Sci* 25:536538 (1847).

Chen, Xucai et al. “Dynamic Behavior of Microbubbles during Long Ultrasound Tone-Burst Excitation: Mechanistic Insights into Ultrasound-Microbubble Mediated Therapeutics Using High-Speed Imaging and Cavitation Detection”. In: *Ultrasound in Medicine & Biology* 42 (Nov. 2015). DOI: 10.1016/j.ultrasmedbio.2015.09.017.

Chollet, Francois. *Deep learning with python*. Manning Publications, 2017. ISBN: 9781617294433.

Clark, Alicia and Alberto Aliseda. “Video: Bjerknes Forces Acting on Ultrasound Contrast Agents”. In: Nov. 2016. DOI: 10.1103/APS.DFD.2016.GFM.V0066.

Crouzet, Sebastien et al. “Whole-gland Ablation of Localized Prostate Cancer with High-intensity Focused Ultrasound: Oncologic Outcomes and Morbidity in 1002 Patients”. In: *European urology* 65 (Apr. 2013). DOI: 10.1016/j.eururo.2013.04.039.

Culjat, Martin, David Goldenberg, and Priyamvada Tewari. “A Review of Tissue Substitutes for Ultrasound Imaging”. In: *Ultrasound in medicine & biology* 36 (June 2010), pp. 861–73. DOI: 10.1016/j.ultrasmedbio.2010.02.012.

Dalecki, Diane. “Mechanical Bioeffects of Ultrasound”. In: *Annual review of biomedical engineering* 6 (Feb. 2004), pp. 229–48. DOI: 10.1146/annurev.bioeng.6.040803.140126.

Disease Control, Centers for and Prevention. *What is Venous Thromboembolism?* Feb. 7, 2020. URL: https://www.cdc.gov/ncbddd/dvt/facts.html.

Dubinsky, Theodore et al. “High-Intensity Focused Ultrasound: Current Potential and Oncologic Applications”. In: *AJR. American journal of roentgenology* 190 (Feb. 2008), pp. 191–9. DOI: 10.2214/AJR.07.2671.

Dutta, Nabanita, Umashankar Subramaniam, et al. “Comparative Study of Cavitation Problem Detection in Pumping System Using SVM and K-Nearest Neighbour Method”. In: June 2020. DOI: 10.1109/EEEIC/ICPSEurope49358.2020.9160689.

Dutta, Nabanita, Arun Shankar Vishnu, et al. “Centrifugal Pump Cavitation Detection Using Machine Learning Algorithm Technique”. In: June 2018. DOI: 10.1109/EEEIC.2018.8494594.

Ellens, Nicholas and K. Hynynen. “High-intensity focused ultrasound for medical therapy”. In: Dec. 2015, pp. 661–693. ISBN: 9781782420286. DOI: 10.1016/B978-1-78242-028-6.00022-3.

EWI. *Inspecting Rails using Matrix Phased Array*. 2010. URL: https://ewi.org/inspecting-rails-using-matrix-phased-array/.

Furie, Bruce and Barbara Furie. “Mechanisms of Thrombus Formation”. In: *The New England journal of medicine* 359 (Sept. 2008), pp. 938–49. DOI: 10.1056/NEJMra0801082.

Furuie, Sérgio et al. “Investigation of Cavitation Detector for Sonothrombolysis”. In: Sept. 2023, pp. 83–87. DOI: 10.1109/CCECE58730.2023.10288666.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep Learning*. The MIT Press, Nov. 18, 2016.

Gray, Henry. “Grays Anatomy, 39th Edition: The Anatomical Basis of Clinical Practice”. In: 2005, pp. 997–1003.

Gregg, David and Pascal Goldschmidt-clermont. “Platelets and Cardiovascular Disease”. In: *Circulation* 108 (Oct. 2003), e88–90. DOI: 10.1161/01.CIR.0000086897.15588.4B.

Guenkawa, Patricia and Sérgio Furuie. “An Analysis of Cavitation in Sonothrombolysis through Convolutional Neural Networks”. In: *2023 Computing in Cardiology (CinC)*. Vol. 50. 2023, pp. 1–4. DOI: 10.22489/CinC.2023.305.

Guenkawa, Patricia, Sérgio Furuie, and Glauco Caurin. “Classification of Cardiac Arrhythmias by Images and Neural Network Model”. In: (2022).

Haworth, Kevin, Kenneth Bader, et al. “Quantitative Frequency-Domain Passive Cavitation Imaging”. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* PP (Oct. 2016), pp. 1–1. DOI: 10.1109/TUFFC.2016.2620492.

Haworth, Kevin, Vasant Salgaonkar, et al. “Using Passive Cavitation Images to Classify High-Intensity Focused Ultrasound Lesions”. In: *Ultrasound in Medicine & Biology* 41 (June 2015). DOI: 10.1016/j.ultrasmedbio.2015.04.025.

Haykin, Simon. *Neural networks and learning machines*. 2009.

He, Kaiming et al. “Deep Residual Learning for Image Recognition”. In: 7 (Dec. 2015).

Herrick, JB. “Clinical features of sudden obstruction of the coronary arteries”. In: *JAMA: the journal of the American Medical Association* LIX (Dec. 1912). DOI: 10.1001/jama.1912.04270120001001.

Hu, Bo et al. “Stable cavitation for efficient thrombolysis in coronary micro-circulation: an in-vitro experimental study with dodecafluropentane acoustic phase-chance nanodroplets”. In: *Journal of the American College of Cardiology* 71 (Mar. 2018), A268. DOI: 10.1016/S0735-1097(18)30809-X.

Huang, Jie et al. “Sign Language Recognition using 3D convolutional neural networks”. In: *2015 IEEE International Conference on Multimedia and Expo (ICME)*. 2015, pp. 1–6. DOI: 10.1109/ICME.2015.7177428.

IEC. *Measurement of cavitation noise in ultrasonic baths and ultrasonic reactors*. Sept. 2019.

— *Measurement of cavitation noise in ultrasonic baths and ultrasonic reactors*. Aug. 2024.

Izadifar, Zahra, Paul Babyn, and Dean Chapman. “Ultrasound Cavitation/Microbubble Detection and Medical Applications”. In: *Journal of Medical and Biological Engineering* 39 (Apr. 2018). DOI: 10.1007/s40846-018-0391-0.

Jolesz, Ferenc. “MRI-Guided Focused Ultrasound Surgery”. In: *Annual review of medicine* 60 (Feb. 2009), pp. 417–30. DOI: 10.1146/annurev.med.60.041707.170303.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Neural Information Processing Systems* 25 (Jan. 2012). DOI: 10.1145/3065386.

Lecun, Yann et al. “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324. DOI: 10.1109/5.726791.

Lee, Hohyeon et al. “Microbubbles used for contrast enhanced ultrasound and theragnosis: a review of principles to applications”. In: *Biomedical Engineering Letters* (Feb. 2017). DOI: 10.1007/s13534-017-0016-5.

Lee, Sunwoo et al. “Parallel Deep Convolutional Neural Network Training by Exploiting the Overlapping of Computation and Communication”. In: *2017 IEEE 24th International Conference on High Performance Computing (HiPC)*. 2017, pp. 183–192. DOI: 10.1109/HiPC.2017.00030.

Li, Fei-Fei, Yunzhu Li, and Ruohan Gao. “CS231n: Deep Learning for Computer Vision”. 2023. URL: http://cs231n.stanford.edu/.

Liao, Ai-Ho et al. “Ultrasound-induced microbubble cavitation via a transcanal or transcranial approach facilitates inner ear drug delivery”. In: *JCI Insight* 5 (Jan. 2020). DOI: 10.1172/jci.insight.132880.

