UNIVERSITÀ
DEGLI STUDI
FIRENZE

Università degli Studi di Firenze
Dipartimento di Ingegneria dell'Informazione (DINFO)
Corso di Dottorato in Ingegneria dell'Informazione
Curriculum: Automatica, Ottimizzazione e Sistemi Complessi (AOSC)

Study of a bone segmentation and 3D modeling workflow for diagnosis and planning

Candidate
Eleonora Tiribilli

Supervisors
Prof. Leonardo Bocchi

Prof. Ernesto Iadanza

Dr. Leonardo Manetti

PhD Coordinator
Prof. Fabio Schoen

ciclo XXXVII, 2021-2024

Università degli Studi di Firenze, Dipartimento di Ingegneria dell'Informazione (DINFO).

To my wonderful family,
to my love Dino,
and to everyone who supported me
in this journey.

Acknowledgments

I would like to acknowledge the efforts and guidance of my supervisor, Prof. Leonardo Bocchi, along with Prof. Ernesto Iadanza, Dr. Leonardo Manetti, and all my colleagues in the R&D department, who provided assistance during my research. My particular thanks go to Dr. Leonardo Manetti, my industrial supervisor, for the incredible opportunity to embark on this journey. My thanks also extend to the company Imaginalis Srl, whose resources, flexibility, and support have been crucial in enabling me to complete my research while balancing professional responsibilities. I am deeply grateful for the opportunities and understanding provided by my colleagues and the management team. Special thanks go to Dr. Elena Denisova, who collaborated closely on the key aspects of my work.

I am deeply grateful to my colleagues and collaborators, of the Bio Lab, for their contributions and for creating an inspiring and stimulating environment, especially Sara Lombardi, who has become a great friend.

To my friends and family, who have been my source of strength and encouragement: your support has been essential to completing this journey. I am deeply grateful for your patience, love, and understanding.

Finally, to my boyfriend, Dino, whose love, patience, and encouragement have been my anchor during this journey: thank you for being my constant source of support.

Contents

List of Figures

List of Tables

AR Augmented Reality

CBCT Cone Beam Computed Tomography

CNN Convolutional Neural Network

CT Computed Tomography

CTA Cognitive Task Analysis

DC Dice Coefficient

DICOM Digital Imaging and Communications in Medicine

GPU Graphics Processing Unit

GUI Graphical User Interface

HCI Human-Computer Interaction

HD Hausdorff Distance

HDMI High-Definition Multimedia Interface

HE Heuristic Evaluation

HFE Human Factors Engineering

HU Hounsfield Units

ICP Iterative Closest Point

JI Jaccard Index

MD Medical Device

MPR Multi-Planar Reformation

MPT Multi-Planar Training

MRI Magnetic Resonance Imaging

MV Majority Voting

SaMD Software as Medical Device

SBM Single Bone Modeler

STL Standard Tessellation Language

VSP Virtual Surgical Planning

Chapter 1

Introduction

Rapid technological development over the past decades has fundamentally influenced many aspects of our lives, including medicine. However, for some parts, the translation of recent technology into everyday clinical practice seems to lag behind. The example of bone segmentation and 3D modeling workflows for diagnosis and planning supports this statement. The main reasons that have been identified in the literature are that these systems lack ease of use and are difficult to integrate into clinical practice. This includes cumbersome segmentation procedures as well as complicated software solutions that require expert users.

This study aims to investigate an easy-to-use bone segmentation and 3D modeling workflows for diagnosis and planning to be integrated in a user-centered Biomedical Imaging Platform All-in-One

1.1 The objective

In the last years, image processing algorithms have played an important role in enhancing information from radiological imaging technologies for healthcare professionals, by enabling deeper and more detailed examination of medical data. Segmentation of bone structures within orthopedics becomes an important task for diagnostics, preoperative planning, and even personalized treatments.

