Universidade do Minho
Escola de Engenharia

Ana Catarina Maio Real

Healthy Bone Reconstruction through Generative Deep Learning

June 2023

Universidade do Minho
Escola de Engenharia

Ana Catarina Maio Real

Healthy Bone Reconstruction through Generative Deep Learning

Master’s Dissertation
Master’s Degree in Biomedical Engineering
Medical Eletronics

Dissertation oriented by
Doctor Professor Carlos Manuel Gregório Santos Lima
Master Pedro Ribeiro

June 2023

DECLARATION

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Licença concedida aos utilizadores deste trabalho

Universidade do Minho, Junho de 2023

Assinatura:

”One day, in retrospect, the years of struggle will strike you as the most beautiful.”

— Sigmund Freud

ACKNOWLEDGEMENTS

Throughout my academic journey, I had the honor and privilege of being surrounded by incredible people who accompanied me at every moment of one of the most important chapters of my life. People who gave me unforgettable and unique moments, as well as, supported me in all circumstances, contributing to the person I am today.

First of all, I would like to express my heartfelt gratitude to my mentor, Professor Carlos Lima, for his support, encouragement, and availability throughout the realization of this project. Additionally, I would like to extend my thanks to PeekMed for the opportunity and trust they have placed in me. I am so grateful for the chance to work with such an exceptional team, and I would especially like to thank Pedro Ribeiro, Daniel Cruz, and André Fabião for their tireless support and promptness in the development of this dissertation. Their contributions were instrumental to the success of this project, and I am truly appreciative of their efforts.

As someone who values personal growth and development, I took part in extracurricular activities such as the B-Side Soft Skills Academy and Portuguese Women in Tech Mentoring Program. In this way, I want to express my gratitude to Paulo Silva and my colleagues at B-Side for helping me develop my soft skills and step outside of my comfort zone. I would also like to thank my mentor at Portuguese Women in Tech, Raquel Carvalho, for her kindness, inspiration, and knowledge-sharing.

My dearest family in Braga - Margarida, Mafalda, Teresa, Luís, Raul, Gabriel, and Manel – as I near the end of our five-year journey, I cannot express enough gratitude for your friendship and support. The memories we have shared will always hold a special place in my heart, and I know that no matter where life takes us, we will continue to create unforgettable experiences together. Thank you from the bottom of my heart. To my longtime friends, I am incredibly grateful for your friendship and constant companionship. Thank you for being a constant source of support and love in my life. In addition, I would like to thank my biomedical friends, my godmother Francisca, Filipa, and my master partner João, for your patience, friendship, and support throughout these years.

To my beloved boyfriend Diogo and his wonderful family, I want to express my deep gratitude for your presence in my life. Diogo, you are the kindest, most supportive, and loving partner I could ever ask for. I feel incredibly lucky to have you by my side and I am excited to continue creating memories with you. Thank you for being an essential part of my life. I love you.

Finally, I want to thank my family for always being there for me and supporting me throughout my journey. Your love and encouragement have meant everything to me, and I couldn't have made it this far without you. To my dearest grandparents Maria, Manuel, António, and Rosa, the most important people in my life, I want

to express my profound gratitude for everything you have done for me. Your unconditional love, support, and guidance have been a constant source of inspiration and strength. To my brave and resilient mother, who has always been by my side and supported me tirelessly in the pursuit of my dreams, I want to express my sincere appreciation for everything you have done for me. Thank you for always being my rock and inspiring me to be the best version of myself. To my little brother, who brings so much joy and laughter into my life, I am so grateful for every moment we spend together. Dad, even though you left me too soon, I want to thank you for always being my guardian angel and for the motivation that you bring me everyday. I miss you so much, but I know that you and Aunt Linda are still looking out for me. To my aunts and cousins, I would like to express my gratitude, as well as, the rest of my family, for their support and presence in my life. Thank you for always being there for me and for making life's journey more meaningful and enjoyable. I love you all deeply and forever.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

ABSTRACT

Artificial Intelligence (AI) is transforming the clinical practice of orthopedic surgeons by combining technology with their technical skills. AI can play a primary role in standardizing pre-surgical planning, assisting orthopedic surgeons in the decision-making process to minimize medical errors, guiding appropriate surgical management, and reducing the cost and duration of surgery through intelligent solutions in the field of orthopedics.

The complexity and variability of the glenoid cavity anatomy have been a challenge for the medical community, especially in reconstructive surgical interventions such as shoulder arthroplasty. This surgery is recommended for the treatment of osteoarthritis (OA), a pathology defined by the progressive degeneration of the articular cartilage of the humeral head and glenoid, causing pain, stiffness, and limitation of movement. In preoperative planning, a 3D reconstruction of glenoid bone defects can play a fundamental role in the comprehension of the patient's native anatomy and, consequently, assist the orthopedic surgeon in the decision-making process, to restore the morphological parameters of the scapula, which is crucial for functional outcomes and the longevity of the implant.

The main objective of this dissertation is the reconstruction of the healthy anatomy of the glenoid from three-dimensional computed tomography (3D CT) images through Generative Deep learning (GDL). In quantitative terms, the goal of this project is to virtually reconstruct glenoid bone defects so that the estimated version is within the range [-5°, 10°] since the purpose of shoulder arthroplasty is to accurately restore a healthy patient's anatomy.

This project explores two approaches for training a 3D image-to-image translation model: Pix2Pix and CycleGAN. In Pix2Pix, a reference image in the original domain X is available for each image in the target domain Y, allowing one-to-one mapping. In contrast, CycleGAN performs training with unpaired data, the images in domain X are semantically related to the images in domain Y, and there is not necessarily a reference image in domain Y for each image in domain X. The distinguishing feature of CycleGAN is the incorporation of cycle consistency loss, which facilitates training without paired data. In other words, this model translates from the original domain to the target domain without a one-to-one mapping. This study aims to investigate and compare the performance of these two architectures in the context of healthy bone reconstruction. Concisely, the generative model (CycleGAN or Pix2Pix) seeks to learn the mapping function between two domains, $G : X \to Y$, i.e., the conversion of an image x from the domain X to an image $G(x)$ from the domain Y, where x is an image of a scapula with the glenoid removed and $G(x)$ the sample produced by the generator, which wants to conceive images similar to those of the real dataset of the healthy domain. The study demonstrates the potential of the CycleGAN and Pix2Pix models to reconstruct a healthy bone from a defective bone.

Taking into consideration a significantly larger dataset, both models are expected to outperform in reconstructing a defective glenoid. That opens up a possibility for the development of an automated and intelligent virtual reconstruction tool that can be used in clinical applications, to ensure that the preoperative planning process of shoulder arthroplasty is efficient and quick, guide an appropriate surgical management, facilitate communication between surgeons, minimize medical errors, provide prognostic information, and optimize the performance of shoulder arthroplasty.

KEYWORDS: Artificial Intelligence, Orthopedic surgery, Preoperative Planning, Shoulder Arthroplasty, Osteoarthritis, Glenoid, Generative Deep Learning, 3D Image-to-image translation, Virtual Bone Reconstruction, Pix2Pix, CycleGAN.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

2D CT two-dimensional computed tomography. 1, 23, 27, 28
3D CT three-dimensional computed tomography. 1, 7, 18, 27–29, 43, 44, 60

AGWA Anterior Glenoid Wall Angle. 1, 31, 55
AI Artificial Intelligence. 1, 7, 17–19
AP anteroposterior. 1, 27, 28
AWS Amazon Web Services. 1, 58

CNNs Convolutional Neural Networks. 1, 58
CT computed tomography. 1, 23, 27, 33
CycleGAN Cycle Generative Adversarial Network. 1, 7–9, 11–13, 37, 38, 40, 41, 44, 46–49, 51–53, 55–57, 59, 60

GAN Generative Adversarial Network. 1, 11, 33, 34, 36, 37, 39, 40, 45, 56
GANs Generative Adversarial Networks. 1, 33, 35–37, 39, 41, 48, 55, 57, 58, 60
GDL Generative Deep learning. 1, 7, 18, 19, 35, 60
GHSI Glenohumeral Subluxation Index. 1, 11, 26, 29, 30, 59

HHM Humeral Head Medialization. 1, 11, 26, 30, 59

MAE Mean Absolute Error. 1, 53, 57
MRI magnetic resonance imaging. 1, 27
MSE Mean Squared Error. 1, 47

OA osteoarthritis. 1, 7, 23–27, 29

PCA Principal Component Analysis. 1, 32
PSNR Peak signal-to-noise ratio. 1, 48, 49, 51–53, 56, 57

SSIM Structural similarity index measure. 1, 48, 49, 51–53, 56
SSM Statistical Shape Model. 1, 11, 31–34, 55, 57

1

INTRODUCTION

The present work was realized in the scope of the master's thesis in Medical Electronics of the Integrated Master's Degree in Biomedical Engineering of the University of Minho. This chapter presents a contextualization and a framework of the theme, as well as the main motivations for its elaboration and the objectives that are intended to be achieved throughout the development of the project. The dissertation will be conducted in a corporate context, in this way, section 1.3 introduces the company PeekMed®. Finally, the last section explains the structure of the document.

1.1 Contextualization and Motivations

With scientific and technological advances, the implementation of Artificial Intelligence (AI) in the healthcare sector is inevitable, especially, due to the rising costs of healthcare [1]. Over the past few years, the application of AI in healthcare has grown considerably, primarily, due to its ability to automate services and, consequently, reduce costs, save time, and increase quality. For example, changing the daily decision-making process in order to minimize human errors and the implementation of systems that provide resource management in an efficient way. In this sense, AI has the potential to ensure the optimization of the performance of healthcare institutions, promoting the success of the health system.

AI is a modern technical approach that develops algorithms that enable machines with problem-solving and decision-making skills, making devices intelligent and efficient to perform specific tasks, which can correspond or exceed human performance [2]. In this domain, AI consists of a set of technologies that allow computers to help understanding, acting and learning in medical situations in order to improve health care at various levels, including diagnosis, management of hospital services, research, preventive medicine, clinical decision-making, among others. One of the future perspectives is the integration of algorithms as indispensable medical devices for patients and healthcare providers [1, 3, 4, 5].

Improvements in computational power, data storage and availability of high-quality data has driven the expansion of AI in the medical field [2]. The increment of digital medical images, information collected from databases and medical records, gives the ideal dataset for designing AI algorithms, especially, in the field of orthopedics [4].

AI is transforming the clinical practice of orthopedic surgeons by combining technology with their technical skills, thus, besides improving clinical outcomes for patients it allows to make the healthcare system more sustainable [3]. However, it remains a relatively new field for most orthopedic surgeons, more research is needed to identify more concrete areas of study and applications that can translate into a significant change in everyday clinical practice [4]. Currently, AI has been employed in orthopedic surgery to assess the surgical prognosis, create patient-specific solutions and reduce safety issues, in seconds and automatically. This emerging technology has also had a huge impact on preoperative planning [6].

Preoperative planning is considered one of the most important parts of surgery, being defined as the process of preparing for surgery with the purpose of optimizing patient, resources, and time management. Preoperative planning facilitates the characterization of customized solutions for each patient and, consequently, decreases risk factors, increases the efficiency and precision of the surgical procedure. Additionally, it helps to improve both short- and long-term clinical outcomes. AI can play a primary role in standardizing pre-surgical planning, assisting orthopedic surgeons in the decision-making process to minimize medical errors, guide appropriate surgical management, and as mentioned previously, reduce the cost and duration of surgery through industrial solutions in the field of orthopedics. The physician should analyze the results granted by AI, considering his experience and knowledge [7, 8].

The main motivation of this project is to optimize the preoperative planning of shoulder arthroplasty through the implementation of an innovative model, in the context of Generative Deep learning (GDL), for the 3D reconstruction of glenoid cavity bone defects, taking advantage of the evolution of medical imaging. In this way, the intention is to respond to the existing scientific gap regarding progressive technical approaches to solve this problem. The reconstruction of the native glenoid anatomy permits orthopedic surgeons to accurately understand the patient's situation and identify the desired outcome. Therefore, there is a reduction in human error, costs, and duration of the surgical intervention, that is, the quality and efficiency of the service delivered is improved. The opportunity to improve the quality of people's lives through technological solutions was one of the main reasons why I opted for the master's in Medical Electronics and the subject under study.

1.2 Objectives

The main objective of this dissertation is the reconstruction of the healthy anatomy of the glenoid cavity from three-dimensional computed tomography (3D CT) scans through GDL. In other words, it is expected to obtain the native shape of the glenoid, which is indispensable in the preoperative planning of surgical interventions that aim to restore the bone defect. After the implementation of this project, it is intended to ensure that the preoperative planning process of shoulder arthroplasty is efficient and quick, guide an appropriate surgical management, facilitate communication between surgeons, minimize medical errors, provide prognostic information, and optimize the performance of shoulder arthroplasty. In quantitative terms, the goal of this project is to virtually reconstruct glenoid bone defects so that the estimated version is within the range [-5°, 10°] since the purpose of shoulder arthroplasty is to accurately restore a healthy patient's anatomy.

1.3 Peekmed ®

The present project will be carried out in a company environment at PeekMed®, in order to explore a possible solution for the reconstruction of glenoid fossa bone defects, contributing to the preoperative planning of shoulder arthroplasty. Peekmed is an AI-based 3D preoperative planning system that aims to assist surgeon decision-making, reducing healthcare costs. In addition, it improves the efficiency and quality of the service provided through its innovative technology, which offers automatic and accurate planning of multiple clinical procedures. The system enables access to several planning tools for measurement and templating. In this way, it can reduce the duration of the process by up to 20% and obtain a more precise surgical prognosis, based on its many functionalities. For example, automatic placement of the most appropriate model, according to the measurements made previously. Such as adding and manipulating, simultaneously, the same object in 2D and 3D environments. Furthermore, 3D technology allows exporting the 3D model of the planning, being compatible with 3D printers, accessing and sharing the final report that helps in the comprehension of the procedure, step by step, and the resources needed for the medical intervention, among other benefits.

1.4 Dissertation Structure

The dissertation is organized into seven chapters, namely, the Introduction, the State of the Art, Research and Development Methodologies, the Case Study, the Results, the Analysis and Discussion, and lastly, the Conclusion.