Libby, Peter. “Current Concepts of the Pathogenesis of the Acute Coronary Syndromes”. In: *Circulation* 104 (Aug. 2001), pp. 365–72. DOI: 10.1161/01.CIR.104.3.365.

Libby, Peter et al. “Atherosclerosis”. In: *Nat Rev Dis Primers* 359 (Aug. 2019). DOI: 10.1038/s41572-019-0106-z.

Liu, Hui and Zaixing Jiang. “Wavelet transform of geophysical well logging signal and its application to sequence division”. In: *Proceedings - 2010 3rd International Congress on Image and Signal Processing, CISP 2010* 7 (Oct. 2010). DOI: 10.1109/CISP.2010.5646751.

Long, Jonathan, Evan Shelhamer, and Trevor Darrell. “Fully convolutional networks for semantic segmentation”. In: June 2015, pp. 3431–3440. DOI: 10.1109/CVPR.2015.7298965.

Lyaker, MR et al. “Arterial embolism”. In: *International journal of critical illness and injury science* 03 (Jan. 2013), pp. 77–87. DOI: 10.4103/2229-5151.109429.

Mallat, Stphane. *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*. 3rd. USA: Academic Press, Inc., 2008. ISBN: 0123743702.

Mashrur, Fazla Rabbi, Amit Roy, and Dabasish Saha. “Automatic Identification of Arrhythmia from ECG Using AlexNet Convolutional Neural Network”. In: Dec. 2019, pp. 1–5. DOI: 10.1109/EICT48899.2019.9068806.

Mast, T Douglas. “Empirical relationships between acoustic parameters in human soft tissues”. In: *Acoustics Research Letters Online* 1 (Nov. 2000). DOI: 10.1121/1.1336896.

MATLAB. *Continuous 1-D wavelet transform*. 2016.

— *Deep Learning Toolbox Model for AlexNet Network*. 2020.

Maxwell, Adam et al. "Non-Invasive Thrombolysis Using Pulsed Ultrasound Cavitation Therapy Histotripsy". In: *Ultrasound in medicine & biology* 35 (Oct. 2009), pp. 1982–94. DOI: 10.1016/j.ultrasmedbio.2009.07.001.

McCarthy, John. "Artificial Intelligence, Logic, and Formalising Common Sense". In: Jan. 2022, pp. 69–90. DOI: 10.1002/9781119815075.ch6.

Medel, Ricky et al. "Sonothrombolysis: An emerging modality for the management of stroke". In: *Neurosurgery* 65 (Nov. 2009), 979–93, discussion 993. DOI: 10.1227/01.NEU.0000350226.30382.98.

Morrow, D and E Braunwald. "Classification and Diagnosis of Acute Coronary Syndromes". In: Jan. 2016, pp. 01–10. ISBN: 9780323359436.

Natrajan, Priyanka et al. "A Transfer Learning based CNN approach for Classification of Horticulture plantations using Hyperspectral Images". In: Dec. 2018, pp. 279–283. DOI: 10.1109/IADCC.2018.8692142.

NHLBI. *Percutaneous Coronary Intervention*. Dec. 31, 2021. URL: https://www.nhlbi.nih.gov/health-topics/percutaneous-coronary-intervention#:~:text=These%5C%20complications%5C%20may%5C%20include%5C%20bleeding,%5C%2C%5C%20stroke%5C%2C%5C%20or%5C%20blood%5C%20clots..

Noltingk, Bernard and E Neppiras. "Cavitation Produced by Ultrasonics". In: *Proceedings of the Physical Society. Section B* 63 (Dec. 1950), p. 674. DOI: 10.1088/0370-1301/63/9/305.

Patterson, Josh and Adam Gibson. *Deep Learning: A Practitioner's Approach*. O'Reilly Media, 2017.

Peng Xand Zhou, J and X Wu. "New Strategies for Myocardial Infarction Treatment: Peng X et al. Novel Treatments for Myocardial Infarction". In: *Journal of Cardiology and Therapy* 4 (Jan. 2017), pp. 664–670. DOI: 10.17554/j.issn.2309-6861.2017.04.127.

Plesset, Milton S. "The dynamics of cavitation bubbles". In: *Trans. ASME J. appl. Phys.* 25 (Jan. 1949), pp. 228–231.

Poritsky, Hillel and Gabriel Horvay. "Closure to Discussion of Stresses in Pipe Bundles (1952, ASME J. Appl. Mech., 19, p. 229)". In: *Journal of Applied Mechanics* 19 (June 1952), pp. 229–230. DOI: 10.1115/1.4010455.

Porter, Thomas and Feng Xie. “Ultrasound, microbubbles, and thrombolysis”. In: *Progress in Cardiovascular Diseases* 44 (Sept. 2001), pp. 101–110. DOI: 10.1053/pcad.2001.26441.

Putterman, Seth et al. “Cavitation science: Is there a simple theory of sonoluminescence?” In: *Nature* 409 (Feb. 2001), pp. 782–783. DOI: 10.1038/35057317.

Quaia, Emilio. “Classification and Safety of Microbubble-Based Contrast Agents”. In: Jan. 2005, pp. 3–14. ISBN: 3-540-40740-5. DOI: 10.1007/3-540-27214-3_1.

Rahko, Peter. “Evaluation of the Skin-To-Heart Distance in the Standing Adult by Two-Dimensional Echocardiography”. In: *Journal of the American Society of Echocardiography : official publication of the American Society of Echocardiography* 21 (June 2008), pp. 761–4. DOI: 10.1016/j.echo.2007.10.027.

Rayleigh, Lord. “On The Pressure Developed in a Liquid During the Collapse of a Spherical”. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, VI. Series* 34 (Jan. 1917). DOI: 10.1080/14786440808635681.

Rezkalla, Shereif and Robert Kloner. “No-Reflow Phenomenon”. In: *Circulation* 105 (Mar. 2002), pp. 656–62. DOI: 10.1161/hc0502.102867.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: (May 2015).

Saalbach, Kai-Alexander, Jens Twiefel, and Joerg Wallaschek. “Self-Sensing Cavitation Detection in Ultrasound-Induced Acoustic Cavitation”. In: *Ultrasonics* 94 (June 2018). DOI: 10.1016/j.ultras.2018.06.016.

Sabatine, Marc et al. “Addition of Clopidogrel to Aspirin and Fibrinolytic Therapy for Myocardial Infarction with ST-Segment Elevation”. In: *The New England journal of medicine* 352 (Mar. 2005), pp. 1179–89. DOI: 10.1056/NEJMoa050522.

Saeed, Mehreen. *A Gentle Introduction To Gradient Descent Procedure*. July 28, 2021. URL: https://machinelearningmastery.com/a-gentle-introduction-to-gradient-descent-procedure/.

Saint Victor, Marie de et al. “Properties, characteristics and applications of microbubbles for sonothrombolysis”. In: *Expert opinion on drug delivery* 11 (Jan. 2014). DOI: 10.1517/17425247.2014.868434.

Schleicher, Nadine et al. “Sonothrombolysis with BR38 Microbubbles Improves Microvascular Patency in a Rat Model of Stroke”. In: *PloS one* 11 (Apr. 2016), e0152898. DOI: 10.1371/journal.pone.0152898.

Sha, Yu et al. “A multi-task learning for cavitation detection and cavitation intensity recognition of valve acoustic signals”. In: *Engineering Applications of Artificial Intelligence* 113 (2022), p. 104904. ISSN: 0952-1976. DOI: https://doi.org/10.1016/j.engappai.2022.104904. URL: https://www.sciencedirect.com/science/article/pii/S0952197622001361.