Segmented bone models derived from medical images, particularly Computed Tomography (CT) scans, are important in visualizing complex bone structures. These models help surgeons plan interventions more accurately, im-

proving patient outcomes and minimizing risks. For example, in orthopedic and reconstructive surgeries 3D-printed bone models based on CT data guide surgical procedures; moreover, they aid in understanding bone strength, osteoporosis, and the creation of personalized implants or prosthetics. Although geometric and dimensional accuracy is crucial for evaluating medical image segmentation and 3D modeling tools, clinical environments also demand high usability, as factors like automation, segmentation time, realistic 3D visualization and training time impact the efficiency of clinicians and the acceptance of the technology.

This thesis is the result of an industrial research project conducted within the R&D department of the company Imaginalis srl [1], a manufacturer of cone beam technology tomography machines (in which the author works as an employee). Therefore, the focus of this research is the segmentation of Cone Beam Computed Tomography (CBCT) images and the creation of 3D models from this kind of data. Unlike spiral CT, CBCT uses a cone-shaped X-ray beam to capture isotropic voxels with higher resolution on hard tissues, making it useful in orthopedic and maxillofacial imaging. Despite its resolution strengths, CBCT introduces several artifacts such as beam hardening, shading, and cupping, complicating the process of segmentation. These artifacts introduce noise in the grayscale values, which results in difficulty with accurate bone segmentation, especially in areas of complex anatomy: joints, hands, feet, and the maxillofacial region. In these cases, poor contrast is found between the bones and soft tissues. Besides that, images are not calibrated in Hounsfield Units (HU), further complicating the creation of models.

The main goal of this industrial research project is to build an image segmentation and three-dimensional modeling workflow for CBCT as part of an user-centered software, the Imaging Platform All-in-One. This multimodal platform processes images from various modalities (CT, fluoroscopy, digital radiography) and is designed for both intraoperative manipulations and pre- and post- operative planning. The platform will include functionalities for image enhancement, visualization, quantitation, segmentation, and registration, all essential features for medical imaging analysis.

The success of the proposed segmentation workflow will depend largely on the analysis of user satisfaction in terms of the interface and functionality of the Biomedical Imaging Platform All-in-One. A good balance between

[1] Imaginalis srl, Sesto Fiorentino, Italy, www.imaginalis.it

research in the field of image segmentation and software usability will enhance the impact of this work. The more user-friendly the tool, the more acceptable it will be to the research and medical world, which will increase its applications.

During my PhD, I have focused on these two key aspects: image segmentation/modeling and medical software usability, ensuring that the final solution addresses both technical and user-centric requirements.

1.2 Contributions

Throughout my PhD, I made contributions to both the technical and usability aspects of medical image visualization and processing. In collaboration with Dr. E. Denisova, we developed and conducted heuristic evaluations and usability tests for the Biomedical Imaging Platform All-in-One. This resulted in the creation of a formal usability testing protocol, adhering to international standards, for evaluating the platform's effectiveness, efficiency, and user satisfaction. The protocol, incorporating 55 exploratory tasks and 4 real-life scenarios, ensures consistency and repeatability in assessing the platform's functionality. Our results showed very high user satisfaction and validated the stability of the Graphical User Interface (GUI)of the Biomedical Imaging Platform All-in-One [31, 32, 89, 131].

In the context of bone segmentation, I initially explored traditional methods such as pixel-based, region-based, and edge-based techniques to evaluate their performance on CBCT data [134]. Recognizing the limitations of these approaches, I designed a graph- cut-based approach with thresholded initialization that provided better segmentation results than those from classical user-driven methods. Despite its sensitivity to CBCT's non-HU calibrated pixel intensities, this approach demonstrated improved user-friendliness and accuracy [132].

To address the challenges posed by fluctuating pixel intensities and complex anatomical structures, my research shifted towards deep learning-based segmentation, which resulted in the development of the Single Bone Modeler (SBM). This tool allows for efficient bone segmentation using a U-Net-based convolutional neural network, trained on multi-planar CBCT data (axial, sagittal, and coronal). The SBM is designed to segment and isolate a single bone, generating a 3D model with just one click. This approach significantly improves segmentation accuracy, particularly in anatomically

complex regions like joints and extremities, surpassing traditional methods, as demonstrated by superior JI and DC metrics.