The Introduction encompasses the contextualization and framing of the topic. The basic motivations that led to the investigation of the subject, its main objectives and presents the enterprise PeekMed®.

The State of Art defines the fundamental theoretical and scientific concepts for a better comprehension of the project. The current state of knowledge regarding the topic under investigation, in other terms, the virtual reconstruction of glenoid bone defects, is compiled in this chapter.

The Research and Development Methodologies chapter describes the technologies that are essential for the technical implementation of the proposed solutions. In particular, the characterization of the proposed solution, in particular, the characterization of the model implemented within the framework of GDL.

The Case Study, describes, in detail, the implemented algorithms, which aim to reconstruct the healthy bone anatomy of the glenoid cavity. It presents the organization, preparation and analysis of the dataset, in advance. The architecture and particularities of the elaborated models are specified in this chapter. In addition, it depicts the qualitative test designed and carried out to assess the feasibility of the developed solutions. The results of the application are shown in Chapter 5, then these are analyzed and discussed, rigorously, in Chapter 6.

The last chapter presents the conclusions and the preponderant contributions of the case study, as well as intends to answer all the objectives stipulated at the beginning of the dissertation. In addition, it proposes future measures to be adopted to improve the solution under consideration.

2

STATE OF ART

The state of the art of theoretical and scientific concepts contributes to the understanding of this dissertation in all its scope. In this chapter, firstly, the anatomy of the glenohumeral joint will be discussed (section 2.1), followed by the etiology and classification of its bone defects, emphasizing degenerative osteoarticular pathology (section 2.2). In section 2.3, the available treatments for osteoarthritis will be presented, inherently related to the preoperative planning explained in section 2.4. Ultimately, a literature review of the virtual reconstruction of bone defects will be elaborated, both of the glenoid and overall.

2.1 Glenohumeral Joint

The glenohumeral joint is formed by the head of the humerus and the glenoid cavity of the scapula (Figure 1). This joint is of the synovial type, so both areas are lined by the hyaline cartilage. The head of the humerus has a hemispherical configuration and interacts with the glenoid fossa, constituting the shoulder joint. In turn, the glenoid consists of a pear-shaped depression located at the lateral angle of the scapula [9]. The glenoid "labrum" consists of a fibrocartilaginous structure incorporated in the extremity of the glenoid increasing its depth and stability [10].

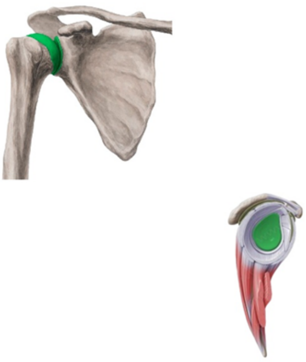

Figure 1: Glenohumeral joint (left) and the glenoid cavity (right). Adapted from [11].

The contact area between the two structures is restricted. Therefore, the joint has six degrees of freedom allowing a wide range of upper limb movements such as abduction, flexion, extension, internal and external rotation, making the shoulder joint inherently unstable and one of the most injured joints in the human body. Its stability depends on the balanced relationship between the static stabilizers, that is, its ligaments and articular capsule, as well as the dynamic stabilizers, i.e., the muscles and tendons of the rotator cuff, among other muscles. The synovial fluid, which fills the subscapular and subacromial bursae, produced in the synovial membrane acts as a lubricant reducing the friction caused by tendons and muscles, as they move over bony areas. A synovial membrane can be found on the inner surface of the joint capsule that internally coats the glenoid cavity. Several muscles act on the glenohumeral joint in order to enable shoulder kinesia, in particular, the scapuloumeral and thoracoumeral muscles. The muscles of the rotator cuff, that is, the supraspinatus, infraspinatus, teres major, and teres minor muscles, act to balance the joint, keeping the humeral head stable and centered in the glenoid fossa and support its movements [12, 11].

The complexity and variability of the glenoid cavity anatomy have been a challenge for the medical community, especially, in reconstructive surgical interventions, in particular, shoulder arthroplasty. There are studies that prove that the incorrect positioning of the glenoid due to, for example, bone loss often observed in patients with osteoarthritis, causes instability in the humeral head and may lead to loosening of the joint [13]. The inclination, the size, and especially, the version are the key parameters for evaluating the state of the glenoid, being essential for the preoperative planning of surgical interventions that aim to restore the bone defect to resemble the native glenoid fossa. Cadaveric studies have shown that these characteristics may vary according to age, race, and gender of the individual [13, 14, 15].

The dimensions of the glenoid, the width, and height, do not vary, substantially, between black and white patients. Contrarily, the difference between males and females is, statistically, significant. By means of a cadaveric study, Churchill et al. has shown in ([15]) that the width and height of the male glenoid were approximately 27.8 ± 1.6 mm and 37.5 ± 2.2 mm (mean ± standard deviation), respectively. The female glenoid width and height were approximately 23.6 ± 1.5 mm and 32.6 ± 1.8 mm (mean ± standard deviation), in that order. The average age of the study group was 25.6 years and was representative in terms of race and sex.

The inclination of the glenoid face is a morphological feature of the scapula with high individual variability, its standard value lies between 0° and 10° [16]. There is no considerable discrepancy in the glenoid cavity inclination between sexes or races [15].

The normal version of the glenoid is 0° ± 4° (mean ± standard deviation) [17]. There is no relevant inequality between men and women of the same race, although black patients exhibit a lower version compared to Caucasian patients [15].

2.2 Etiology and Classification of Bone Defects

The bones represent a fundamental element of the musculoskeletal system, associated dysfunctions often affect quality of life, particularly, at the professional and social levels. Bone defects in the glenoid occur mostly due to chronic dislocations, congenital anomalies, degenerative bone diseases of the glenohumeral joint, such as osteoarthritis, repercussions of autoimmune diseases, for example, rheumatoid arthritis, fixation devices used in the treatment of a proximal humeral fracture that damage the humeral surface or unsuccessful shoulder arthroplasty. The main goal of orthopedic surgeons is to reconstruct the bone defect to safeguard the function of the affected limb and the regression of pain in the patient [18].

Bone defects in the glenoid fossa can be classified according to Seebauer et al. [19], although there are multiple classifications present in the literature. After excision of the labrum or glenoid component, bone defects in the glenoid cavity are classified, regardless of their primary etiology, as centric, notably, superficial (C1), deep (C2), cavitary (C3), or destructive (C4); or in eccentric, these being divided based on size and location, into small (E1), medium (E2), large (E3) or massive (E4) and anterior (A), posterior (P), superior (S) or inferior (I), respectively [15, 19]. The classification of Antuna et al. divides them into central, peripheral, and combined, with each group subcategorized into mild, moderate, and severe [20].

2.2.1 *Degenerative Osteoarticular Pathology and Walch Classification*

Osteoarthritis is considered the most common cause of disability in the United States, it has been demonstrated that OA of the shoulder joint affects about 32.8% of patients over the age of 60. Glenohumeral OA is defined by progressive degeneration of the articular cartilage of the humeral head and glenoid component leading to pain, stiffness, and limitation of range of motion. Its diagnosis is based on medical history, physical examination, and imaging. Subsequently, can be divided into primary or secondary. Primary OA does not have a specific cause, its prevalence increases with age, women are more susceptible than men and genes are a preponderant factor in this condition. On the other hand, secondary OA can result from infection, trauma, surgery, avascular necrosis or a rotator cuff tear [17, 21].

Walch's classification describes the morphological alterations of the arthritic glenoid on 2D axial CT scans, being the most widely used system concerning shoulder replacement. Its main purpose was to prove that the glenoid version and the relationship between the humeral head and the glenoid are vital for preoperative planning and surgery. The existence of categorization allows guiding an appropriate surgical management, facilitates communication between surgeons, and provides prognostic information [22]. However, the original classification system had some limitations due to the use of two-dimensional computed tomography (2D CT), so it was slightly modified by Bercik et al. [23] to be applicable to 3D reconstructions, improving its reliability.

The modified Walch classification (Figure 2) consists of adding defects B3 and D, and defines more precisely the defect A2 [23]. The glenoid morphology of type A is described by a symmetrical arthritis without posterior subluxation of the humeral head, subtype A1 comprises a slight central erosion, and subtype A2 is a deep central erosion. Type B is defined by asymmetrical arthritis with posterior subluxation of the humeral head. The subtype B1 has no bone erosion but has posterior joint space narrowing and osteophyte, while the subtype B2 is characterized by a biconcave glenoid as a result of posterior glenoid erosion. The subtype B3 is depicted as a concave glenoid cavity with posterior bone wear and pathologic retroversion (>15°) or posterior subluxation or both. The type C defect manifests a retroversion greater than 25° due to a dysplastic glenoid. Finally, the type D morphology presents a significant anteversion (>5°) [17, 22].

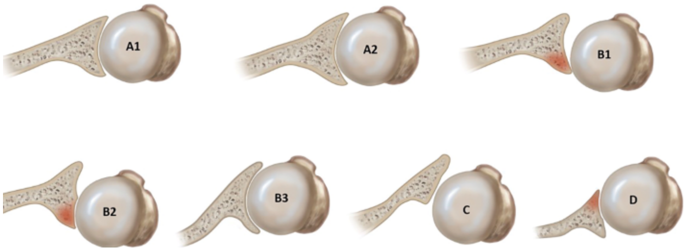

Figure 2: Modified Walch Classification [17].

2.3 Treatments for Glenohumeral Osteoarthritis

The optimal treatment for OA of the glenohumeral joint is frequently controversial and is chosen according to the patient's age, symptoms severity, activity level, imaging results, and medical comorbidity [17, 21].

2.3.1 *Nonsurgical Strategies*

Conventional procedures, that is, those that do not require medical surgical intervention, should be considered first, especially, in patients with mild to moderate osteoarthritis, when pain and functional limitation are reasonable, although there may be alarming radiographic alterations. This method can be effective in improving mobility and reducing pain, however, it does not modify the progression of the disease. The treatment options involve lifestyle and occupational changes, and physiotherapy compatible with the patient's needs, usually, including isometric exercises for rotator cuff and scapulothoracic muscle strengthening. Oral medication such as salicylates, paracetamol, also called acetaminophen, non-steroidal anti-inflammatory drugs, can help relieve pain and inflammation. At last, intra-articular corticosteroid injections provide pain relief, however, given the lack of evidence to prove their effectiveness and adverse effects, no more than three injections per year in any joint are recommended [17, 21, 24]. If nonsurgical measures are ineffective, especially, in the case of pain, surgical options are considered.

2.3.2 *Surgical Techniques*

Arthroscopy

Arthroscopy has been employed in the management of mild to moderate OA, particularly, in young people and athletes. This procedure permits the diagnosis, i.e., the characterization of the lesions related to joint dysfunction, and is used as a therapeutic tool for debridement and soft tissue repair. Arthroscopic treatment improves limb function, decreases pain, and consequently, improves the patient's quality of life. This is achieved by stabilizing the cartilage abnormalities, eliminating mechanical symptoms, and relieving capsular contractures. This intervention has few complications, low morbidity, and can postpone more complex surgeries. Unfortunately, it does not prevent arthritic progression [21].

Shoulder Resurfacing

Shoulder resurfacing is considered an alternative to conventional arthroplasty for the treatment of OA. This method involves reaming the proximal portion of the humeral head and resurfacing the arthritic articular surface by applying a metal alloy over the residual humeral head. The main advantages are the reduction in operative time, the extent of bone excision, and periprosthetic fractures (bone fracture around the implant after arthroplasty). The angle between the head and neck of the humerus remains intact and an osteotomy is not performed, so it is simpler to restore the version and inclination of the glenohumeral joint [21, 25]. The bone stock is mostly preserved, therefore, this intervention is, essentially, indicated for young people who may need, in the future, a total shoulder replacement [21]. It is important to mention, that bone loss in the humeral head should not be significant and the glenoid should not be injured, otherwise, the surgical treatment will be compromised [25].

Hemiarthroplasty

Hemiarthroplasty is the replacement of the humeral head by a metallic prosthesis, which is indicated when the humeral head is deteriorated or fractured and the glenoid face is intact [17, 26]. In addition, is also recommended when humeral arthritis and rotator cuff deficiencies coexist. The condition of the glenoid is critical for a successful treatment, a glenoid with severe bone loss and a non-functional rotator cuff cannot stabilize a glenoid component and can lead to its loosening [26]. Patients with primary OA and concentric bone defects have better surgical outcomes than those with secondary OA and eccentric wear [21]. The intervention exhibits a high rate of dissatisfaction, particularly, in young individuals, since the need for total shoulder arthroplasty arises in the long term [17, 21].

Reverse Shoulder Arthroplasty

Reverse shoulder arthroplasty is, predominantly, advised for the treatment of glenohumeral arthritis in patients with arthropathy of the rotator cuff, the group of muscles that provides active stabilization of the shoulder joint [21, 27]. If a rotator cuff tear occurs, the joint's center of rotation migrates upward, so the joint stresses become off-center, which may be the reason for glenoid loosening in total shoulder arthroplasty. Furthermore, the progressive dislocation of the humeral head causes degradation of the coracoacromial arch, leading to possible long-term post-surgical complications [21]. The inverted implant design of reverse shoulder arthroplasty, in other words, the use of a glenosphere and a humeral socket, enables stabilization of the glenohumeral joint, enhances deltoid function and range of motion. In detail, the glenosphere provides the mediation of the center of rotation, as well as, the torque reduction of the glenoid component. The humeral socket moves the joint's center of rotation distally, maximizing the length and tension of the deltoid muscle, increasing limb function and stability. This medical intervention is also indicated as a revision surgery of an unsuccessful anatomical total shoulder replacement [21, 27, 28].

Total Shoulder Arthroplasty

Total shoulder arthroplasty involves the replacement of the humeral head and glenoid with a prosthesis, usually, of cobalt chromium and polyethylene, respectively [26]. This surgery involves the excision of the humeral head and, subsequently, the placement of a titanium intramedullary stem in the proximal portion of the humerus and the humeral implant [25]. Considered the gold standard treatment for the management of advanced OA, total shoulder arthroplasty provides pain relief and the improvement of shoulder motion [21]. The most common complications are the loosening of the prosthesis, glenohumeral instability, periprosthetic fracture, and rotator cuff tear. There is a potential for infection, nerve damage, and dysfunction of the deltoid muscle [29]. The longevity and effectiveness depend on the selection of the patients who will be submitted to this intervention, given the risk of loosening of the glenoid [21].