Sharma, Deepak Kumar et al. “3 - Deep learning applications for disease diagnosis”. In: *Deep Learning for Medical Applications with Unique Data*. Ed. by Deepak Gupta et al. Academic Press, 2022, pp. 31–51. ISBN: 978-0-12-824145-5. DOI: https://doi.org/10.1016/B978-0-12-824145-5.00005-8. URL: https://www.sciencedirect.com/science/article/pii/B9780128241455000058.

Sharma, Vidish. *Machine Learning*. Dec. 18, 2018. URL: https://medium.com/@vidishsharma1311/machine-learning-70f6f15d8b3d.

Shen, Xuedong et al. “Therapeutic Ultrasound-Enhanced Thrombolysis in Patients With Acute Myocardial Infarction”. In: *Angiology* 61 (Nov. 2009), pp. 253–8. DOI: 10.1177/0003319709343287.

Shi, William et al. “Investigation of effectiveness of microbubble stable cavitation in thrombolysis”. In: *Proceedings - IEEE Ultrasonics Symposium* (Oct. 2010), pp. 330–333. DOI: 10.1109/ULTSYM.2010.5935628.

Shoeb, Ali and Gari Clifford. “Wavelets; Multiscale activity in physiological signals short-time fourier transform”. In: *Biomedical signal and image processing* (2006), pp. 1–29.

Shpak, Oleksandr et al. “Droplets, Bubbles and Ultrasound Interactions”. In: vol. 880. Jan. 2016, pp. 157–174. ISBN: 978-3-319-22535-7. DOI: 10.1007/978-3-319-22536-4_9.

Silva, Wilton and Sérgio Furuie. “Ultrasound Coupled Radial Vibration Mode: Influence on Cardiac Sonothrombolysis”. In: Jan. 2024, pp. 523–533. ISBN: 978-3-031-49403-1. DOI: 10.1007/978-3-031-49404-8_50.

Simonyan, Karen and Andrew Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition”. In: *arXiv 1409.1556* (Sept. 2014).

Song, Jae, Kristoffer Johansen, and Paul Prentice. “An analysis of the acoustic cavitation noise spectrum: The role of periodic shock waves”. In: *The Journal of the Acoustical Society of America* 140 (Oct. 2016), pp. 2494–2505. DOI: 10.1121/1.4964633.

Souza, Vitoria and Sérgio Furuie. “Spectral Profiles of Sonothrombolysis Bubble Radiation”. In: Nov. 2023. DOI: 10.22489/CinC.2023.182.

Sun-young, Kim et al. “Uterine Fibroids: Postsonication Temperature Decay Rate Enables Prediction of Therapeutic Responses to MR Imaging-guided High-Intensity Focused Ultrasound Ablation”. In: *Radiology* 270 (Sept. 2013). DOI: 10.1148/radiol.13130380.

Szabo, Thomas. “Diagnostic Ultrasound ImagingInside Out”. In: (Sept. 2004).

Tachibana, Katsuro. “Enhancement of Fibrinolysis with Ultrasound Energy”. In: *Journal of vascular and interventional radiology : JVIR* 3 (June 1992), pp. 299–303. DOI: 10.1016/S1051-0443(92)72029-6.

Tanikic, Dejan and Vladimir Despotovic. “Artificial Intelligence Techniques for Modelling of Temperature in the Metal Cutting Process”. In: Sept. 2012. ISBN: 978-953-51-0736-1. DOI: 10.5772/47850.

Taye, Mohammad Mustafa. “Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions”. In: *Computation* 11.3 (2023). ISSN: 2079-3197. DOI: 10.3390/computation11030052. URL: https://www.mdpi.com/2079-3197/11/3/52.

Thygesen, Kristian et al. “Third Universal Definition of Myocardial Infarction”. In: *European heart journal* 33 (Oct. 2012), pp. 2551–2567. DOI: 10.1093/eurheartj/ehs184.

Tiwari, Rajiv, D.J. Bordoloi, and Aakash Dewangan. “Blockage and cavitation detection in centrifugal pumps from dynamic pressure signal using deep learning algorithm”. In: *Measurement* 173 (2021), p. 108676. ISSN: 0263-2241. DOI: https://doi.org/10.1016/j.measurement.2020.108676. URL: https://www.sciencedirect.com/science/article/pii/S0263224120311878.

Treeby, Bradley, Ben Cox, and Jiri Jaros. *k-Wave: A MATLAB toolbox for the time-domain simulation of acoustic wave fields*. 2020.

Trübestein, Gustav et al. “Thrombolysis by Ultrasound”. In: *Clinical science and molecular medicine. Supplement* 3 (Jan. 1977), 697s–698s. DOI: 10.1042/cs051697s.

Uijlings, Jasper et al. “Selective Search for Object Recognition”. In: *International Journal of Computer Vision* 104 (Sept. 2013), pp. 154–171. DOI: 10.1007/s11263-013-0620-5.

Vaillant, Régis, Christophe Monrocq, and Yann Lecun. “An Original Approach for the Localization of Objects in Images”. In: vol. 141. June 1993, pp. 26–30. ISBN: 0-85296-573-7.

Venkatesan, Ragav and Baoxin Li. *Convolutional Neural Networks in Visual Computing: A Concise Guide*. Oct. 2017, p. 187. ISBN: 9781315154282. DOI: 10.4324/9781315154282.

Vrana, Johannes and Ripudaman Singh. "The NDE 4.0: Key Challenges, Use Cases, and Adaption". In: (Mar. 2020).

Wang, Tao et al. "Automatic ECG Classification Using Continuous Wavelet Transform and Convolutional Neural Network". In: *Entropy* 23 (Jan. 2021), p. 119. DOI: 10.3390/e23010119.

Wang, Tzu-Yin et al. "Ultrasound and Microbubble Guided Drug Delivery: Mechanistic Understanding and Clinical Implications". In: *Current pharmaceutical biotechnology* 14 (Dec. 2013). DOI: 10.2174/13892010146661312261146 11.

White, Harvey and Derek Chew. "Acute myocardial infarction". In: *Lancet* 372 (Sept. 2008), pp. 570–84. DOI: 10.1016/S0140-6736(08)61237-4.

WHO. *Cardiovascular diseases (CVDs)*. June 21, 2021. URL: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

Wood, Paul. "Acute and Subacute Coronary Insufficiency". In: *British medical journal* 1 (July 1961), pp. 1779–82. DOI: 10.1136/bmj.1.5242.1779.

Wu, Jun and Ren-Ke Li. "Ultrasound-targeted microbubble destruction in gene therapy: A new tool to cure human diseases". In: *Genes & Diseases* 4 (Aug. 2016). DOI: 10.1016/j.gendis.2016.08.001.

Wu, Pengfei, Weijun Lin, and Xiuming Wang. "Mechanism and dynamics of hydrodynamic-acoustic cavitation (HAC)". In: *Ultrasonics Sonochemistry* 49 (July 2018). DOI: 10.1016/j.ultsonch.2018.07.021.

Wu, Pengfei, Xiuming Wang, and Weijun Lin. "Acoustic characterization of cavitation intensity: A review". In: *Ultrasonics Sonochemistry* 82 (Dec. 2021), p. 105878. DOI: 10.1016/j.ultsonch.2021.105878.

Xie, Feng et al. "Diagnostic Ultrasound Induced Inertial Cavitation to Non-Invasively Restore Coronary and Microvascular Flow in Acute Myocardial Infarction". In: *PloS one* 8 (July 2013), e69780. DOI: 10.1371/journal.pone.0069780.