Additionally, the 3D modeling process is powered by the Marching Cubes algorithm, enabling the creation of precise models that are validated using the Hausdorff distance. By automating much of the segmentation and modeling process, the SBM minimizes manual intervention and enhances usability [130, 133].

Lastly, a novel workflow for preoperative planning in the reduction of complex bone fractures using CBCT is presented. The proposed approach introduces a semi-automatic virtual fracture reduction system, which integrates the segmentation and modeling techniques detailed in the previous chapter with registration methods to align bone fragment models to a reference template. The method is based on the Iterative Closest Point (ICP) algorithm, enhanced with a new initialization strategy to improve accuracy. Experimental tests conducted on bovine femur fractures demonstrate the system's accuracy, efficiency, and usability, providing a comprehensive evaluation of the results.

Overall, my PhD research advanced the accuracy, efficiency, and usability of bone segmentation and 3D modeling workflow in CBCT imaging. The SBM workflow offers a robust, user-friendly solution for improving preoperative planning, and enhancing the practicality of CBCT in clinical and academic settings.

1.3 Thesis outline

The following chapters cover the contributions of this thesis. Firstly, in terms of the usability of the imaging platform then, in the field of image segmentation and modeling. Chapter 2 introduces the multimodal biomedical imaging platform all-in-one, designed to streamline data management, visualization, and processing across the surgical workflow, from preoperative planning to postoperative follow-up. The platform aims to integrate advanced technologies such as 3D segmentation and modeling within a user-centered environment. Special attention is given to intuitive design and optimized user experience, as such conditions enhance adoption by health professionals. This chapter sets the ground for the usability evaluations discussed in chapters 3 and 4, developed in collaboration with Dr. E. Denisova. Chapter 3 presents the application of heuristic evaluation on the imaging

platform during its development phase. The analysis reveals potential usability problems, guiding design improvements in subsequent iterations. This chapter emphasizes the iterative nature of usability testing as part of platform development. Focusing on the formalization of usability testing for Software as Medical Device (SaMD), chapter 4 addresses a gap in the literature regarding standardized protocols for evaluating medical software. The chapter introduces a reproducible framework that was applied to the imaging platform, demonstrating the software's effectiveness, efficiency, and user satisfaction. The results confirm the protocol's robustness and enable comparative analysis across different SaMDs. The usability findings from this chapter contribute to refining the imaging platform's interface and functionality. Transitioning from usability to image processing, chapter 5 explores bone segmentation and 3D modeling as essential tools for diagnostic and planning purposes within the imaging platform. It provides a survey of segmentation techniques, and their medical applications and presents the challenges for working with CBCT.

Chapter 6 presents a novel approach for segmenting bones in extremities from CBCT images, addressing the difficulty of segmenting small, closely positioned bones. The method involves an automatic graph cut initialization based on pixel-based techniques, followed by morphological post-processing. The results are compared to traditional user-dependent methods, demonstrating the advantages in accuracy and usability. Due to some limitations of the previous strategy the following chapters refer to the use of Convolutional Neural Network (CNN).

The focus of chapter 7 is on finding the best training strategy for CNNs for bone segmentation in CBCT. It compares axial training, majority voting, and Multi-Planar Training (MPT), identifying MPT as a resource-efficient strategy that maintains accuracy while reducing computational demands. Chapter 8 introduces the SBM, a deep learning-based workflow for precise bone segmentation and 3D modeling in CBCT, with a focus on complex extremities. The SBM workflow includes segmentation, separation, and 3D modeling stages, employing a specialized U-Net architecture and comparisons with other methods such as SegNet, Graph Cut, and Thresholding. Chapter 9 presents a semi-automatic workflow for virtual fracture reduction, incorporating the segmentation and modeling techniques established in the previous chapter. The system aligns 3D bone fragments with reference models, facilitating preoperative planning. The workflow's accuracy and us-

ability are validated through experimental tests on bovine femur fractures. The final chapter concludes this thesis. It summarizes the presented contributions and puts their findings into a broader context of possible future directions.