2.4 Preoperative Planning

Preoperative planning is an essential prerequisite for the success of orthopedic procedures. Radiology plays a crucial role in pre-surgical planning, particularly, imaging studies. Medical imaging evaluates morphological aspects of the glenohumeral joint, such as bone stock, glenoid version, Glenohumeral Subluxation Index (GHSI), and Humeral Head Medialization (HHM), which often suffer alterations due to degenerative osteoarticular pathology [17, 30]. The comprehension of the relationship between soft and bone tissues is also indispensable for preoperative planning.

2.4.1 *Imaging Protocols*

Conventional radiographs, i.e., using the anteroposterior (AP) and axial views, are fundamental for the confirmation of the diagnosis and initial evaluation of shoulder OA [24]. Frontal radiography enables the localization of osteophytes and the observation of humeral head collapse. In turn, the views, mentioned above, are useful in determining joint space narrowing [17]. This medical imaging exam permits the identification of bone loss, but it is not as precise and reliable as computed tomography (CT) and magnetic resonance imaging (MRI) for bone loss and version calculation, as it has some limitations regarding patient positioning, overlapping structures and lack of three-dimensional perception [17, 31, 32].

The 2D CT may under- or overestimate glenoid cavity bone loss if the patient is not correctly oriented in the gantry, consequently, three-dimensional computed tomography (3D CT) is considered the gold standard for preoperative planning [32]. This imaging modality allows automatic image segmentation, in volume, of the glenoid and humerus and image reconstructions with high spatial resolution. Its digitalization is fast, having an acquisition time, of approximately, 5 seconds, which offers a better evaluation of the cortex and the bone contour. Its advantages come from its main disadvantage, the high dose of ionizing radiation [17].

Three-dimensional MRI has been suggested as an alternative to 3D CT for estimation of glenoid bone loss, it has been proven that there is no statistically significant discrepancy between bone loss measurements in patients with glenohumeral instability, comparing both examinations [32]. Nevertheless, MRI has a high time of acquisition (7-8 min) and postprocessing. In addition, image reconstructions may exhibit lower spatial resolution and the examination might be contraindicated for some patients due to factors, such as claustrophobia, the presence of cochlear implants, and pacemakers [17]. Its benefits include multiplanar 3D reconstructions along the scapula without the appearance of artifacts seen in two-dimensional images (stair-step artifact), the need for only one scan, subtle detection of bone contours, and more precise calculations. It does not involve ionizing radiation and, thoroughly, evaluates soft tissue structures, such as muscles, ligaments, and tendons [17, 32].

It should be noted that the monetary cost of standard radiography is significantly lower in comparison to CT and, especially, MRI.

2.4.2 *Evaluation of Morphological Abnormalities of the Glenohumeral Joint*

The analysis of the glenoid version is fundamental for the understanding of the pathological state of the shoulder, since an abnormal variation of the glenoid cavity version can have an impact on the biomechanics and stability of the glenohumeral joint. In shoulder arthroplasty, orthopedic surgeons should normalize the version, as it has been shown to increase stress and wear on the glenoid component in the upper limb prosthetic [17].

Despite the existence of multiple methods to measure the glenoid version, Friedman's line evidence greater interobserver reliability [33]. The Friedman line technique (Figure 3) determines the glenoid version on 2D CT, but it can be applied on 3D CT, in that case, the results were more precise in detecting retroversion and bone erosion of the posterior glenoid. A vertical line is drawn on the 3D surface of the glenoid face in the AP direction, then a 2D transverse plane is produced, perpendicular to the midpoint of the vertical line, which passes through the scapular axis. The neutral glenoid version line is defined by the line perpendicular to this plane. This process provides an image for measuring the relevant angle [30]. If the anterior margin is medial to the neutral version line, the glenoid cavity exhibits anteversion, otherwise, if it is the posterior margin, the glenoid fossa is retroverted [17, 33].

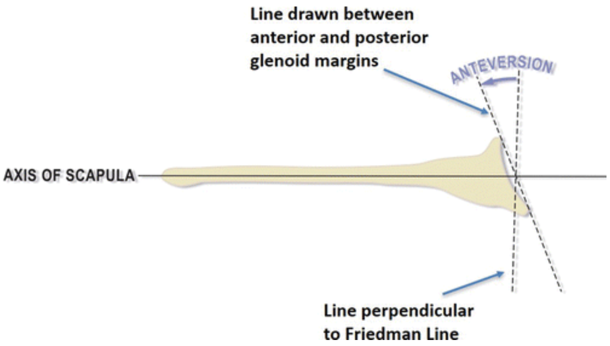

Figure 3: Friedman's line method for the calculation of the glenoid fossa version, in this case, the glenoid exhibits anteversion [17].

A deficiency in the glenoid bone stock causes a perturbation in the normal kinematics of the shoulder joint [17]. Glenoid bone loss should be analyzed when planning shoulder arthroplasty, its quantification is essential to determine the bone reconstruction surgical technique to be employed, with the goal of restoring the neutral glenoid version due the implant. If bone erosion is significant (>17.3%) a bone graft should be applied, otherwise Bankart's repair, a minimally invasive surgery that provides stabilization and restoration of shoulder joint function, will be sufficient [32].

The Pico method (Figure 4) involves defining a circumference in the contralateral joint, that is, asymptomatic, encompassing the 3-9h margin of the inferior glenoid, in the sagittal plane. Subsequently, the circle is transferred to the pathological shoulder, the bone defect is manually traced and its superficial area is calculated through a software [31]. The contralateral joint could have suffered previous injuries or surgeries, moreover, the extra dose of radiation induced on the patient should be avoided. In this way, an approach has emerged, which applies the circumference that best fits the anteroinferior or posteroinferior curvature of the glenoid cavity, just to the glenohumeral joint under study. Afterwards, the measurement of bone erosion is done in the same way as in the Pico method [17].

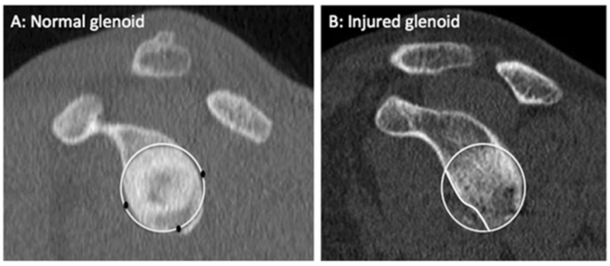

Figure 4: Pico method for measuring bone erosion [31]

.

The paleoglenoid line (Figure 5) consists of a line perpendicular to Friedman's line positioned at the lateral extremity of the glenoid. Thereby, an approximation of its native surface is obtained. In the presence of osteophytes, the axial sections inferior and superior to the junction between the coracoid process and the scapular spine should be analyzed to precisely limit the original face. In the following example, the bone loss was measured at the central point and 5 mm from the anterior and posterior margins of the glenoid fossa, in order to avoid osteophytes, that may generate incorrect measurements [34].

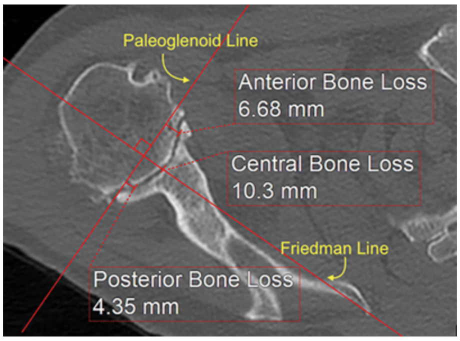

Figure 5: Practical example of the application of the paleoglenoid line technique [17].

GHSI evaluates the centrality of the humeral head within the glenoid cavity, being critical to direct OA treatment. It allows the assessment of viable surgical options, the need for corrective measures during arthroplasty, and its prognosis. Posterior glenohumeral subluxation correlates with the increase of the rate of loosening of the glenoid component, raising concerns about the longevity of substitution surgery [17, 35].

The GHSI measurement technique was originally defined by Walch et al. [36], however, in order to improve interobserver reliability it was slightly modified (Figure 6). On a 3D CT axial image, a line tangent to the anterior and posterior margins of the glenoid face is traced, followed by the dissection of the glenoid through a line perpendicular to the preceding one [17, 36]. The circumference that best fits the humeral head articular surface is defined and its diameter is calculated. To determine the GHSI, the posterior portion of the line that dissects the glenoid is divided by the diameter determined, previously [17, 34]. An index higher than 0.55 indicates a posterior subluxation [17, 35, 36].

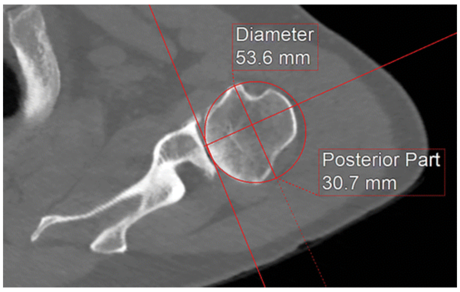

Figure 6: Practical example of the calculation of the GHSI using the modified Walch technique (GHSI=30.7 mm/53.6mm=0.57) [17].

A high HHM is compatible with central and anterior glenoid bone erosion, as well as fat infiltration in all muscles that constitute the rotator cuff [17, 34, 37]. HHM is measured at the apex of the humeral head medial to the paleoglenoid line (Figure 7) [37].

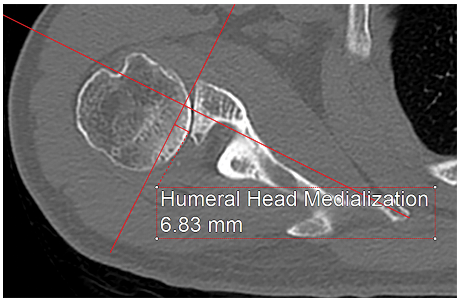

Figure 7: Practical example of HHM determination [17].

Fat infiltration and rotator cuff atrophy are the soft tissue factors that should be examined in pre-surgical planning, as they have been found to be related to poor clinical results, such as humeral head subluxation and postoperative loosening of the glenoid component. Fat infiltration and rotator cuff atrophy are associated with the increased of HHM, pathologic version, glenoid bone loss and, a more severe Walch classification (Type B3) [37].

2.5 Virtual Reconstruction of Glenoid Bone Defects

In preoperative surgical planning, the 3D reconstruction of a bone defect can perform a fundamental role in the comprehension of the patient's native anatomy and, consequently, support the orthopedic surgeon's decision to restore the morphological parameters of the scapula, crucial in the functional outcome and longevity of the implant. The morphological properties emphasize the individual anatomical variability of the glenoid, so it is challenging to find a technique to determine accurately its original shape.

The most commonly adopted method is to use the patient's healthy scapula as a template to guide glenoid bone reconstruction in unilateral arthropathy [14]. Mirroring the contralateral joint estimated the glenoid version, and inclination in 94% and 96% of the scapular pairs, respectively, with a difference under 5° and with a mean error of 4° for both parameters [38]. Despite the excellent results achieved, the contralateral joint may have suffered previous injuries or surgeries and the extra dose of radiation induced on the patient must be avoided, since the scan is, usually, unilateral [14, 39].

To overcome these limitations, it has been proven that the glenoid version can be predicted using linear regression models of the Resch angle (angle between the plane of the anterior glenoid wall and the plane of the glenoid fossa) and of Anterior Glenoid Wall Angle (AGWA), with 52.6% and 78.9% of the predictions, respectively, comprised in a range inferior to 5° [40]. The results show less reliability relatively to those of the mirroring of the contralateral joint, since they revealed a mean error of 4.70° for the linear regression equation of the Resch angle and 3.23° for the AGWA [38, 40].

Estimation of the glenoid version using a standard template of the healthy glenoid has been demonstrated to be less effective than using the patient's healthy scapula, since the standard template does not take into account the variability of the glenoid anatomy [14]. This technique measured, with a discrepancy of under 5°, the version in 57.9% of the pathological scapulae and exhibited a mean error of 3.68° [40].

2.5.1 *Statistical Shape Model [14]*

A Statistical Shape Model (SSM) can be applied to reconstruct glenoid fossa bone defects and to predict its morphological parameters, such as the version, the inclination, and its center point. This method was firstly proposed and evaluated by Vanden Berghe et al. [41] for the bone reconstruction of the acetabulum and prediction of anatomical properties of the hip joint.

A database consisting of 66 scans of healthy scapulae, i.e., without signs of glenohumeral arthropathy, was manipulated for the development of the SSM. Initially, the images were segmented using image processing software and converted to a 3D model with a triangular border of 1.5 mm. The design of the model integrates three phases (Figure 8). The first phase consists of registering a model of the dataset in the other models to acquire the corresponding surfaces. Afterwards, keeping the size reference, the models were aligned to suppress rotational and translational variations. Finally, through a Principal Component Analysis (PCA) the variations and the average shape of the models were extracted.

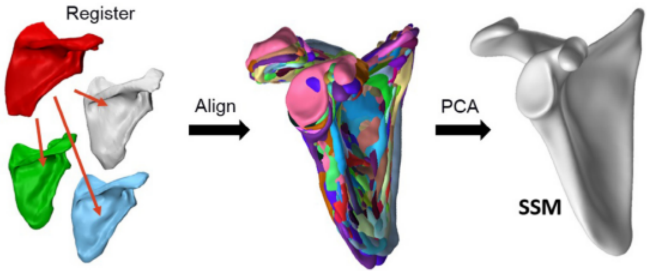

Figure 8: Steps for SSM development [14].

The SSM assumes that the native shape of the scapula can be obtained by extrapolating its healthy parts. The reconstruction procedure involves two steps, firstly, the bone defects are removed manually with the aid of a software. Secondly, the model is fitted to the residual healthy anatomy, using MATLAB (Figure 9). The proposed fitting algorithm optimizes the model coefficients individually, minimizing the discrepancy between the SSM and the healthy structures of the scapula.

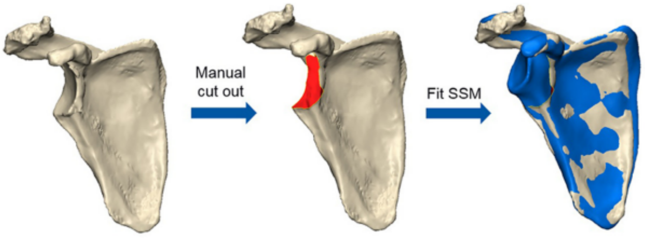

Figure 9: Application of SSM for virtual bone reconstruction [14].