Xu, Zhen. *Histotripsy for Treatment of Cancer and Neurological Diseases*. Focused Ultrasound Foundation. Mar. 25, 2020. URL: https://www.youtube.com/watch?v=y-TiigoPpzk&ab_channel=FocusedUltrasoundFoundation.

Yang, Bo-Suk et al. “Cavitation detection of butterfly valve using support vector machines”. In: *Journal of Sound and Vibration* 287 (Oct. 2005), pp. 25–43. DOI: 10.1016/j.jsv.2004.10.033.

Yang, Yaoheng et al. “Cavitation dose painting for focused ultrasound-induced blood-brain barrier disruption”. In: *Scientific Reports* 9 (Feb. 2019). DOI: 10.1038/s41598-019-39090-9.

Zafari, A Maziar and Mahmoud H Abdou. *Myocardial Infarction*. May 7, 2019. URL: https://emedicine.medscape.com/article/155919-overview#a6.

Zhang, Xv et al. “Survey of Convolutional Neural Network”. In: *Proceedings of the 2018 International Conference on Network, Communication, Computer Engineering (NCCE 2018)*. Atlantis Press, 2018, pp. 93–97. ISBN: 978-94-6252-517-7. DOI: 10.2991/ncce-18.2018.16. URL: https://doi.org/10.2991/ncce-18.2018.16.

Zhu, Michael et al. “Primary stent implantation compared with primary balloon angioplasty for acute myocardial infarction: A meta-analysis of randomized clinical trials”. In: *The American journal of cardiology* 88 (Sept. 2001), pp. 297–301. DOI: 10.1016/S0002-9149(01)01645-9.

A

DEEP LEARNING

In this Appendix, we describe the main concepts related to Deep Learning as a complement to the details present in Section 3. As deep learning is a wide field, we prioritize Neural Networks (structure and components). The objective is to provide an overview of the concepts applied during the development of this work. We start by describing the inspiration behind the abstraction of a Neural Network.

BIOLOGICAL INSPIRATION

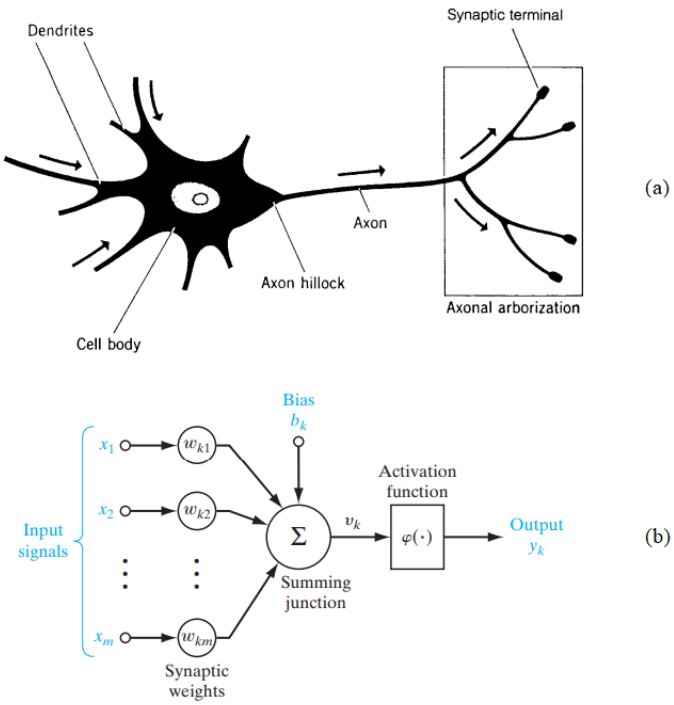
Neurons are responsible for carrying information throughout the human body, and they communicate with each other by sending electrical and chemical signals at a synapse. There are three kinds of neurons:

1. The sensory neurons conduct impulses from a sense organ to the central nervous system;
2. The motor neurons are responsible for voluntary and involuntary movements;
3. The interneurons pass signals from sensory neurons to motor neurons and other

interneurons.

Each neuron consists of a cell body, dendrites, and an axon, and the structure can be seen in Figure A.1a. With the intention to mimic the anatomy and function of the human brain, the organization of artificial neural networks is similar to that of biological neural networks. The comparison can be seen in Figure A.1, where the weight factors w_k correspond to synapses, and the stimuli coming to the cell $(x_1, x_2, ..., x_m)$ perform the same function as the information that is carried in biological cells.

Figure A.1 – Similarity between biological and artificial neural networks. (a) Representation of a biological neuron cell. (b) an artificial neuron model



Source: Arbib 2023, Haykin 2009.

NEURAL NETWORKS

A neural network is a learning representation whose goal is to approximate some function f^*, and it is a structure composed of processing units called neurons (Haykin 2009). Within the data collected from a problem analyzed, the neural network defines an approximate mapping function $y = f^*(x)$ that outlines an input x to an output y. During the neural network training, the objective is to adjust $f(x)$ to match f^*, by using the training dataset.

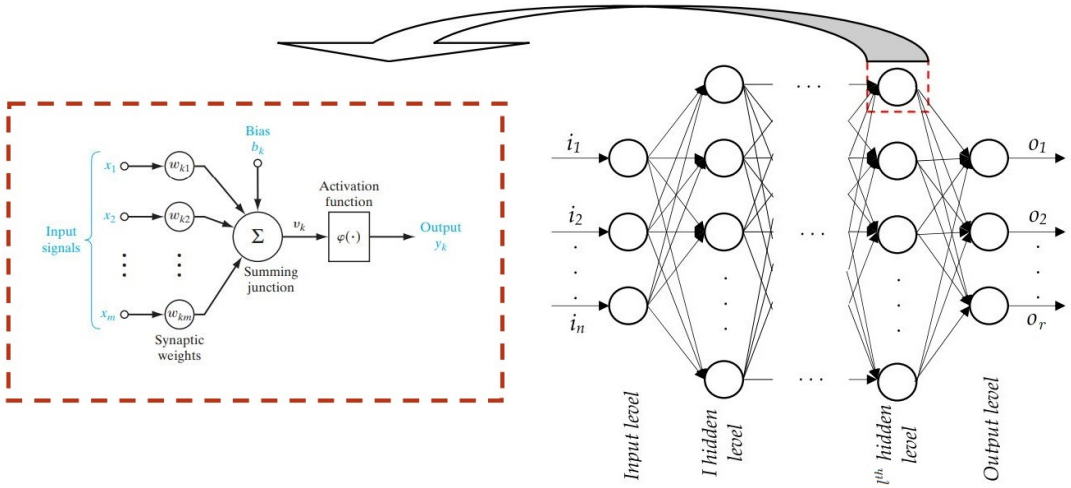
A neuron is structured by a linear combiner, an activation function, and a set of synapses - designated as synaptic weights, which store the knowledge achieved through a learning algorithm. The governing equations of each single neuron are equations A.1 and A.2:

$$v_k = \sum_{j=1}^{m} w_{kj} x_j + b_k \tag{A.1}$$

$$y_k = \varphi(v_k) = \varphi\left(\sum_{i=1}^{m} w_{kr} x_r + b_k\right) \tag{A.2}$$

The output y_k represents an input to the neurons of another layer, or an element of the output vector of the neural network, as we can observe in Figure A.2. The standard architecture has as the first layer an input vector $I = [i_1, i_2, ..., i_n]$, which is called the input layer, and the last layer is featured by a vector $O = [O_1, O_2, ..., O_r]$, called the output layer. The other layers are called hidden layers, and they are not seen directly from the input or output of a network, being the depth of a neural network related to the number of layers in the architecture. The linear combiner is represented by Σ, $\varphi(\cdot)$ represents an activation function, and b_k represents the bias, which has the effect of applying an affine transformation[1] to the result of the linear combiner.