Chapter 2

Multimodal Biomedical Imaging Platform All-in-One

This chapter aims to introduce the Multimodal Biomedical Imaging Platform All-in-One designed to simplify intraoperative data management, visualization, and processing, preoperative planning, and postoperative follow-up. This platform is conceived as a multimodal, all-in-one environment in which advanced technologies such as 3D segmentation and modeling will converge. When integrating biomedical technologies into clinical practice, providing operators with well-designed and easy-to-use tools is essential for the success and effectiveness of technological advancements. An intuitive interface and optimized user experience not only facilitate the adoption of new technologies by healthcare professionals but also enhance diagnostic accuracy and operational efficiency, contributing to higher-quality patient care. The work and ideas described below were conceived and developed in collaboration with Dr. E. Denisova, data collection and analysis were conducted as a team effort. [1]

[1]This chapter has been published as " Multimodal Biomedical Imaging Platform All-in-One" [32] in *IFMBE Proceedings of IUPESM World Congress on Medical Physics and Biomedical Engineering 2022.*

2.1 Introduction

Recently, rapid technological progress has offered a lot of opportunities for the development of novel tools and methods in the biomedical field, first of all, thanks to the hardware performance [105]. The usage of powerful workstations together with high-performing Graphics Processing Unit (GPU) permits advancing of the diagnostic by imaging techniques [143]. As imaging plays a central role in the healthcare process [114] and is particularly important in domains of visual diagnosis such as radiology, an additional effort should be put in the design and development of applications that help the physicians in interpreting the biomedical imaging data and therefore improving their everyday medical practice. However, it was noted that the healthcare system is slow in adopting information technology into the work environment [69]. Moreover, according to a survey including 336 radiologists, strong majority of radiologists were dissatisfied with their current computing workflow and setup ([119], [61]). The causes of this dissatisfaction are multiple and include the complexity of the proposed solutions, "messy and chaotic" [33] GUI, the lack of understanding of the workflow and users' mental model. Many issues are due to the fact that there is no single way to represent the biomedical knowledge and often there is no theoretical foundation while designing and developing the solutions [106].
The task that we set to resolve is to introduce a system - a multimodal all-in-one imaging platform - designed to be more intuitive and confluent with everyday orthopedics' practice (the intra-op data managing, visualization, processing, pre-op planning, post-op follow-up, etc.). To achieve the objective we will:

- evaluate different hardware solutions (like multiscreens and touch screens);
- in-depth study the Human-Computer Interaction (HCI), Cognitive Task Analysis (CTA), usability and workflow;
- ask to contribute the group of practical orthopedics in two phases: collection of the requirements during the oral interviews and execution of the specific tasks using the proposed imaging platform.

The system should permit to process the images of different nature/-modalities: computed tomography, fluoroscopy, digital radiography, magnetic resonance imaging, nuclear medicine imaging (*multimodal*). Also, it

should be designed to be used during intra-operation manipulations as for pre-operation planning and post-operation follow-up, providing the capabilities for image enhancement, visualization, quantitation, registration, segmentation and modeling (*all-in-one*). We will describe the methods we use in our approach and propose the strategy to follow during the design of such kinds of systems, discussing, in conclusion, the next steps of our research.

2.2 Methods

Assuming that "any system will be easier and less burdensome to use if is co-extensive with the user's mental models" [105], we started our research by dipping into the theoretical foundations of HCI. HCI is a multifaceted discipline devoted to the study and practice of design and usability [106].
According to Nielsen, usability is a quality attribute that assesses how easy user interfaces are to use. He suggests that usability can be defined by five quality components [99]:

1. learnability: the system should be relatively easy to learn;
2. efficiency: once learned, the system should be productive;
3. memorability: users can easily reestablish proficiency after some period the system was not in use;
4. errors: the system should be designed to minimize errors and support error detection and recovery;
5. satisfaction: the user experience should be subjectively satisfying.