The SSM has proven to be efficient in the reconstruction of glenoid bone defects and in the prognosis of its morphological parameters, being considered a successful mechanism to assist surgeons in the preoperative planning of shoulder arthroplasty. The degree of severity of bone defects leads to an increase in reconstruction errors, consequently, it is essential to carefully select the bone defect area. The weighting of the border had a positive effect on the results, indicating that the healthy regions adjacent to the bone defect contain important information.

One of its main advantages is that it allows the estimation of several anatomical features of the glenoid, including the version, unlike most of the methods previously discussed. The results presented a mean error of 2.9° and 2.4° in the version and inclination calculation, in that order, outperforming the alternative techniques formerly described (Table 1). The SSM-based reconstruction method discussed in [39], which allows the automatic analysis of glenoid bone defects using 3D measurements, exhibits a statistically equivalent performance, in comparison, to the contralateral joint mirroring technique in measuring subluxation distance, defect depth, total and local vault loss percentages, except for the defect area percentage. Therefore, this methodology can help orthopedic surgeons in the selection of the best treatment option in the preoperative stage and compare the postoperative results of similar bone defects, not requiring the healthy scapula of the patient as a reference.

The reconstruction technique has several drawbacks, such as the dependence on the process of registering the shape of the bone. In addition, it was exclusively evaluated in bone defects created artificially. This model assumes that osteoarthritis is restricted to the glenoid cavity and the other structures are only slightly affected by the pathology, so it may lead to inaccuracies in the fit of the SSM over the healthy regions.

Table 1: Comparison of the different methods in glenoid version prediction, based on the mean error and the results that presented a difference under 5°

Method	Mean Error	Difference <5°
Mirroring of the contralateral joint	4°	94%
Linear regression model of Resch angle	4.70°	52.6%
Linear regression model of AGWA	3.23°	78.9%
Standard model	3.68°	57.9%
SSM	2.9°	n.d.

2.6 Virtual Reconstruction of Bone Defects [42]

With the purpose of finding the appropriate solution to meet the objectives defined in section 1.2, it is necessary to perform a meticulous analysis of the various techniques suitable for the virtual reconstruction of bone defects. In this sense, there was a technology that aroused interest.

Xiong et al. [42] proposed to assess the accuracy of Generative Adversarial Networks (GANs) for the virtual reconstruction of midface bone defects, a complex bone structure essential for facial function and aesthetics. Methodologies for facial bone reconstruction include the mirroring of the healthy skull on the affected side and the use of a standard model or a deformable model, however, these have numerous disadvantages. For example, the mirroring technique can only be applied to unilateral bone defects. Therefore, there is a need to design an intelligent reconstruction of the healthy bone anatomy of the midface in clinical practice. The authors explore a model named GAN, in the context of Deep Learning, which demonstrates an excellent performance in image generation and is widely applied in medical image processing.

According to the anatomy, the midfacial bone was divided into five structural subunits, then the bone defects were manually created on CT images of patients with a healthy midface. The real images and the corresponding images with artificial bone defects were used to train the GAN. In the data preprocessing phase, there is a horizontal inversion in random axial slices, random resizing, and rotation of the images. Moreover, the addition of Gaussian noise improves the training process and avoids overfitting.

The architecture of the GAN consists of a generator and a discriminator (Figure 10). The generator is composed by a Resnet 101, previously trained. The input data of the generator consist of an image with the defective facial bone x and random noise z. The generator aims to obtain the native shape of the bone, that is, the reconstructed image $G(z)$, from the corresponding image. The discriminator has x and $G(z)$ as inputs and is composed of three 3x3 Convolution-BatchNorm-LeakyReLU layers, being trained to discriminate between real and synthetic data. The model is implemented using the Adam optimizer, with a learning rate of 0.002.

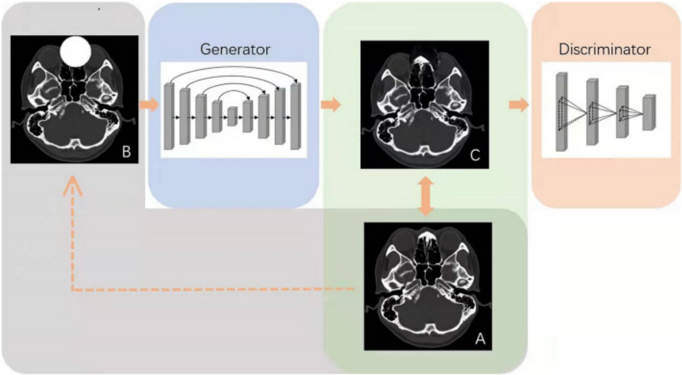

Figure 10: Basic architecture of the GAN. (A) Image with healthy bone anatomy. (B) Image with a manually created artificial bone defect. (C) The generator produces the image with the reconstructed medial facial bone. The discriminator ascertains the similarity between the normal image (A) and the reconstructed image (C) [42].

The evaluation content involves anatomical similarity, the continuity of the border, and whether the global shape meets physiological and aesthetic requirements. Firstly, the reconstructed images are compared with the real images through the cosine similarity calculation. This indicator evaluates the similarity between two objects, the closer to 1 the higher the accuracy. Afterwards, the 3D models of the images are imported into a software that enables the registration and calculation of the mean error. The study revealed a cosine similarity of 0.97 ± 0.01 mm and a mean error of 0.59 ± 0.31 mm in the virtual reconstruction of the artificial bone defects. Furthermore, it demonstrated that the discrepancy between the results of reconstructions with artificial defects and those with clinical bone defects is not statistically relevant. A cosine similarity of 0.96 ± 0.01 mm and a mean error of 0.48 ± 0.08 mm were obtained for the clinical data. This technology has greater accuracy, stability, and feasibility than SSM-based methods.

The article claims that this method can be employed in other bones and digital imaging techniques, as well as, the GAN-based technologies exhibit a higher accuracy compared to existing ones and have the potential to become the gold standard in bone defect reconstruction.

3

RESEARCH AND DEVELOPMENT METHODOLOGIES

This chapter provides a comprehensive overview of the technologies implemented, that aim to achieve the objectives defined in Chapter 1. In the context of GDL, it focuses on the specifications of GANs and identifies the models that satisfy the project requirements. Additionally, section 3.3 discusses the technical challenges of GANs.

3.1 Generative Deep Learning

In recent years, with the constant scientific and technological innovations, exponential growth has been observed in the application of Deep Learning for the analysis of medical images. First of all, it is important to distinguish discriminative models from generative models. Discriminative models learn a conditional distribution, $p(y|x)$, with x as the input variable and y as the target variable. As an example, given a medical image, in order to segment anatomical structures, a convolutional neural network tries to discriminate images or voxels that correspond to distinct classes. In contrast, generative models learn a joint distribution, $p(y,x)$ or $p(x)$. These models aim to learn the underlying distribution of the data and its generative process, through the factors that regulate the process and the natural properties of the dataset. If the model training is successful, it allows the generation of new data by sampling. The most popular approach in this domain is Generative Adversarial Networks (GANs), the subject of our analysis [43, 44].

3.2 Generative Adversarial Networks

The essence of GANs consists in optimizing two neural networks with different goals, i.e., in Adversarial Training (Figure 11) [44].

The first network corresponds to a generative model G that learns the probability distribution of the input data (p_{data}) and, by transforming noise vectors z with probability distribution p_z, attempts to generate samples, perceptually, similar ($x_{fake} = G(z)$). The p_z consists of a latent space with low dimensionality, such as the Gaussian distribution. The second neural network represents a discriminative model D that seeks

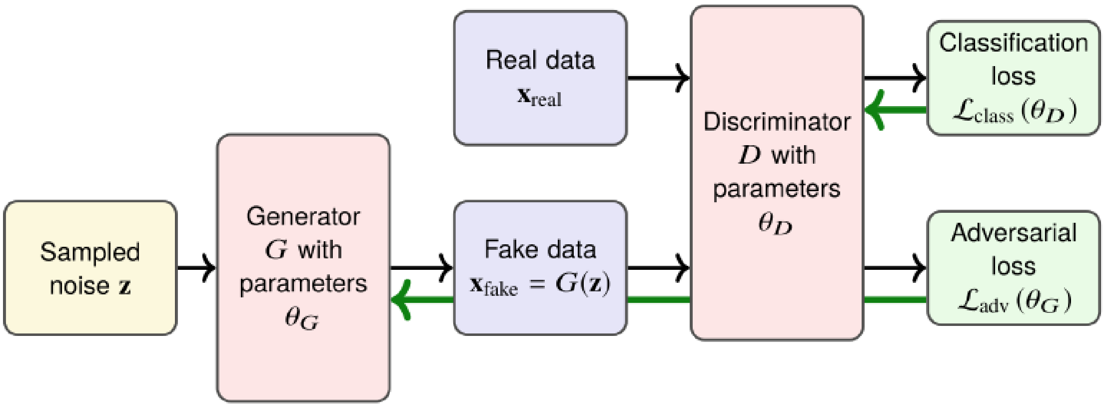

Figure 11: The diagram represents the basic architecture of a GAN. The green arrows indicate the flow of gradients. The training process of the discriminator considers the classification loss and the generator training process takes into account the adversarial loss, which also depends directly on the discriminator's performance [43].

to differentiate between samples from the real dataset and the artificial samples produced by the generator, with a probability distribution p_g, from the calculation of $D(x_{real})$, which represents the probability of the sample being a resultant of the input data. In general, the discriminator construction is based on binary classification. The last layer of the discriminator includes a sigmoid activation function, such that $D(x_{real})$, $D(G(z)) \in [0,1]$, where $D(G(z))$ is the probability that the image is derived from the generator [43, 44, 45, 46].

The training process for the generator is to maximize the probability of the discriminator making an error, if it can easily distinguish samples, the generator weights need to be adjusted, accordingly [44, 45]. On the other hand, the discriminator is trained in order to maximize the probability of accurately determining the origin of the image [45]. The competition between the two models allows them to improve their methods until the convergence between real and synthetic samples ($p_g = p_{data}$) [43, 45]. The system can be trained by backpropagation if the neural networks consist of at least three layers [45].

Mathematically, the GANs are trained to solve the following optimization problem, which the discriminator tries to maximize and the generator to minimize:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] , \tag{1}$$

where $V(D,G)$ represents the objective function, under study. The generator aims to minimize the possibility of the discriminator correctly identifying the generated images as false. In opposition, the discriminator minimizes the error that is associated with the classification of the samples, maximizing the objective function [44]. Denote that in the elementary formulation $\mathcal{L}_{adv}$ and $\mathcal{L}_{class}$ are practically the same function, however, optimized in different directions[43].

In practice, the optimization problem turns out to be very unstable for the generator, as a result the saturation of $\log(1 - D(z)))$ occurs and the convergence between p_g and p_{data} is extremely slow. At the beginning of training, the discriminator is able to easily specify the provenance of the image, in this scenario, the gradient for $\log(1 - D(z)))$ is approximately zero. Therefore, the generator does not optimize its parameters and minimize its loss [43, 44, 45]. Therefore, an alternative loss function for the generator has emerged:

$$\min_G V(D, G) = -\,\mathbb{E}_{z\sim p_z(z)}\left[\log(D(G(z)))\right]. \tag{2}$$

As mentioned earlier, equation (1) intends to minimize the probability of the discriminator identifying the generated samples as false. In contrast, the goal of equation 2 is to maximize the probability that the discriminator recognizes the generated samples as real [44]. This objective function allows the intensification of the gradients at the beginning of the training [45].

However, model training may remain unstable, mainly due to the way the discriminator calculates the discrepancy between p_g and p_{data}. In the original definition (Equation 1), the difference is estimated as the Jensen-Shannon divergence. Unfortunately, in the presence of two disjoint distributions, this type of symmetric divergence is not effective. As a result, several alternative objective functions have been proposed to optimize the model, such as an objective function based on the Wasserstein distance (WGAN/WGAN-GP) [44].

The most widely employed models with the highest success rate in generating volumetric data are, in the following order, CycleGAN, cGAN (conditional GAN), DCGAN (Deep Convolutional GAN), and WGAN/WGAN-GP based architectures. Additionally, in the medical domain, the main applications of GANs are image translation and volumetric reconstruction, with the Cycle Generative Adversarial Network (CycleGAN) based architecture being the preferred choice for these purposes [47].

3.2.1 *Cycle Generative Adversarial Network*

The CycleGAN, given the training samples $x \in X$ and $y \in Y$, with probability distribution $p_{data}(x)$ and $p_{data}(y)$, respectively, aims to learn two mapping functions between two domains, $G : X \to Y$ and $F : Y \to X$ (Figure 12) [48].

The CycleGAN comprises a generator G that transforms images from the source domain X to the target domain Y, being the most relevant function of the model, and a generator F that converts the images from the target domain Y to the original domain X [43, 48]. The generator network seeks to minimize the difference between the input samples, from the respective domain, and the generated samples, as well as, to maximize the loss of the discriminators [44].

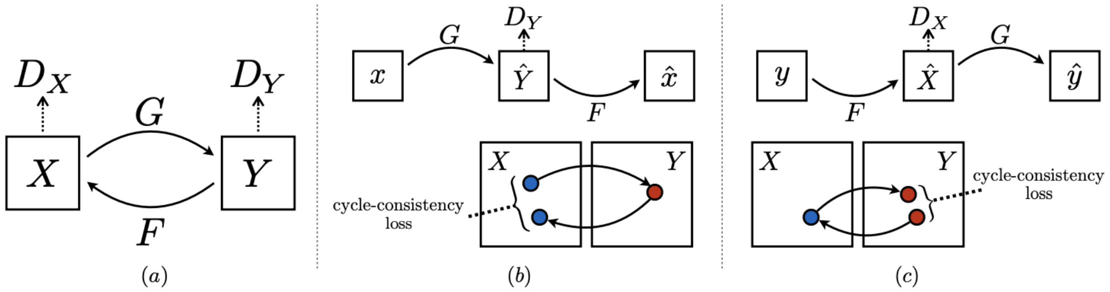

Figure 12: (a) The model is set out to train two mapping functions $G : X \rightarrow Y$ and $F : Y \rightarrow X$, consisting of two generators and two discriminators. In order to regularize the mappings, CycleGAN integrates two cycles. (b) Forward Cycle Consistency: $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ e (c) Backward Cycle Consistency: $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$ [48].