Figure A.2 – Artificial neuron and the structure of the feed-forward artificial neural network



Source: Modified from (Tanikic and Despotovic 2012) and (Haykin 2009).

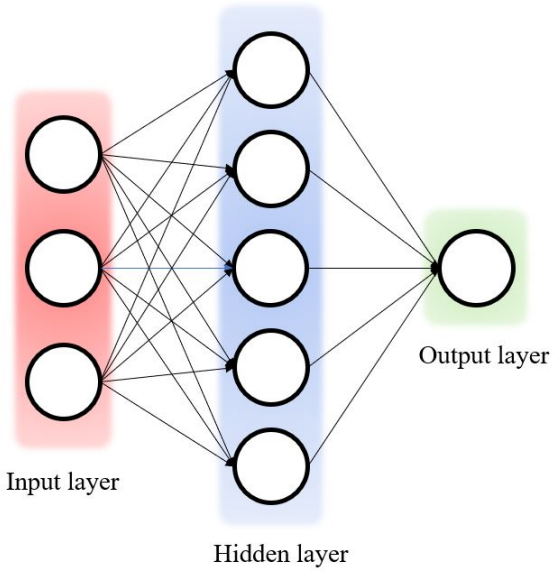
The artificial neural networks can hold an arbitrary number of layers and a random number

[1] Affine transformation is a linear mapping method that preserves points, straight lines, and planes.

of neurons. Besides those parameters, the performance of an artificial neural network also depends on the chosen transfer function, the presence of a bias, and the scheme in which the neurons are connected. Since there are no formal rules to drive the selection of the mentioned arguments, the recommendation is to use personal skills and experience in a manner to obtain the least error.

To understand how the layers are connected to one another, we have a simple example of a neural network in Figure A.3, with an input layer (red), a hidden layer (blue), and an output layer (green).

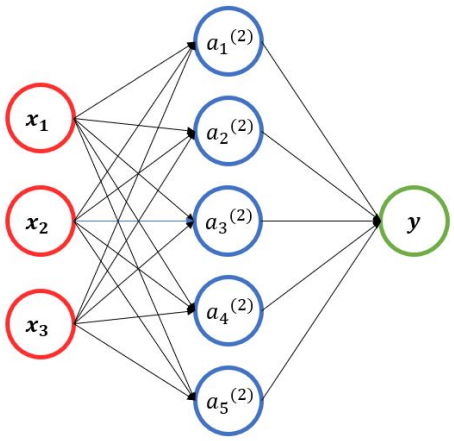
Figure A.3 – Layers representation: input level (red), hidden level (blue), and output level (green)



Source: Own work.

For this case, the architecture has three inputs (x_1, x_2, x_3), a hidden layer, and one output (green). Let us define $a_j^{(l)}$ and $b_j^{(l)}$ as the activation and bias of neuron j^{th} in layer l^{th}, respectively. And since the first layer in every neural network is the input layer, $a_j^{(1)} = x_j$, where x_j is an input, as we can see in Figure A.4. The weight parameter from the k^{th} neuron in the $(l-1)^{th}$ layer to the j^{th} neuron in the l^{th} layer is denoted by $w_{jk}^{(l)}$.

Figure A.4 – General flow graph of a neural network with 3 layers



Source: Own work.

Therefore, for the equations that govern the neural network from Figure A.4, the linear combinations of the inputs weighted by $w^{(2)}$ of the second layer are:

$$v_1^{(2)} = w_{11}^{(1)} x_1 + w_{12}^{(1)} x_2 + w_{13}^{(1)} x_3 + b_1^{(1)} \tag{A.3}$$

$$v_2^{(2)} = w_{21}^{(1)} x_1 + w_{22}^{(1)} x_2 + w_{23}^{(1)} x_3 + b_2^{(1)} \tag{A.4}$$

$$v_3^{(2)} = w_{31}^{(1)} x_1 + w_{32}^{(1)} x_2 + w_{33}^{(1)} x_3 + b_3^{(1)} \tag{A.5}$$

$$v_4^{(2)} = w_{41}^{(1)} x_1 + w_{42}^{(1)} x_2 + w_{43}^{(1)} x_3 + b_4^{(1)} \tag{A.6}$$

$$v_5^{(2)} = w_{51}^{(1)} x_1 + w_{52}^{(1)} x_2 + w_{53}^{(1)} x_3 + b_5^{(1)} \tag{A.7}$$

And the outputs of each neuron of the second layer are presented by the equations from A.8 to A.12:

$$a_1^{(2)} = \varphi\left(v_1^{(2)}\right) = \varphi\left(w_{11}^{(1)} x_1 + w_{12}^{(1)} x_2 + w_{13}^{(1)} x_3 + b_1^{(1)}\right) \tag{A.8}$$

$$a_2^{(2)} = \varphi\left(v_2^{(2)}\right) = \varphi\left(w_{21}^{(1)} x_1 + w_{22}^{(1)} x_2 + w_{23}^{(1)} x_3 + b_2^{(1)}\right) \tag{A.9}$$

$$a_3^{(2)} = \varphi\left(v_3^{(2)}\right) = \varphi\left(w_{31}^{(1)} x_1 + w_{32}^{(1)} x_2 + w_{33}^{(1)} x_3 + b_3^{(1)}\right) \tag{A.10}$$

$$a_4^{(2)} = \varphi\left(v_4^{(2)}\right) = \varphi\left(w_{41}^{(1)} x_1 + w_{42}^{(1)} x_2 + w_{43}^{(1)} x_3 + b_4^{(1)}\right) \tag{A.11}$$

$$a_5^{(2)} = \varphi\left(v_5^{(2)}\right) = \varphi\left(w_{51}^{(1)} x_1 + w_{52}^{(1)} x_2 + w_{53}^{(1)} x_3 + b_5^{(1)}\right) \tag{A.12}$$

For the output layer, we have:

$$v_1^{(3)} = w_{11}^{(2)} a_1^{(2)} + w_{12}^{(2)} a_2^{(2)} + w_{13}^{(2)} a_3^{(2)} + w_{14}^{(2)} a_4^{(2)} + w_{15}^{(2)} a_5^{(2)} + b_1^{(2)} \tag{A.13}$$

$$a_1^{(3)} = \varphi\left(v_1^{(3)}\right) = \varphi\left(w_{11}^{(2)} a_1^{(2)} + w_{12}^{(2)} a_2^{(2)} + w_{13}^{(2)} a_3^{(2)} + w_{14}^{(2)} a_4^{(2)} + w_{15}^{(2)} a_5^{(2)} + b_1^{(2)}\right) \tag{A.14}$$

More generally, we can rewrite the equations as:

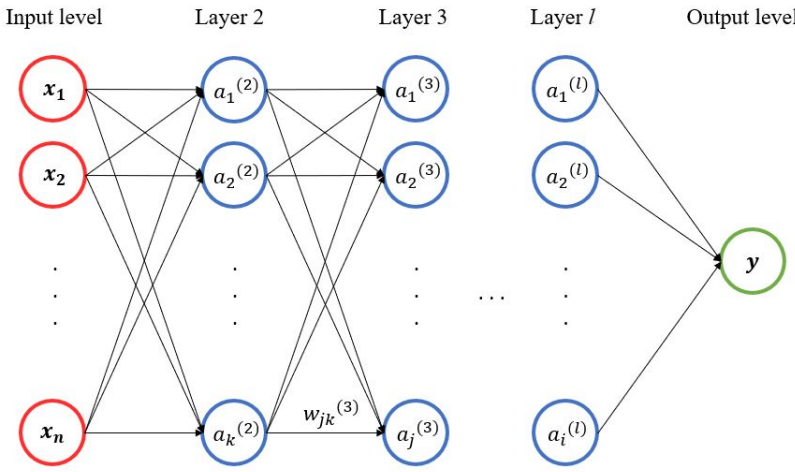
$$a_i^{(2)} = \varphi\left(v_1^{(2)}\right) = \varphi\left(\sum_{j=1}^{3} w_{ij}^{(1)} x_j + b_i^{(1)}\right), i \in \{1,2,3,4,5\} \tag{A.15}$$

$$a_i^{(3)} = \varphi\left(v_1^{(3)}\right) = \varphi\left(\sum_{j=1}^{5} w_{ij}^{(2)} a_j^{(2)} + b_i^{(2)}\right), i = 1 \tag{A.16}$$

We can define $a^{(l)}$ a vector with each neuron output a_i from a layer l, $W^{(l)}$ as a vector compounded by the $w_{ij}^{(l)}$, and $b^{(l)}$ the vector within each bias $b_i^{(l)}$ from layer l. Then, following the reasoning presented for architecture with few layers, for a general architecture with $l+1$ layers ($l > 1$), as in Figure A.5 - the output, or activation, can be calculated recursively as a function of the outputs of the previous layer ($a^{(l)}$), the matrix of weights ($W^{(l)}$), and the vector of biases ($b^{(l)}$), according to A.17:

$$a^{(l+1)} = \varphi\left(W^{(l)} a^{(l)} + b^{(l)}\right) \tag{A.17}$$

Figure A.5 – General flow graph of a neural network with $(l+1)$ layers



Source: Own work.

Now we introduce in more detail the concept behind activation functions since they play an integral role in neural networks by introducing nonlinearity.

ACTIVATION FUNCTIONS

The activation functions are responsible for transforming the summed weighted inputs from the neuron and addressing them to the next layer. They implement the nonlinear component to the model, permitting the performance of more complex tasks. In the absence of an activation function, the network is reduced to a model of linear regression.

The most common activation functions are the sigmoid, the hyperbolic tangent, and the rectified linear unit, and each one is more suited to the specific problem.

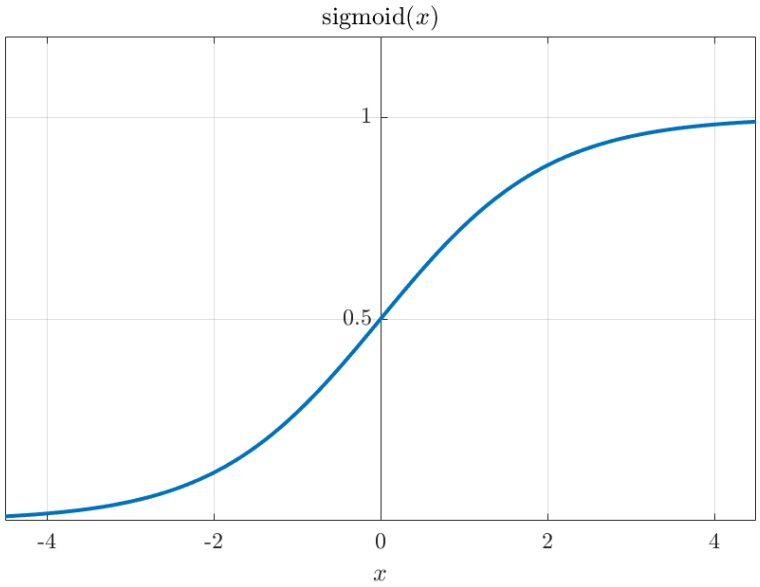
SIGMOID FUNCTION

The sigmoid function assumes values only in the interval $(0, +1)$, and it is defined by the equation (A.18):

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \tag{A.18}$$

The sigmoid function is the most common form of activation function used in the construction of neural networks (Haykin 2009), however, its considerable disadvantage is that the sigmoid's derivative is almost zero when it approaches 0 or 1, which impacts the iterative calculations required for the network optimization. Figure A.6 is the plot of the sigmoid function for $x \in [-4, +4]$.

Figure A.6 – Sigmoid function for $x \in [-4, +4]$



Source: Own work.

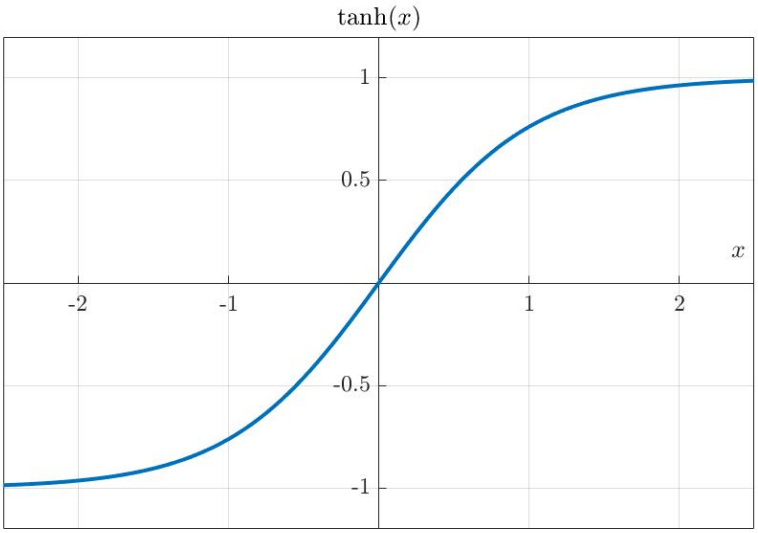
HYPERBOLIC TANGENT FUNCTION

The hyperbolic tangent function is given by equation (A.19).

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{A.19}$$

The hyperbolic tangent function is similar to the sigmoid function, having similar shapes, but tanh is symmetric in the x axis, and $\in [-1, +1]$. Figure A.7 plots the function.

Figure A.7 – Hyperbolic Tangent Function for $x \in [-2, +2]$



Source: Own work.

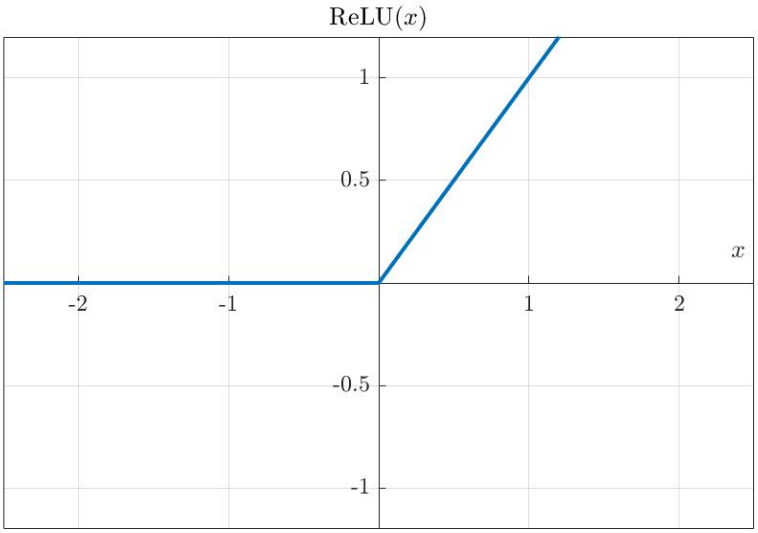
RECTIFIED LINEAR UNIT

The rectified linear unit (ReLU) function is defined in equation A.20. Contrasting this function with the two presented before, its greatest advantage is the non-saturation of the gradient, which accelerates the convergence of the algorithm optimization.

$$\text{ReLU}(x) = max(0, x) \tag{A.20}$$

Figure A.8 represents the function's plot. We can observe that the activation is a simple threshold $x = 0$, making ReLu much more efficient than the functions mentioned above (Krizhevsky et al. 2012).