Based on this definition, Patel et al. [106] proposed the classification of the usability evaluation methods dividing them into two categories: analytic evaluation approaches and usability testing. They suggest to apply task-analytic, inspection-based and/or model-based approaches to the stage of analytic evaluation. During this stage the usability experts, domain experts or/and software designers are used as participants in laboratory-based settings. Then, a general category of methods (field observation, interviews, focus groups, surveys, etc.) is applied on the usability testing stage. It captures situated and contextual use of technology in field-based, laboratory-based or online studies (see Figure2.1).
In this paper we cover the analytic approach we used for design and usability evaluation.

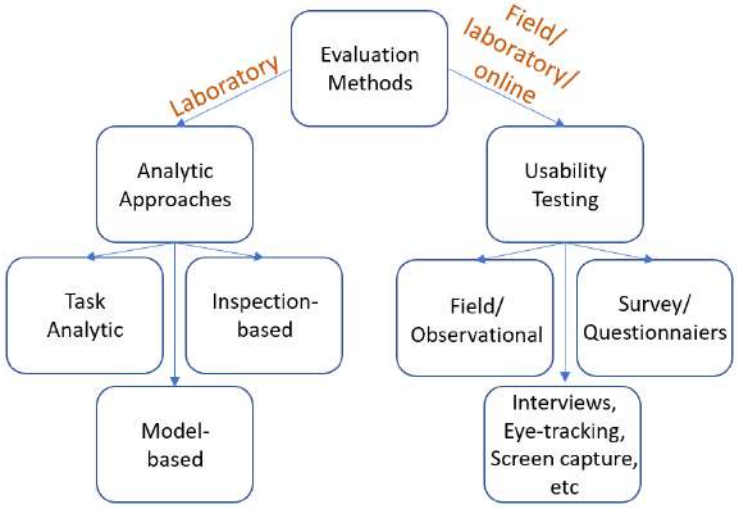


Figure 2.1: Classification of Evaluation Methods [32]

2.2.1 Cognitive Task Analysis

The first step of the research consisted of requirements collection, based on the analytic CTA. CTA is a family of psychological research methods for uncovering and representing people's state of mind, used in a variety of different fields, including the design of human-computer interfaces [124]. It uses a number of interview and observation strategies to capture a description of the knowledge that experts use to perform complex tasks [29].
A group of 10 practical physicians including radiologists and orthopedics was interviewed about their current experience with the third-party imaging systems. Based on their response the list of the desired environment features was collected, e.g.:

- multiscreen and touch screen usage;
- multimodality support;
- comparison and automated registration
- automated segmentation and three dimensional modeling;
- mini-invasive intervention planning;

- medical tools tracking;
- 3D easy cropping and transfer functions customization, etc.

2.2.2 Graphical User Interface. Gestalt Principle, Fitts and Hick-Hyman Laws

Once requirements were collected, we prepared the proposal of the GUI. The following aspects were considered (see Figure2.2):

- distribution of different amounts of visual information between different amounts of free space (in the case when two screens are available, the visual information was split between them, otherwise all information was presented on the single screen);
- scalability (support of different monitor configurations from the smallest dimensions equal to 1280x720, up to the highest dimensions of 4096x2160 [4K], plus mixed solutions, e.g. 4K main monitor with 200% of elements re-dimension and a 1920x1080 secondary monitor without re-dimension);
- mouse/keyboard versus touchscreens and the switch between control mechanisms (e.g. use of the graphical controls of bigger sizes, placement of controls in the way the hand does not cover them and runs the shortest path to activate basic functionality, support for left-handed users; more, the text tips, custom cursor appearance, highlightening when elements are under mouse cursor were avoided due to the specificity of touchscreen which is controlled by fingers and not by mouse);
- concurrent usage of the same instance of imaging platform by users with different speaking languages (e.g. during intervention effectuated by international teams; achieved by avoiding the text on the most of controls and taking advantage of the "speaking" icons instead);
- pleasant visual aspect and easy to understand interface (based on the principals of human perception).