In addition, the model has a discriminator D_X that intends to differentiate between real images x and synthetic images $F(y)$ and a discriminator D_Y which acts, reciprocally, in domain Y, that is, it intends to discriminate between y and $G(x)$ [43, 48].

To ensure that the generated samples are equivalent to the actual samples in the considered domain, the objective function applies adversarial losses to both mapping functions. For example, for the mapping function $G : X \rightarrow Y$ and for its discriminator D_Y:

$$\mathcal{L}_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{data}(y)} \left[\log D_Y(y)\right] + \mathbb{E}_{x \sim p_{data}(x)} \left[\log(1 - D_Y(G(x)))\right]. \quad (3)$$

The generator G, in equation (3), aspires to reproduce images similar to the images in domain Y. The D_Y differentiates between the artificial samples $G(x)$ and the input samples y. It is important to note that a loss function is also employed for the mapping function $F : Y \rightarrow X$ and its discriminator D_X:

$$\mathcal{L}_{GAN}(F, D_X, Y, X) = \mathbb{E}_{x \sim p_{data}(x)} \left[\log D_X(x)\right] + \mathbb{E}_{y \sim p_{data}(y)} \left[\log(1 - D_X(F(y)))\right], \quad (4)$$

where the generator F seeks to conceive samples identical to the samples in domain X. The discriminator D_X determines the source of the images, i.e., whether the samples come from generator $F(y)$ or from the set of real data x [43, 44, 48].

To prevent mapping functions G and F from contradicting each other, the objective function contains loss functions that evaluate the consistency of the cycle since the adversarial losses cannot guarantee, independently, the translation of the image domain:

$$\mathcal{L}_{cycle}(G, F) = \mathbb{E}_{x \sim p_{data}(x)} \left[\|F(G(x)) - x\|_1\right] + \mathbb{E}_{y \sim p_{data}(y)} \left[\|G(F(y)) - y\|_1\right]. \quad (5)$$

The model infers that an image x from the domain X that is converted to an image $G(x)$ in the domain Y can be reconstructed, $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$, this step is called Forward Cycle Consistency. Similarly, for each image y of the domain Y, the generator network performs Backward Cycle Consistency: $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$ [48]. One of the goals of the generator is to minimize the discrepancy between the original image and its reconstruction in the original domain [44].

Despite not being considered an essential loss function in the official objective function, the generator can be regularized by an identity mapping loss, when real images of the target domain are given as input to the generator:

$$\mathcal{L}_{identity}(G,F) = \mathbb{E}_{y \sim p_{data}(y)} \left[\|G(y) - y\|_1 \right] + \mathbb{E}_{x \sim p_{data}(x)} \left[\|F(x) - x\|_1 \right]. \tag{6}$$

The generator network, without $\mathcal{L}_{identity}$, can inappropriately alter the tonality of the samples [48].

In this way, the complete objective function is obtained, taking into account the loss of identity:

$$\begin{aligned} \min_{G,F} \max_{D_X,D_Y} \mathcal{L}(G,F,D_X,D_Y) = \mathcal{L}_{GAN}(G,D_Y,X,Y)+ \\ \mathcal{L}_{GAN}(F,D_X,Y,X)+ \\ \lambda \mathcal{L}_{cycle}(G,F)+ \\ 0.5\lambda \mathcal{L}_{identity}(G,F), \end{aligned} \tag{7}$$

the term λ represents the relative importance of the losses that evaluate cycle consistency, as well as, the identity mapping loss, comparatively, to the Adversarial loss functions. According to the literature, $\lambda = 10$ [48].

3.2.2 *Conditional Generative Network*

Generally, regular GAN does not enable control over the features of the generated samples. In practice, it can be advantageous to have more control over what is represented in the generated images [44]. GANs can be extended to a conditional model, both the generator and discriminator are conditioned with extra information that is fed as an additional input layer [49]. In a conditional GAN (Figure 13), the generator uses the information in y in addition to the noise vector z to generate plausible samples. On the other hand, the discriminator evaluates whether the synthetic samples resemble the real images with probability distribution p_{data}, given the information provided in y [44]. A cGAN can use the same loss functions as a standard GAN [43].

The objective function can be represented by the following equation:

$$\min_{G} \max_{D} V(D,G) = \mathbb{E}_{x \sim p_{data}(x)} \left[\log D(x|y) \right] + \mathbb{E}_{z \sim p_z(z)} \left[\log(1 - D(G(z|y)|y)) \right]. \tag{8}$$

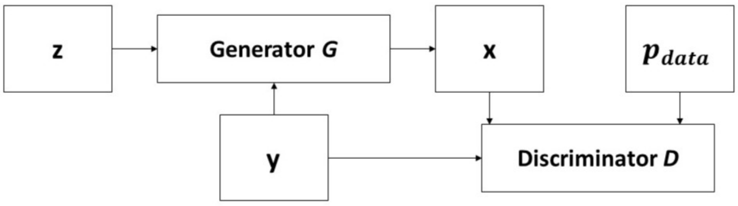

Figure 13: In a conditional GAN, the generator and discriminator receive an additional conditioning vector y as input. Adapted from [44].

3.2.3 *Pix2Pix Generative Adversarial Network*

Pix2pix (Figure 14) was proposed by Isola et al. ([50]) in 2017, this model refers to a conditional GAN (cGAN). It is the most straightforward approach for translating images from one domain to another, as for each image x in the domain X there is a reference image y' in the domain Y, i.e., it requires a paired dataset [43, 44, 50].

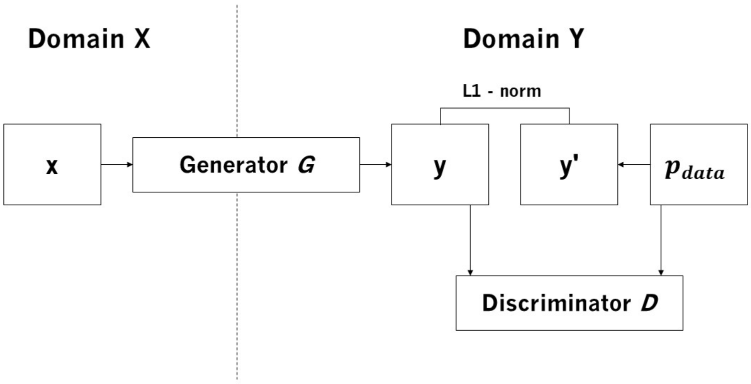

Figure 14: The generator G minimizes the voxelwise loss between its output y and a reference image y'. The discriminator D provides an adversarial loss that represents how y matches the real images in the domain Y. Adapted from [44].

There are two possibilities for training an image translation model. In the first case, a reference image in domain X is available for each image in domain Y (Pix2Pix). In the second situation, the images in domain X are semantically related to the images in domain Y, however, there is not necessarily a reference image in domain Y for each image in domain X. In other words, training can be performed with unpaired data (CycleGAN) [44]. The main difference between the CycleGAN architecture and the Pix2Pix architecture is the addition of the cycle consistency loss to enable training without the need for paired data. Therefore, a CycleGAN translates from the source domain to the target domain without a one-to-one mapping, as opposed to Pix2Pix [51].

As the CycleGAN, this model uses an adversarial network to assess if the predictions are, perceptually, similar to the real images. The generator network is trained to minimize the voxelwise discrepancy between its output and the corresponding reference image [44]. The discriminator receives two images, the real image or the synthetic image and the condition (reference image), and seeks to differentiate between the real images of the target domain and the generated images conceived by the generator [43].

The objective function can be represented by:

$$\begin{aligned}\min_{G}\max_{D} V(D,G) = \lambda_1 \mathbb{E}_{y\sim p_{data}}\left[\log D(y,x)\right] + \\ \mathbb{E}_{x\sim p_x}\left[\log(1 - D(G(x),x))\right] + \\ \lambda_2 \mathbb{E}_{y'\sim p_{data},x\sim p_x}\left[\|y' - G(x)\|_1\right],\end{aligned} \tag{9}$$

where the images x and y' are spatially aligned. When $\lambda_1 > 0$ and $\lambda_2 > 0$, the output image $G(x)$ is connected to the reference image y' and the target domain through L1-norm and the adversarial loss [44].

CycleGAN's ability to learn from unpaired datasets makes it attractive in scenarios where acquiring paired images for training is difficult, such as in the medical domain. On the other hand, Pix2Pix is a conditional model that requires paired datasets for supervised learning. It learns a direct mapping from input images to output images based on the paired examples provided during training. Pix2Pix is suitable when paired data is available and can provide more precise control over the translation process. Accordingly, the performance of these two architectures will be investigated.

3.3 Technical Disadvantages [47]

Technically, GANs can exhibit some disadvantages:

- Mode collapse: if perchance, the discriminator is fixed at a local minimum, the generator learns that it is plausible to conceive the same set of results to fool the discriminative network.

- Imperception: given the heterogeneous applications of GANs, it is challenging to compare the models without human intervention, since there are no loss functions and evaluation metrics capable of mimicking human judgment.

- Non-convergence: the convergence point may be defined as the point where the discriminator is no longer effective in differentiating between real data and synthetic data, since the generator, progressively, produces realistic samples. Given this, the discriminator's feedback is not consistent, causing the results to deteriorate.

- Diminished gradient: On the other hand, if the discriminator outperforms the generator, in other words, if the generator does not follow the evolution of the discriminator, its feedback becomes useless.

4

CASE STUDY

This chapter describes in detail the organization, analysis, and preparation of the dataset to satisfy the implementation requirements. The generator and discriminator architectures, which are crucial to achieving the intended results, are described in this chapter. Likewise, the training details of both models to obtain high-quality 3D images are also addressed. Section 4.4 presents the metrics applied to evaluate the performance of the proposed methods. Finally, section 4.5 describes the qualitative test that was designed and carried out in order to evaluate the condition of the glenoid cavity. That is, to validate if the estimated glenoid version presents a healthy value since the aim of shoulder arthroplasty is to accurately restore the patient's healthy anatomy.

4.1 Dataset

The dataset consists of 204 3D CT scans of the glenohumeral joint, with a shape of 512x512xn° of slices, with the n° of slices being variable. Its categorization was based on the modified Walch classification, which describes morphological alterations of the arthritic glenoid on 3D CT, as previously mentioned in subsection 2.2.1. If the glenoid cavity presented a pathological retroversion or anteversion, >10° and >5°, respectively, it would be considered defective. In other words, if it exhibited bone defects of type B, C, and D, the most likely to be submitted to reconstructive surgery. For this purpose, an algorithm was developed, which applies Friedman's line method to calculate the glenoid version from the landmarks associated with each medical image in the dataset. The results of the classification revealed that the dataset was not balanced, as only 93 of the medical images corresponded to unhealthy scapulae and 111 to healthy scapulae. The images of scapulae with signs of glenohumeral arthropathy will only be used to test the models.

Besides the images, the dataset contains the labels associated with each bone and the relevant anatomical landmarks, annotated and reviewed by independent professionals.

4.1.1 *Data Preprocessing*

Data preprocessing is a crucial step in developing a successful deep learning model, as the quality and relevance of the data directly impact the training process. To achieve the objectives of this study, the 3D CT segmentations were manipulated, with a specific focus on the patient's scapulae, by keeping only the labels 5 (right scapula) and 6 (left scapula). The left scapulae were submitted to a vertical flip, so that the images presented the same orientation. The images were cropped around the area of interest and resized to a standard configuration of 64x64x64 for the CycleGAN and 128x128x128 for the Pix2Pix. The difference in image resolution is due to the fact that CycleGAN has a higher computational cost. Finally, the glenoid cavity was removed in each healthy domain image by applying a spherical mask with a radius of 50, so that the networks learn to reconstruct healthy scapulae from the images of interest (Figure 15).

Figure 15: Example of a scapula with the glenoid fossa removed.

Concisely, two distinct domains are obtained, the unhealthy domain (X) and the healthy domain (Y). The generative model (CycleGAN or Pix2Pix) seeks to learn the mapping function between these two domains, $G : X \rightarrow Y$, i.e., the conversion of an image x from the domain X to an image $G(x)$ from the domain Y, where x is an image of a scapula with the glenoid removed and $G(x)$ the sample produced by the generator, which wants to generate images similar to those of the real dataset of the healthy domain.

4.2 Networks Architecture

The architecture of the generator and the discriminator network is identical for both models, except that Pix2Pix's discriminator takes the reference image into account. For each discriminator network (Figure 16), a PatchGAN was adopted for the classification of overlapping image patches (subvolumes) into real or fake, instead of classifying the whole volume. In a PatchGAN, the input image is passed through a series of convolutional layers, each of which downsamples the image and increases the number of filters. The final convolutional layer produces a single output value for each local image patch [48, 52, 53]. In this case, the last layer of the discriminator includes a sigmoid activation function, such that $D(x_{real})$, $D(G(x)) \in [0, 1]$, where a value close to 1 indicates that the subvolume is real, and a value close to 0 indicates that the subvolume is fake. The PatchGAN discriminator approach limits the ability of the discriminator to use arbitrary information from unexpected locations within the image to make its decisions [53]. Gaussian noise was added in the input of the discriminator in order to prevent overfitting of the training data and force the discriminator to learn more robust features. This can lead to a more stable and effective training process of the GAN [54, 55].

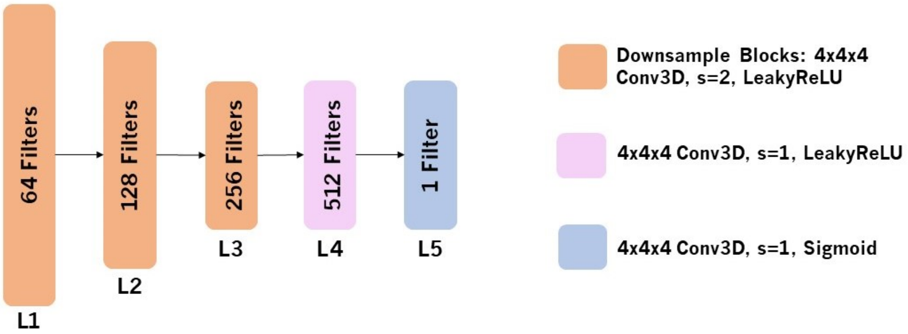

Figure 16: Diagram of the discriminator architecture.