Figure A.8 – ReLU Function for $x \in [-4, +4]$



Source: Own work.

SOFTMAX

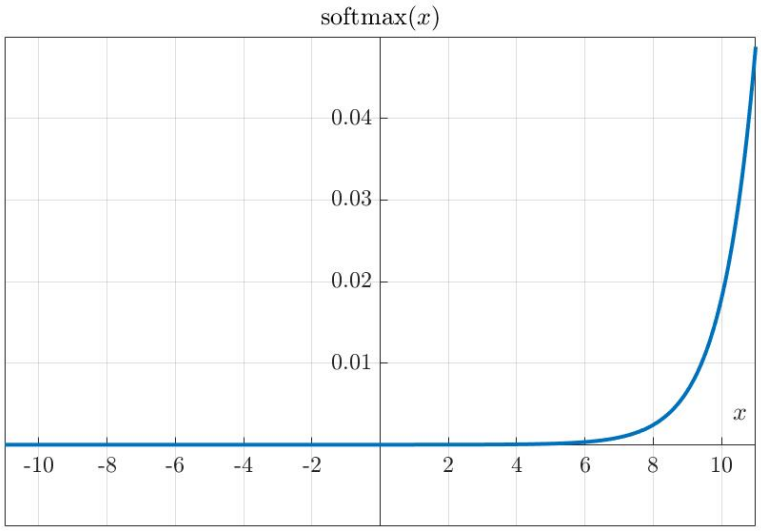
The softmax function is defined to be

$$\text{Softmax}(x)_i = \frac{e^{x_{i=1\ldots n}}}{\sum_{j=1}^{n} e^{x_j}} \tag{A.21}$$

with $n \geqslant 1$.

Figure A.9 represents the function's plot. The softmax function, also known as soft-argmax, takes a vector x of n real numbers as input and normalizes it into a probability distribution. The advantage is that after applying the function, a vector that could have components negative or greater than one, now each element will be in the interval $(0, 1)$, expressing the likelihood of each value.

Figure A.9 – Softmax function for $x \in [-10, +10]$



Source: Own work.

LOSS FUNCTION

The loss function, or cost function, is a method of evaluating the neural network model applied to predict the expected outcome. It is responsible for optimizing the values of the weight (W) and bias (b) since those parameters are calculated iteratively when the running model compares the predicted outcomes to the actual output values.

Considering x the input vector and y the output vector, the loss function is conditional to x, y, W, and b, and the purpose is to minimize it. Let us call the loss function $\mathcal{L}$, thus, the algorithm optimization involves finding the W^* and b^* which decrease the value of $\mathcal{L}$, as we can verify in (A.22).

$$(W^*, b^*) = \underset{W,b}{\operatorname{argmax}}\, \mathcal{L}(W, b, x, y) \tag{A.22}$$

There are several types of loss functions, each more suited to certain problems. Selecting an appropriate cost function is extremely important since neural networks operate those calculations in order to learn parameters.

To choose a loss function, we first have to classify the problem we are trying to solve.

There are mainly two different categories: Classification loss and Regression loss. The first group includes problems of binary or multiclass classification, for example, if we want to label a dataset compounded of handwritten images of digits from 0 to 9. Whereas, in the regression loss, we are dealing with continuous data problems. For instance, if we want to predict the prices of houses on the basis of some features.

CLASSIFICATION LOSS

In classification, we are trying to predict the output from a set of finite categorical values. Below we have the two main loss functions for this category.

• Cross-Entropy Loss

It estimates the performance of a classification model through the evaluation of the predicted output's probability (value between 0 and 1). As the probability converges to the actual label, the cross-entropy decreases, and it can be calculated by

$$L = -\frac{1}{m}\sum_{i=1}^{m} y_i \cdot \log(\hat{y}_i) \tag{A.23}$$

where y is the true probability distribution, $\hat{y}$ is the predicted probability distribution, m is the number of training samples, and i is the i^{th} training sample in the dataset.

• Hinge Loss

It is the second most common loss function used for classification problems after the cross-entropy. The function penalizes both the wrong and uncertain predictions, and it is evaluated by

$$L = max(0, 1 - t \cdot y)) \tag{A.24}$$

where $t = \pm 1$ is the intended output for class labels of $+1$ or -1, and y is the true probability distribution.

REGRESSION LOSS

The other category of loss function is the regression. Below we have two types of regression losses.

• Mean Square Error

The mean square error (MSE) loss function is defined as the average of squared differences between the actual and the predicted output. The model is robust because the penalties occur by squaring the errors, and it can be calculated by

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{A.25}$$

where y is the true probability distribution, $\hat{y}$ is the predicted probability distribution, n is the number of training samples, and i is the i^{th} training sample in the dataset.

• Mean Absolute Error

We define the Mean Absolute Error (MAE) loss function as the average of absolute differences between the actual and the predicted value, i.e., the average magnitude of errors in a set of predictions. It is calculated as

$$MAE = \frac{\sum_{i=1}^{n} |haty_i - y_i|}{n} \tag{A.26}$$

where y is the true probability distribution, $\hat{y}$ is the predicted probability distribution, n is the number of training samples, and i is the i^{th} training sample in the dataset.

GRADIENT DESCENT AND BACKPROPAGATION

As we discussed in the Loss Function topic, the main goal of neural network training is to find the optimized W and b parameters that suit the function and data distribution represented. For that, the cost function needs to be minimized, and the most popular algorithm to perform optimization is called Gradient Descent (Cauchy 1847).

The Gradient Descent technique is an iterative method that calculates the derivative of the cost function and applies them as step updates according to a learning rate formerly

set, α. The parameter α adjusts the dimension of the modifications at each update step of the model. The method admits the evaluation of the cost function at an arbitrary point, and with a process named Backpropagation, it is possible to measure its negative gradient, which results in a step in that direction.

Equation (A.27) brings the algorithm's approach in the $(k+1)$ iteration, with θ being a set of parameters for a target function $f(\theta)$.

$$\theta[k+1] \leftarrow \theta[k] - \alpha \nabla f(\theta[k]) \tag{A.27}$$

For updating the neural network, not only does the gradient for the output layer need to be calculated, but for each layer as well. For this, we require the chain rule of calculus.

Let $x \in \mathbb{R}$, and we have two functions: $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$. Suppose that $y = g(x)$ and $z = f(g(x)) = f(y)$. Then the chain rule asserts that

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx} \tag{A.28}$$

We can generalize to a vectorial structure. Let $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$, $g : \mathbb{R}^n \to \mathbb{R}^p$, and $f : \mathbb{R}^m \to \mathbb{R}^n$. If $y = g(x)$ and $z = f(y)$, then

$$\frac{\partial z}{\partial x_i} = \sum_j \frac{\partial z}{\partial y_j}\frac{\partial y_j}{\partial x_i} \tag{A.29}$$

Now we can apply the gradient for each layer of the neural network. As we saw in Section A, the weighted linear combination of the $(l+1)^{th}$ layer of a neural network, with n neurons, can be expressed by

$$v_i^{(l+1)} = \sum_{j=1}^{n} w_{ij}^l a_j^{(l)} + b_i^{(l)}, l > 1 \tag{A.30}$$

And its activation is

$$a_i^{(l+1)} = \varphi\left(\sum_j^n w_{ij}^l a_{j=1}^{(l)} + b_i^{(l)}\right), l > 1 \tag{A.31}$$

Combining equations (A.30) and (A.31), the output v can be written recursively by (A.32).

$$v_i^{(l+1)} = \sum_{j=1}^{n} w_{ij}^{l} \varphi(v_k^{(l)}) + b_i^{(l)},\ l > 1 \tag{A.32}$$

Then, we can apply the formula present in (A.28), and we get:

$$\frac{\partial v_i^{(l+1)}}{\partial v_j^{(l)}} = \varphi^{'}(v_j^{(l)}) w_{ij}^{(l)},\ l > 1 \tag{A.33}$$

Now we can calculate the derivative of the cost function $\mathcal{L}$ (equation (A.34)).

$$\frac{\partial \mathcal{L}}{\partial v_i^{(l)}} = \sum_{j=1}^{n+1} \frac{\partial \mathcal{L}}{\partial v_j^{(l+1)}} \frac{\partial v_j^{(l+1)}}{\partial v_i^{(l)}} = \varphi^{'}(v_i^{(l)}) \sum_{j=1}^{n+1} \left(w_{ij}^{(l)} \frac{\partial \mathcal{L}}{\partial v_j^{(l+1)}} \right) \tag{A.34}$$

Since we are interested in the derivatives of the weights and biases, equations (A.35) and (A.36) are calculated.

$$\frac{\partial \mathcal{L}}{\partial w_{ij}^{(l)}} = \frac{\partial \mathcal{L}}{\partial v_i^{(l+1)}} \frac{\partial v_i^{(l+1)}}{\partial w_{ij}^{(l)}} \tag{A.35}$$

$$\frac{\partial \mathcal{L}}{\partial b_i^{(l)}} = \frac{\partial \mathcal{L}}{\partial v_i^{(l+1)}} \frac{\partial v_i^{(l+1)}}{\partial b_i^{(l)}} \tag{A.36}$$

Considering that the derivatives for weight and bias are respectively

$$\frac{\partial v_i^{(l+1)}}{\partial w_{ij}^{(l)}} = \varphi(v_j^{(l)}) \tag{A.37}$$

$$\frac{\partial v_i^{(l+1)}}{\partial b_i^{(l)}} = 1 \tag{A.38}$$

Substituting equation (A.37) in (A.35) and equation (A.38) in (A.36), we have

$$\frac{\partial \mathcal{L}}{\partial w_{ij}^{(l)}} = \varphi(v_j^{(l)}) \frac{\partial \mathcal{L}}{\partial v_i^{(l+1)}} \tag{A.39}$$

$$\frac{\partial \mathcal{L}}{\partial b_i^{(l)}} = \frac{\partial \mathcal{L}}{\partial v_i^{(l+1)}} \tag{A.40}$$

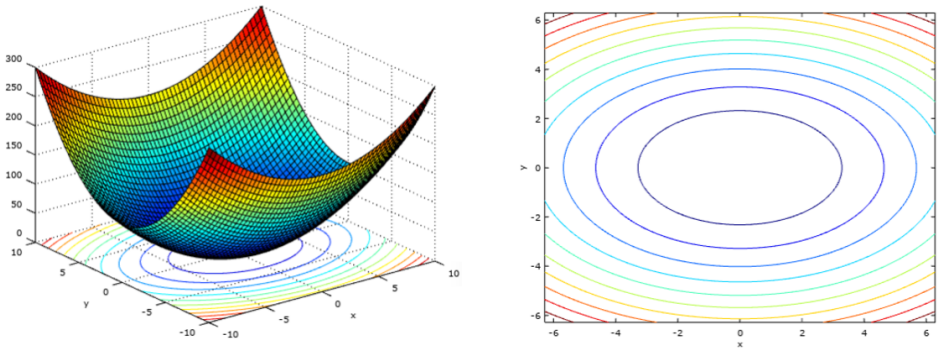
Therefore, the values obtained in equations (A.39) and (A.40) are used to update W and b, as described in equation (A.27). The model "recognizes" the parameters' adjustment through the backpropagation algorithm for training feedforward, and it takes the equation (A.27) recursively until the gradient is transmitted to all layers of the neural network. Thus, the updated equations are described below.

$$w_{ij}^{(l)} \leftarrow w_{ij}^{(l)} - \alpha \varphi(v_j^{(l)}) \frac{\partial \mathcal{L}}{\partial v_i^{(l+1)}} \tag{A.41}$$

$$b_i^{(l)} \leftarrow b_i^{(l)} - \alpha \frac{\partial \mathcal{L}}{\partial v_i^{(l+1)}} \tag{A.42}$$

As an example of the method, figure A.10 brings the graphs and contours of a function of two variables (x and y).

Figure A.10 – Gradient Descent example



Source: Adapted from (Saeed 2021).

Since the function exemplified is described by equation (A.43), the general form of the gradient vector is given by equation (A.44).

$$f(x, y) = x^2 + 2y^2 \tag{A.43}$$

$$\nabla f(x, y) = 2x_i + 4y_j \tag{A.44}$$

The Algorithm 1 brings two iterations of the procedure with $\alpha = 0.1$. In this example, we have only two iterations, but normally the calculations run until there is no change in the values computed, if they are below a certain threshold, or we settle a defined maximum number of iterations.

Algorithm 1 Gradient Descent algorithm example

Require: $t \geq 1$

Require: Learning rate α

Require: Initial parameters W and b

1. Initial $t = 0$

   $x[0] = (2, 4)$ ▷ This is a randomly chosen point

2. At $t = 1$

   $x[1] = x[0] - \alpha \nabla f(x[0])$

   $x[1] = (2, 4) - 0.1 \times (4, 16)$

   $x[1] = (1.6, 2.4)$

3. At $t = 2$

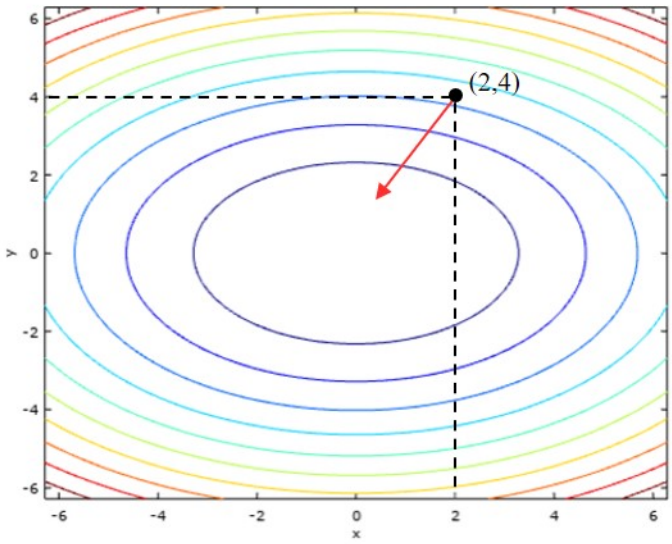
   $x[2] = x[1] - \alpha \times \nabla f(x[1])$

   $x[2] = (1.6, 2.4) - 0.1 \times (3.2, 9.6)$

   $x[2] = (1.28, 1.44)$

While the Gradient Descent updates the parameters for each θ, which increases the computational cost and memory allocation size, an alternative is the optimizer stochastic gradient descent (SGD). SGD performs a parameter update for each training example, which makes the processing much faster.

Figure A.11 – Red arrow illustrates the direction of the negative gradient descent at iteration $t = 1$



Source: Adapted from (Saeed 2021).