The architecture of the discriminator (Table 2) of the first three layers is composed by three downsample blocks, that reduce the spatial dimensions of the feature maps while increasing the number of filters. Each downsample block consists of a 3D convolutional layer with a stride of 2, a kernel size of 4, an instance normalization, except for the first layer, and a LeakyReLU activation function with a slope of 0.2. The fourth layer is similar to the previous ones but with a stride of 1. The last layer is a 3D convolution with a stride of 1, a kernel size of 4, and as abovementioned, includes a sigmoid activation function.

Table 2: Architecture used for the discriminators

Layer	Layer Type	Filters	Filter Size	Stride	Activation Function	Normalization
1	Downsampling (Conv3D)	64	4	2	LeakyReLU(0.2)	None
2	Downsampling (Conv3D)	128	4	2	LeakyReLU(0.2)	Instance Normalization
3	Downsampling (Conv3D)	256	4	2	LeakyReLU(0.2)	Instance Normalization
4	Conv3D	512	4	1	LeakyReLU(0.2)	Instance Normalization
5	Conv3D	1	4	1	Sigmoid	None

In a CycleGAN, the generator architecture (Figure 17) plays a crucial role in the quality and realism of the generated output. In this case, the generator is responsible for mapping the unhealthy domain to the healthy domain and creating realistic samples of healthy scapulae.

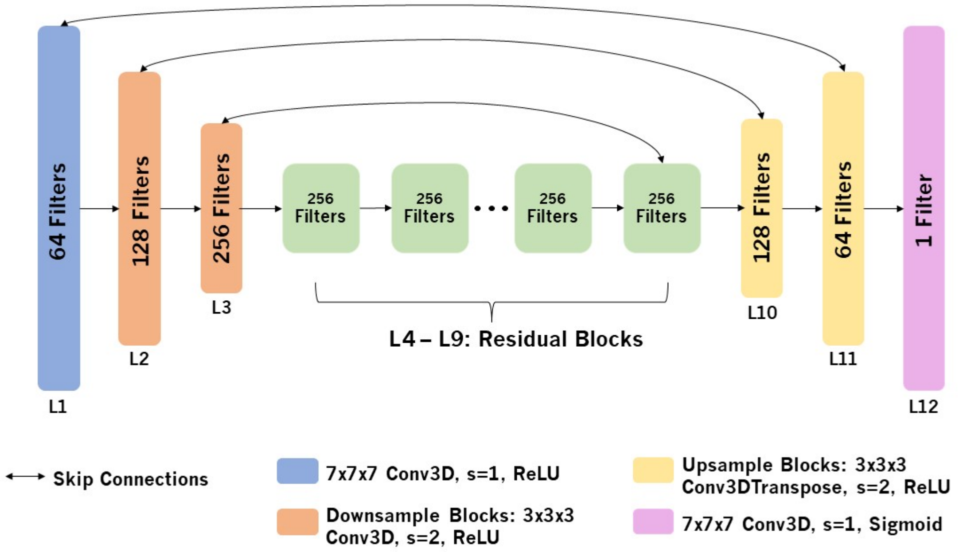

Figure 17: Diagram of the generator architecture.

The generator networks are expected to be stable and capable of converging to a good solution. Table 3 presents the generator architecture of the proposed solution.

Table 3: Architecture used for the generators

Layer	Layer Type	Filters	Filter Size	Stride	Activation Function	Normalization
1	Conv3D	64	7	1	ReLU	Instance Normalization
2	Downsampling (Conv3D)	128	3	2	ReLU	Instance Normalization
3	Downsampling (Conv3D)	256	3	2	ReLU	Instance Normalization
4-9	Residual (Conv3D)	256	3	1	ReLU	Instance Normalization
10	Upsampling (Conv3DTranspose)	128	3	2	ReLU	Instance Normalization
11	Upsampling (Conv3DTranspose)	64	3	2	ReLU	Instance Normalization
12	Conv3D	1	7	1	Sigmoid	None

The first layer of the generator architecture consists of a 3D convolutional layer with a stride of 1, a kernel size of 7, an instance normalization, and a ReLU activation function. Then, two downsample blocks are applied sequentially, these incorporate a 3D convolution with a stride of 2, a kernel size of 3, an instance normalization, and a ReLU activation function. Subsequently, there are six residual blocks, each block includes two 3D convolutional layers with a kernel size of 3, an instance normalization, and a ReLU activation function. After the residual blocks, a skip connection is applied between the output of the last residual block and the output of the second downsample block. Two upsample blocks are implemented, which increase the spatial dimensions of the feature maps while decreasing the number of filters. Each upsample block is composed by a 3D transposed convolutional layer with a stride of 2, a kernel size of 3, an instance normalization, and a ReLU activation function. Skip connections are applied between the output of the first upsample block and the output of the first downsample block and between the output of the last upsample block and the first convolutional layer. Finally, the last layer consists of a 3D convolutional layer with a stride of 1, a kernel size of 7, and a sigmoid activation function.

According to the literature, the first and last layers are critical for the preservation of anatomical structures in medical imaging [53]. Moreover, the addition of skip connections allows for low-level information from the input to be fed directly to the output, preserving and reconstructing details from the source images. These are believed to be particularly important to reduce the loss of high-resolution information in deeper layers of the network. Overall, the use of skip connections can enhance the preservation of detailed information in images and improve the quality of the generated images [56].

4.3 Training Details

Pre-training the generator has been demonstrated to improve the performance of both networks. The generator was pre-trained for 150 epochs, using the Binary Cross Entropy loss function and the Adam optimizer with a learning rate of 2e-4.

4.3.1 *CycleGAN*

The CycleGAN was trained for 150 epochs. The weights of the pre-trained generator were used to initialize the CycleGAN generators. The batch size, learning rate, and momentum parameters values used were those suggested in the literature for the CycleGAN, 1, 2e-4, $\beta_1 = 0.5$ and $\beta_2 = 0.999$, respectively [48].

It should be noted that during the training step it is necessary to define different loss functions to guide the model learning process:

- Generator loss — computes the Mean Squared Error (MSE) between the generated images and a tensor of ones (target output). This loss depends, directly, on the performance of the discriminator, and by minimizing the generator's loss function, it will produce more realistic samples [44].

- Discriminator loss — computes the MSE between the real and generated images and tensors of ones and zeros, respectively. By minimizing this loss, the discriminator learns to better distinguish between real and generated images.

- Cycle consistency loss — calculates the mean absolute difference between the real image and its reconstruction. This loss function helps to ensure that consistent samples are generated across domains, and prevents the generators from producing contradictory samples [48].

- Identity loss — calculates the mean absolute difference between the real image and the output image. By minimizing this loss function the CycleGAN ensures that the output content is consistent with the input [48].

4.3.2 *Pix2Pix*

The Pix2Pix was trained for 100 epochs. The weights of the pre-trained generator were used to initialize its generator. The batch size, learning rate, and momentum parameters values used were those suggested in the literature for the Pix2Pix, 1, 2e-4, $\beta_1 = 0.5$ and $\beta_2 = 0.999$, respectively [50].

To guide the learning process, the generator loss and discriminator loss, described in the previous subsection, were applied. Additionally, the generator loss function includes the calculation of the mean absolute difference between the desired output and the generated image. This is denominated L1-norm and provides a measure of dissimilarity between the two samples, encouraging the generator to generate samples similar to the target images.

Furthermore, in both networks, the objective function was divided by 2 during the optimization of the discriminators, which decreases the learning rate of the discriminators compared to the generators [50].

4.4 Evaluation Metrics [47]

The following metrics were applied to evaluate the performance of the proposed methods:

- Peak signal-to-noise ratio (PSNR) — metric, commonly, applied to evaluate reconstructions, one of the main applications of volumetric GANs. It evaluates the quality of images/volumes and for this reason is widely used in medical imaging. The higher the PSNR, the better the quality of the reconstructed image [57].

- Structural similarity index measure (SSIM) — the second most used metric in reconstruction problems. It assesses the quality of the data, taking into account human perception, that is, it extracts structural information, such as human vision. A value closer to 1 indicates a better image quality [58].

- Visual — visual evaluation is one of the best tools for evaluating the images generated by GANs and is broadly utilized as a metric in CycleGAN-based architectures.

- Dice similarity coefficient — a statistical tool that determines the similarity between two datasets. Despite being mostly used for model evaluation of segmentation algorithms, it is a general concept that can be used in many applications [59].

In the generator pre-training, apart from the metrics, mentioned above, the accuracy of the healthy scapula reconstruction was calculated. In the training phase of CycleGAN and Pix2Pix, the SSIM and PSNR metrics were determined between the real images of the healthy domain and the generated images. The dice similarity coefficient was also computed between the target image and the generated output. Furthermore, all reconstructions were evaluated visually to assess the quality and reliability of the results.

In the test step, the SSIM and PSNR metrics provide an overall assessment of the similarity and quality of the generated images compared to the real images in the test dataset.

4.5 Qualitative Test

As mentioned previously, an abnormal deviation of the glenoid fossa version can affect the biomechanics and stability of the glenohumeral joint. A pathological version was proven to increase stress and wear on the glenoid component. Therefore, it is crucial to normalize the glenoid version during shoulder arthroplasty to improve the functional outcome and longevity of the implant [17]. Consequently, to evaluate the condition of the glenoid in the acquired reconstructions, the relevant landmarks were manually annotated to calculate the glenoid version (Figure 18). Then, the algorithm used in the classification of the dataset was applied, which employed Friedman's line method to evaluate the version of the glenoid.

Figure 18: Example of a manual annotation of landmarks.

5

RESULTS

This chapter presents the metrics obtained to evaluate the performance of the models in the reconstruction of glenoid bone defects. The focus is on presenting the quantitative measures achieved to evaluate the accuracy and efficiency of the models. Additionally, the chapter includes qualitative test results, which are crucial for evaluating the medical validity and clinical applicability of the obtained outcomes. By combining quantitative metrics and qualitative assessments, this chapter provides a comprehensive overview of the models' performance and their potential in the field of glenoid bone defect reconstruction. The dataset used in this study was divided into three subsets: the training set, the validation set, and the test set. The training set consisted of 111 images of healthy scapulae. The validation set comprised 10% of the training set. Finally, the test set comprised 24 images of unhealthy scapulae, which were reserved for evaluating the performance of the models in reconstructing glenoid bone defects. It is important to note that the validation and test sets in this study did not include volumes from the same individuals.

Table 4 presents the results of the generator pre-training. By observing the values obtained for each metric, it can be inferred that the training and validation phases were both successful. Moreover, a training loss of 7.4354e-4 and a validation loss of 0.0040 were obtained. Visually, the reconstructions appeared to be correct.

Table 4: Generator pre-training results

Generator Pre-training	**Train**	**Validation**
SSIM	0.9976	0.9946
PSNR	75.5020	77.8550
Dice similarity coefficient	0.9942	0.9838
Accuracy	0.9997	0.9990

Table 5 shows the relevant evaluation metrics achieved during the training and testing of CycleGAN. In the training phase, the G and F generators obtained losses of 0.3747 and 0.4204, respectively. On the other hand, the X and Y discriminators got losses of 0.0709 and 0.0884, in that order. Visually, only a few promising reconstructions were registered (Figure 19).

Table 5: Results of the training and test steps of the CycleGAN model

CycleGAN	**Train**	**Test**
SSIM	0.9820	0.9740
PSNR	53.5711	27.6460
Dice similarity coefficient	0.9621	–

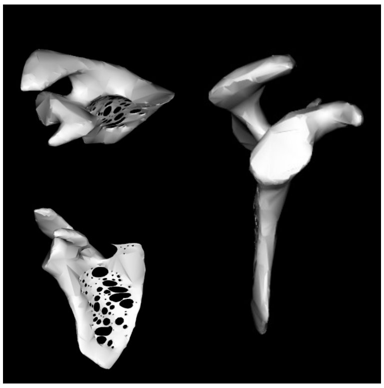

Figure 19: Example of a reconstruction after testing the CycleGAN model.

Table 6 presents the pertinent evaluation metrics achieved during Pix2Pix training and testing. The generator and the discriminator exhibited losses of 0.2502 and 0.2500, in that order. Most of the reconstructions seemed, visually, correct (Figure 20).

Table 6: Results of the training and test steps of the Pix2Pix model

Pix2Pix	**Train**	**Test**
SSIM	0.9987	0.9902
PSNR	114.6688	29.1981
Dice similarity coefficient	0.9971	–

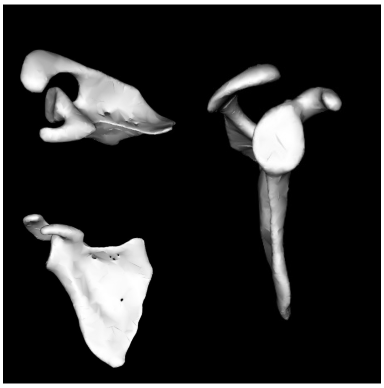

Figure 20: Example of a reconstruction after testing the Pix2Pix model.

It should be emphasized, during training, the SSIM and PSNR metrics correspond to the values determined between the real images of the healthy domain and the generated images of healthy scapulae. In the test, the metrics were calculated between the real images (unhealthy scapulae) and the generated images, samples produced by the generator, which wants to generate images similar to those of the real dataset of the healthy domain. Additionally, the SSIM and PSNR test values refer to the mean values obtained after testing the test dataset. During testing, the calculation of the dice similarity coefficient was not performed due to the absence of a target image for reconstruction.

5.1 Qualitative Test Results

The primary goal of this study is to reconstruct glenoid bone defects so that the predicted version is comprised in the range [-5°, 10°] since the purpose of shoulder arthroplasty is to accurately restore the patient's healthy anatomy. Each test phase consisted in evaluating the performance of the model using a dataset of 24 images of unhealthy scapulae. It is important to note that these images were submitted to the same preprocessing as the training images of the respective model, including the removal of the glenoid cavity.

The Pix2Pix model successfully reconstructed the glenoid bone defects in the majority of cases. Out of the 24 artificial bone defects, the model was able to accurately reconstruct 21 cases, resulting in a success rate of 87.5%. These reconstructions presented a healthy version within the desired range of [-5°, 10°], indicating the model's ability to closely approximate the normal glenoid version. However, there was one case where the model generated a reconstruction that did not fit the healthy range. Considering the normal version as 0° (true value), a Mean Absolute Error (MAE) of 2.6875 was obtained.

Unfortunately, the CycleGAN model results are insufficient to assess its quality. The absence of a reference image imposes limitations on the ability of the CycleGAN model to learn and capture the specific features required for glenoid bone defect reconstruction, i.e., as a model designed to handle unpaired data, it cannot directly compare the original and target domains. Furthermore, the reliance on a small dataset further amplifies this restriction, as the model has fewer examples to learn from. Consequently, the model may have difficulty in effectively learning the mapping functions necessary for generating high-quality images.

6

RESULTS ANALYSIS AND DISCUSSION

Virtual reconstruction technology provides a crucial reference for the reconstruction of glenoid bone defects, irrespective of their primary etiology, facilitating preoperative planning of reconstructive surgical interventions. The positioning of the glenoid component in excessive retroversion leads to altered biomechanics and an increase in glenoid shearing forces. It has been proven that a retroversion greater than 10° will result in more than 5 mm of posterior translation of the humeral head. Additionally, a glenoid placed in 15° of retroversion causes an increase in the peak strain, liftoff, and slippage of the implant, which can lead to its loosening and failure. Based on previous studies, the goal of surgery is to restore the glenoid version to within 10° of a patient's native glenoid version [40].

The virtual reconstruction techniques present in clinical practice and literature manifest some drawbacks. The mirroring of the contralateral joint, the gold standard, can only be applied in the case of unilateral arthropathy and assumes that the scapula is symmetrical, which can lead to errors in the glenoid version calculation. It should be noted that the contralateral joint must not have been subject to previous surgeries and injuries [14, 39]. The use of a standard model of the healthy glenoid does not consider the individual anatomical variability of the scapula. The results obtained from the linear regression models of Resh's angle and AGWA are not reliable, compared to the conventional technique [38, 40]. On the other hand, the SSM relies on the process of registering the shape of the bone structure. Consequently, there is a need to develop an intelligent solution to reconstruct the glenoid healthy bone, to improve clinical practice [14].

Recently, there has been an increasing interest in the application of CycleGAN for 3D image-to-image translation, particularly, in the field of medical imaging. In the context of healthy bone reconstruction, a CycleGAN-based approach can be used to generate a healthy bone image from an unhealthy bone image (an image of a scapula with the glenoid removed), preserving the underlying anatomical structure and features of the original image. The training process involves using two GANs that are trained in a cyclic way to learn the mapping between two distinct domains, in this case, the unhealthy bone domain and the healthy bone domain. On the other hand, Pix2Pix, previously mentioned, is an image translation model, just like CycleGAN. However, it performs training with paired data, i.e., it translates from the source domain to the target domain via a one-to-one mapping, unlike CycleGAN. The major difference between the CycleGAN architecture and the Pix2Pix architecture is the addition of cycle consistency loss to allow training with unpaired data. In both models, during training, the generators are optimized to minimize the discrepancy between the generated images and the real images, while the discriminators are adjusted to discern between real and synthetic images. For image translation and volumetric reconstructions, the CycleGAN-based architecture is the preferred choice for such ends, but the study dataset is insufficient to evaluate this model at its full potential. Thereby, the performance of Pix2Pix will also be examined, since it takes into consideration a reference image at the training step and, therefore, the size of the dataset may not be so critical to achieving good results.

The pretrained generator used to initialize the CycleGAN and Pix2Pix generators, achieved a high SSIM and dice similarity coefficient of 0.9976 and 0.9942, respectively, indicating a strong similarity between the generated and reference images. The high PSNR value of 75.5020, implies that the generated images present a low level of noise and distortion compared to the target images. The validation results show that the pretrained generator can generalize to unseen data.

During training and testing, the Pix2Pix model attained a high SSIM of 0.9987 and 0.9902, in that order, suggesting a great structural similarity between the generated images and the real images. The training phase yielded a high PSNR of 114.6688, indicating a low level of noise and distortion in the generated images compared to the target images. Nevertheless, in the test phase, the PSNR decreased, significantly, to 29.1981, pointing to a higher level of noise and distortion in the generated images. The dice similarity coefficient reached a high value of 0.9971 during the training of the Pix2Pix model, this value suggests that the model can preserve structural information and details. The generator and discriminator losses of 0.2502 and 0.2500, respectively, indicate a balanced and competitive training process between the two GAN components. Overall, the results show that the Pix2Pix model is able to translate images from the unhealthy domain to the healthy domain.

During the training and testing processes, the CycleGAN model achieves an SSIM of 0.9820 and 0.9740, respectively, suggesting a substantial similarity in structure between the generated images and the real images. In the training phase, the model attained a PSNR of 53.5711, which indicates a low level of noise and distortion between the generated images and the target domain images. However, in the test step, the PSNR value drops notably to 27.6460. The dice similarity coefficient of 0.9621 attained during the training period is typically considered a good value, suggesting a proper alignment of the generated images with the real images of the healthy domain. However, it is important to note that this value primarily reflects the overall similarity of the entire scapula rather than specifically focusing on the glenoid region, that is, this metric may not be reliable. During the training phase, the G and F generators obtained losses of 0.3747 and 0.4204, respectively, while the X and Y discriminators achieved losses of 0.0709 and 0.0884, correspondingly. These loss values imply that the discriminators had a better performance in comparison to the generators. The slightly higher loss values of the generators indicate that they had difficulty generating images similar to those in the healthy domain.

In the testing phases of the Pix2Pix and CycleGAN models, there was a noticeable decrease in PSNR values, indicating that there may be a loss in the detail and quality of the images generated during the translation process. This may be due to the usage of images with glenohumeral arthropathy, which may have a different probability distribution.

In the case of the Pix2Pix model, most of the reconstructions seemed, visually, correct, as opposed to the CycleGAN model, where only a few promising reconstructions were registered. Visual image analysis is the most, commonly, used and one of the best tools to evaluate images generated by GANs. Visual assessment of the quality of the images may not be the best option, as the person in question may not have the necessary skills to analyze medical images. Therefore, it depends on expert opinion, which can be expensive, time-consuming, and not always easy to obtain [47].

The results of this study emphasize the promising performance of the Pix2Pix model in reconstructing glenoid bone defects and restoring the pathological version in unhealthy scapulae. The Pix2Pix model demonstrates its potential as a valuable technique in this domain, with a success rate of 87.5% and a MAE of 2.6875. It is important to note that a direct comparison with existing techniques was not conducted as the models have different evaluation criteria and practical limitations. Furthermore, it should be noted that due to the inaccessibility of the models, it was not possible to test and evaluate their performance using the specific dataset used in this study. Consequently, a direct comparison of the Pix2Pix model with these techniques was not feasible. Future investigations should be centered on comparing the Pix2Pix model with other established methods, particularly the mirroring of the contralateral joint and SSM (Table 1).

Although the study demonstrates the potential of the CycleGAN and Pix2Pix models for 3D image-to-image translation in medical imaging, it is crucial to recognize some limitations and areas that can be improved. The categorization of the dataset images, based on the modified Walch classification, may have led to inaccuracies in the results since it only considers one morphological parameter of the scapula. Similarly, the qualitative test that calculates the glenoid version to determine the condition of the glenoid also relies on the same classification algorithm. It is important to note that in this process, I annotated the 3D landmarks, which can introduce errors and impact the accuracy of the results. Ideally, these tasks should have been performed by an expert in the field of medical imaging, which can be a time-consuming and costly solution. The complexity of clinical cases, eventual artefacts, and the overfitting of the training set has the potential to impact the accuracy of the reconstructions [42]. The reduced size of the dataset amplified the limitations of the CycleGAN and Pix2Pix models. Regarding CycleGAN, the dependence on a smaller dataset enhances its constraints, since this model handles unpaired data and cannot compare directly between the source and target domains. This restriction can impact the ability of the model to learn the mapping functions and produce high-quality images. Likewise, even though Pix2Pix enables training with paired data, the small size of the dataset can affect its performance. Capturing the complexity of the reconstruction task can be challenging and lead to non-optimal results and limited generalization. Moreover, processing volumetric data requires a computer with high processing and storage capacity [47]. For example, the CycleGAN involves the training of four CNNs simultaneously and several filter responses need to be stored, for each layer of every CNN [52]. Thus, hardware limitation remains an adversity [47]. The model training and testing processes were conducted using an Amazon Web Services (AWS) EC2 instance type g5.xlarge (GPU 24GB RAM 16GB), taking advantage of its computational resources.

Looking at future perspectives, there are various potential areas for improvement and research in this study. The problems mentioned in section 3.3, i.e., mode collapse, non-convergence, and diminished gradient, are even harder to overcome when it concerns volumetric data. Likewise, due to the lack of an appropriate metric to compare the different models, it would be crucial to find a metric similar to human intelligence to enhance the results of volumetric data generation using GANs [47]. In addition, evaluating the performance of the models on a larger dataset would allow a more comprehensive assessment of their results.

Taking into consideration a significantly larger dataset, both models are expected to outperform in reconstructing a defective glenoid. This opens up a possibility for the development of an automated and intelligent virtual reconstruction tool that can be effectively used in clinical applications.

7

CONCLUSION

The present project was conducted in a company environment at PeekMed®, with the aim of investigating a possible solution for the reconstruction of glenoid cavity bone defects, contributing to the preoperative planning of shoulder arthroplasty.

The complexity and variability of the glenoid anatomy have posed significant challenges for the medical community, especially in reconstructive surgeries. Patients with osteoarthritis often present alterations in the morphological aspects of the glenohumeral joint, which may lead to loosening of the joint. Glenohumeral osteoarthritis is defined by progressive degeneration of the articular cartilage of the humeral head and glenoid component leading to pain, stiffness, and limitation of range of motion. Its diagnosis is based on medical history, physical examination, and imaging. Walch's classification is a method for categorizing the morphological alterations observed in arthritic glenoids and is widely acknowledged as the predominant system in the context of shoulder replacement procedures. Its primary objective is to emphasize the critical role of the glenoid version and the relationship between the humeral head and the glenoid in preoperative planning and surgical interventions. The optimal treatment for osteoarthritis of the glenohumeral joint is frequently controversial and is chosen according to the patient's age, symptoms severity, activity level, imaging results, and medical comorbidity. If nonsurgical measures are ineffective, particularly in the case of pain, surgical options are considered. Preoperative planning is crucial for the success of orthopedic procedures. Radiology, especially imaging studies, plays a vital role in this planning process. Medical imaging assesses important aspects of the glenohumeral joint, such as bone stock, glenoid version, GHSI, and HHM. These morphological parameters can be affected by degenerative osteoarticular pathology. Understanding the relationship between soft tissues and bone structures is also essential for effective preoperative planning.

The virtual reconstruction methods in literature and clinical practice have some limitations. Consequently, there is a need to develop an intelligent solution to reconstruct the glenoid healthy bone, to improve clinical practice. The performance of generative models was investigated, in this case, the CycleGAN and Pix2Pix.

Overall, the results emphasize the effectiveness of the Pix2Pix model in reconstructing glenoid bone defects, with a high success rate and an accurate restoration of a healthy version in the desired range, compared to the CycleGAN model.

The objectives defined in section 1.2 have been successfully accomplished. The main objective was the virtual reconstruction of the healthy anatomy of the glenoid cavity from 3D CT images, using GDL techniques. The quantitative aspects reinforce its validity, the method under study aligns perfectly with the fundamental target of shoulder arthroplasty, which is to accurately restore the patient's healthy anatomy. Although the Pix2Pix model has not been directly compared with existing methods, it is expected to have the potential to surpass them and become the gold standard for glenoid bone reconstruction.

While the study highlights the potential of CycleGAN and Pix2Pix models to reconstruct a defective glenoid, it is important to acknowledge certain limitations and identify areas for further improvement. It is crucial to acknowledge that the limited size of the dataset was a constraint in training the CycleGAN and Pix2Pix models and may have affected their overall performance in terms of their ability to generalize and produce high-quality images. The algorithm used for the classification of the dataset images and the qualitative test only takes into account one morphological parameter of the glenoid to assess its condition, which may have impacted the results obtained. Additionally, the manual annotation of the landmarks may also have influenced the accuracy of the results. These tasks should have been carried out by an expert in the field of medical imaging, which can be a time-consuming and expensive solution. Lastly, it would be paramount to overcome the hardware limitations and find a metric that mimics human intelligence to improve GANs results and compare different models.

Considering a substantially larger dataset, both models are expected to outperform in the reconstruction of a defective glenoid. This presents an opportunity to develop an automated and intelligent virtual reconstruction tool and incorporate it into a 3D preoperative planning system.

BIBLIOGRAPHY

[1] Tamra Lysaght, Hannah Yeefen Lim, Vicki Xafis, and Kee Yuan Ngiam. AI-Assisted Decision-making in Healthcare. *Asian Bioethics Review*, 11(3):299–314, September 2019. doi: 10.1007/s41649-019-00096-0.

[2] Thomas G. Myers, Prem N. Ramkumar, Benjamin F. Ricciardi, Kenneth L. Urish, Jens Kipper, and Constantinos Ketonis. Artificial Intelligence and Orthopaedics. *Journal of Bone and Joint Surgery*, 102 (9):830–840, February 2020. doi: 10.2106/jbjs.19.01128.

[3] PeekMed. Intelligence (AI) in Healthcare: how it's used, jan 2022. URL `https://blog.peekmed.com/ai-in-healthcare`.

[4] Simon J. Federer and Gareth G. Jones. Artificial intelligence in orthopaedics: A scoping review. *PLOS ONE*, 16(11):e0260471, November 2021. doi: 10.1371/journal.pone.0260471.

[5] Silvana Secinaro, Davide Calandra, Aurelio Secinaro, Vivek Muthurangu, and Paolo Biancone. The role of artificial intelligence in healthcare: a structured literature review. *BMC Medical Informatics and Decision Making*, 21(1), April 2021. doi: 10.1186/s12911-021-01488-9.

[6] PeekMed. Artificial Intelligence (AI) in Orthopedics: How to use it, jan 2022. URL `https://blog.peekmed.com/artificial-intelligence-in-orthopedics`.

[7] PeekMed. Preoperative Planning: Why is it important?, jan 2022. URL `https://blog.peekmed.com/pre-operative-planning`.

[8] Elena G. BIGNAMI, Federico COZZANI, Paolo DEL RIO, and Valentina BELLINI. Artificial intelligence and perioperative medicine. *Minerva Anestesiologica*, December 2020. doi: 10.23736/s0375-9393.20.14999-x.

[9] Henry Knipe and Jeremy Jones. Humerus. August 2013. doi: 10.53347/rid-24562.

[10] Magdalena Chmiel-Nowak and Henry Knipe. Glenoid labrum. January 2014. doi: 10.53347/rid-27109.

[11] Nicola McLaren MSc. Glenohumeral joint, September 2022. URL `https://www.kenhub.com/en/library/anatomy/the-shoulder-joint`.

[12] Chirag Udhani and Jeremy Jones. Glenohumeral joint. August 2013. doi: 10.53347/rid-24606.

[13] Thay Q. Lee. Clinical anatomy and biomechanics of the glenohumeral joint (including stabilizers). In *Shoulder Instability: A Comprehensive Approach*, pages 1–19. Elsevier, 2012. doi: 10.1016/b978-1-4377-0922-3.00010-1.

[14] Katrien Plessers, Peter Vanden Berghe, Christophe Van Dijck, Roel Wirix-Speetjens, Philippe Debeer, Ilse Jonkers, and Jos Vander Sloten. Virtual reconstruction of glenoid bone defects using a statistical shape model. *Journal of Shoulder and Elbow Surgery*, 27(1):160–166, January 2018. doi: 10.1016/j.jse.2017.07.026.

[15] R.Sean Churchill, John J. Brems, and Helmuth Kotschi. Glenoid size, inclination, and version: An anatomic study. *Journal of Shoulder and Elbow Surgery*, 10(4):327–332, July 2001. doi: 10.1067/mse.2001.115269.

[16] L. Favard, J. Berhouet, G. Walch, J. Chaoui, and C. Lévigne. Superior glenoid inclination and glenoid bone loss. 46(12):1015–1021, November 2017. doi: 10.1007/s00132-017-3496-1.

[17] Arghavan Sharifi, Matthew J. Siebert, and Avneesh Chhabra. How to Measure Glenoid Bone Stock and Version and Why It Is Important: A Practical Guide. *RadioGraphics*, 40(6):1671–1683, October 2020. doi: 10.1148/rg.2020200008.

[18] G. Porcellini, G. M. Micheloni, L. Tarallo, P. Paladini, G. Merolla, and F. Catani. Custom-made reverse shoulder arthroplasty for severe glenoid bone loss: review of the literature and our preliminary results. *Journal of Orthopaedics and Traumatology*, 22(1), January 2021. doi: 10.1186/s10195-020-00564-6.

[19] Ashish Gupta, Christoph Thussbas, Michael Koch, and Ludwig Seebauer. Management of glenoid bone defects with reverse shoulder arthroplasty—surgical technique and clinical outcomes. *Journal of Shoulder and Elbow Surgery*, 27(5):853–862, May 2018. doi: 10.1016/j.jse.2017.10.004.

[20] Samuel A. Antuna, John W. Sperling, Robert H. Cofield, and Charles M. Rowland. Glenoid revision surgery after total shoulder arthroplasty. *Journal of Shoulder and Elbow Surgery*, 10(3):217–224, May 2001. doi: 10.1067/mse.2001.113961.

[21] Claudio Chillemi and Vincenzo Franceschini. Shoulder Osteoarthritis. *Arthritis*, 2013:1–7, January 2013. doi: 10.1155/2013/370231.

[22] Kiet V. Vo, Daniel J. Hackett, Albert O. Gee, and Jason E. Hsu. Classifications in Brief: Walch Classification of Primary Glenohumeral Osteoarthritis. *Clinical Orthopaedics & Related Research*, 475(9):2335–2340, September 2017. doi: 10.1007/s11999-017-5317-6.

[23] Michael J. Bercik, Kevin Kruse, Matthew Yalizis, Marc-Olivier Gauci, Jean Chaoui, and Gilles Walch. A modification to the Walch classification of the glenoid in primary glenohumeral osteoarthritis using

three-dimensional imaging. *Journal of Shoulder and Elbow Surgery*, 25(10):1601–1606, October 2016. doi: 10.1016/j.jse.2016.03.010.

[24] Michael Thomas, Amit Bidwai, Amar Rangan, Jonathan L Rees, Peter Brownson, Duncan Tennent, Clare Connor, and Rohit Kulkarni. Glenohumeral osteoarthritis. *Shoulder and Elbow*, 8(3):203–214, April 2016. doi: 10.1177/1758573216644183.

[25] Derrick L. Burgess, Mike S. McGrath, Peter M. Bonutti, David R. Marker, Ronald E. Delanois, and Michael A. Mont. Shoulder Resurfacing. *The Journal of Bone and Joint Surgery-American Volume*, 91 (5):1228–1238, May 2009. doi: 10.2106/jbjs.h.01082.

[26] Joseph A. Abboud, Charles L. Getz, and Gerald R. Williams. Chapter 11 - Shoulder Arthroplasty. In JONATHAN P. GARINO and PEDRO K. BEREDJIKLIAN, editors, *Core Knowledge in Orthopaedics: Adult Reconstruction Arthroplasty*, pages 180–210. Mosby, Philadelphia, 2007. ISBN 978-0-323-03370-1. doi: 10.1016/B978-0-323-03370-1.50016-2.

[27] Naoko Mizuno, Patrick J Denard, Patric Raiss, and Gilles Walch. Reverse Total Shoulder Arthroplasty for Primary Glenohumeral Osteoarthritis in Patients with a Biconcave Glenoid. *The Journal of Bone and Joint Surgery-American Volume*, 95(14):1297–1304, July 2013. doi: 10.2106/jbjs.l.00820.

[28] Thomas Kozak, Stefan Bauer, Gilles Walch, Saad Al-karawi, and William Blakeney. An update on reverse total shoulder arthroplasty: current indications, new designs, same old problems. *EFORT Open Reviews*, 6(3):189–201, March 2021. doi: 10.1302/2058-5241.6.200085.

[29] L. De Wilde and A. Van Tongel. The history of shoulder arthroplasty. In *Joint Replacement Technology*, pages 571–601. Elsevier, 2014. doi: 10.1533/9780857098474.4.571.

[30] Jonelle Petscavage-Thomas. Preoperative Planning and Postoperative Imaging in Shoulder Arthroplasty. *Seminars in Musculoskeletal Radiology*, 18(04):448–462, September 2014. doi: 10.1055/s-0034-1384833.

[31] David J. Saliken, Troy D. Bornes, Martin J. Bouliane, David M. Sheps, and Lauren A. Beaupre. Imaging methods for quantifying glenoid and hill-sachs bone loss in traumatic instability of the shoulder: a scoping review. *BMC Musculoskeletal Disorders*, 16(1), July 2015. doi: 10.1186/s12891-015-0607-1.

[32] Justin Rabinowitz, Richard Friedman, and Josef K. Eichinger. Management of Glenoid Bone Loss with Anterior Shoulder Instability: Indications and Outcomes. *Current Reviews in Musculoskeletal Medicine*, 10(4):452–462, October 2017. doi: 10.1007/s12178-017-9439-y.

[33] Dominique M. Rouleau, Jacob F. Kidder, Juan Pons-Villanueva, Savvas Dynamidis, Michael Defranco, and Gilles Walch. Glenoid version: How to measure it? Validity of different methods in two-dimensional

computed tomography scans. *Journal of Shoulder and Elbow Surgery*, 19(8):1230–1237, December 2010. doi: 10.1016/j.jse.2010.01.027.

[34] Matthew J. Siebert, Majid Chalian, Arghavan Sharifi, Parham Pezeshk, Yin Xi, Parker Lawson, and Avneesh Chhabra. Correction to: Qualitative and quantitative analysis of glenoid bone stock and glenoid version: inter-reader analysis and correlation with rotator cuff tendinopathy and atrophy in patients with shoulder osteoarthritis. *Skeletal Radiology*, 49(6):995–1003, February 2020. doi: 10.1007/s00256-020-03386-z.

[35] Matthijs Jacxsens, Alexander Van Tongel, Heath B. Henninger, Robert Z. Tashjian, and Lieven De Wilde. The three-dimensional glenohumeral subluxation index in primary osteoarthritis of the shoulder. *Journal of Shoulder and Elbow Surgery*, 26(5):878–887, May 2017. doi: 10.1016/j.jse.2016.09.049.

[36] Gilles Walch, Roger Badet, Aziz Boulahia, and Alfred Khoury. Morphologic study of the Glenoid in primary glenohumeral osteoarthritis. *The Journal of Arthroplasty*, 14(6):756–760, September 1999. doi: 10.1016/s0883-5403(99)90232-2.

[37] Kenneth W. Donohue, Eric T. Ricchetti, Jason C. Ho, and Joseph P. Iannotti. The Association Between Rotator Cuff Muscle Fatty Infiltration and Glenoid Morphology in Glenohumeral Osteoarthritis. *Journal of Bone and Joint Surgery*, 100(5):381–387, March 2018. doi: 10.2106/jbjs.17.00232.

[38] Filip Verhaegen, Katrien Plessers, Olivier Verborgt, Lennart Scheys, and Philippe Debeer. Can the contralateral scapula be used as a reliable template to reconstruct the eroded scapula during shoulder arthroplasty? *Journal of Shoulder and Elbow Surgery*, 27(6):1133–1138, June . doi: 10.1016/j.jse.2017.12.024.

[39] Katrien Plessers, Filip Verhaegen, Christophe Van Dijck, Roel Wirix-Speetjens, Philippe Debeer, Ilse Jonkers, and Jos Vander Sloten. Automated quantification of glenoid bone defects using 3-dimensional measurements. *Journal of Shoulder and Elbow Surgery*, 29(5):1050–1058, May . doi: 10.1016/j.jse.2019.10.007.

[40] Asvin Ganapathi, Jesse A. McCarron, Xi Chen, and Joseph P. Iannotti. Predicting normal glenoid version from the pathologic scapula: a comparison of 4 methods in 2- and 3-dimensional models. *Journal of Shoulder and Elbow Surgery*, 20(2):234–244, March . doi: 10.1016/j.jse.2010.05.024.

[41] Peter Vanden Berghe, Jan Demol, Frederik Gelaude, and Jos Vander Sloten. Virtual anatomical reconstruction of large acetabular bone defects using a statistical shape model. *Computer Methods in Biomechanics and Biomedical Engineering*, 20(6):577–586, December 2016. doi: 10.1080/10255842.2016.1265110.

[42] Yu-Tao Xiong, Wei Zeng, Lei Xu, Ji-Xiang Guo, Chang Liu, Jun-Tian Chen, Xin-Ya Du, and Wei Tang. Virtual reconstruction of midfacial bone defect based on generative adversarial network. *Head & Face Medicine*, 18(1), June 2022. doi: 10.1186/s13005-022-00325-2.

[43] Sergey I. Nikolenko. Chapter 4 - Generative Models in Deep Learning. In *Synthetic Data for Deep Learning*, pages 107–147. Springer International Publishing, 2021. doi: 10.1007/978-3-030-75178-4.

[44] Jelmer M. Wolterink, Konstantinos Kamnitsas, Christian Ledig, and Ivana Išgum. Deep learning: Generative adversarial networks and adversarial methods. In *Handbook of Medical Image Computing and Computer Assisted Intervention*, pages 547–574. Elsevier, 2020. doi: 10.1016/b978-0-12-816176-0.00028-4.

[45] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Networks. 2014. doi: 10.48550/ARXIV.1406.2661.

[46] Lars Ruthotto and Eldad Haber. An Introduction to Deep Generative Modeling. 2021. doi: 10.48550/ARXIV.2103.05180.

[47] André Ferreira, Jianning Li, Kelsey L. Pomykala, Jens Kleesiek, Victor Alves, and Jan Egger. GAN-based generation of realistic 3D data: A systematic review and taxonomy. 2022. doi: 10.48550/ARXIV.2207.01390.

[48] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. 2017. doi: 10.48550/ARXIV.1703.10593.

[49] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets, 2014.

[50] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976, 2017. doi: 10.1109/CVPR.2017.632.

[51] Cyclegan. URL `https://www.tensorflow.org/tutorials/generative/cyclegan`.

[52] David Abramian and Anders Eklund. Generating fMRI volumes from T1-weighted volumes using 3D Cyclegan. 2019. doi: 10.48550/ARXIV.1907.08533.

[53] Zizhao Zhang, Lin Yang, and Yefeng Zheng. Multimodal medical volumes translation and segmentation with generative adversarial network. In *Handbook of Medical Image Computing and Computer Assisted Intervention*, pages 183–204. Elsevier, 2020. doi: 10.1016/b978-0-12-816176-0.00013-2.

[54] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein GAN. 2017. doi: 10.48550/ARXIV.1701.07875.

[55] Lars Mescheder, Andreas Geiger, and Sebastian Nowozin. Which Training Methods for GANs do actually Converge? 2018. doi: 10.48550/ARXIV.1801.04406.

[56] Lihong Huang, Zixia Zhou, Yi Guo, and Yuanyuan Wang. A stability-enhanced CycleGAN for effective domain transformation of unpaired ultrasound images. *Biomedical Signal Processing and Control*, 77: 103831, August 2022. doi: 10.1016/j.bspc.2022.103831.

[57] Compute peak signal-to-noise ratio (psnr) between images - simulink. URL `https://www.mathworks.com/help/vision/ref/psnr.html`.

[58] Structural similarity (ssim) index for measuring image quality - simulink. URL `https://www.mathworks.com/help/images/ref/ssim.html`.

[59] Daniel Bell and Candace Moore. Dice similarity coefficient. March 2020. doi: 10.53347/rid-75056.

Este trabalho foi efetuado sobre supervisão do Professor Doutor Carlos Manuel Gregório Santos Lima e Mestre Pedro Ribeiro (Universidade do Minho; PeekMed).