First, the Gestalt visual design principles were applied during the GUI desing [62]. The main aspects of the application of Gestalt theory to the design of interface are: the most effective placement of elements; the direction of users' attention to specific points of focus; the realisation of beautiful, pleasing, and

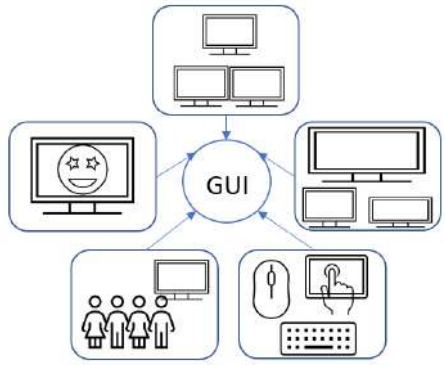


Figure 2.2: Graphical User Interface Design Aspects [32]

intuitive to use system [112].
Second, Fitts and Hick-Human Laws was taking into account. Fitts Law is used to predict human motor behavior; it is used to predict the time taken to acquire a target ([106], [39]). Instead, Hick-Hyman Law is used to predicts the human choice reaction time ([106], [51], [53]).

2.2.3 Zhang Heuristic

After completing the first integration of the GUI design, the inspection-based heuristics evaluation approach was applied. Heuristic Evaluation (HE) involves experts appraising user interfaces, and identifying potential usability and interaction issues [106]. Originally, the set of heuristics was proposed by Nielsen [99]. Then Zhang et al. [156] modified the original set of heuristics to fit better the medical devices and evaluate the patient safety of those devices. The list of the Zhang heuristics description, together with a severity rating scale and the tasks list, was distributed between the usability experts and software design experts. After the installation of the imaging platform, the experts executed the tasks. Then they gave back the list of usability problems found, describing the potential consequences and giving a severity rate. We analyzed the result of the evaluation and added the number of the modifications/fixes to the current interface design. The list of the tasks was updated according to the interface modifications and, again, was distributed and evaluated (see Figure2.3).

Figure 2.3: Design Iteration Process [32]

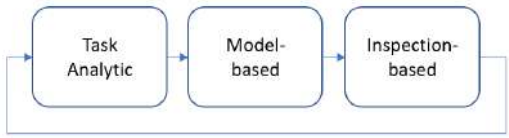


Figure 2.4: Analytic Evaluation Iterations [32]

2.2.4 Proposed Strategy

The iterative application of the analytic approach gave us a set of methods for evaluating the design and interface of our all-in-one biomedical imaging platform. We concluded that the iterative process applied to the analytic approach steps in practice produced effective results: in our research, we repeated the iteration process described in Sect. 2 until only *minor issues* remained (by *minor issues* we mean the issues that do not require the fix unless extra time is available).
Putting together the Figure2.1 and Figure2.3, we can propose to order the analytic approaches presented by Patel et al. [106] and iterate its application until the desired usability evaluation level is achieved (see Figure2.4).

2.3 Conclusion

As the proposed method is applied during the design and implementation and not on the fully developed system, the introduction of changes and fixes takes minimum time and permits to design more flexible system. The analytic approach should be iteratively repeated during all the next steps of the design and development of the Biomedical Imaging Platform All-in-One and the usability testing should be introduced to produce the measurement for the system's usability and utility. We underline that the analysis we effectuated and the preliminary results we achieved would not be possible

without the participation of the physicians, orthopedics and radiologists, because as we mentioned in the introduction, to design more intuitive and confluent with everyday physicians' practice one needs to understand the workflow and users' mental model.

Chapter 3

Heuristic evaluation of the Biomedical Imaging Platform All-in-One

The objective of this chapter is to present the application of heuristic evaluation to the Biomedical Imaging Platform All-in-One. during its development, specifically focusing on the first iteration of the iterative analysis introduced in the previous chapter. Heuristic evaluation was conducted using Zhang's well-established heuristics, together with a task list, which helped identify design characteristics that violated one or more principles. For each violation, the evaluators analyzed the usability issues likely to arise and assessed their potential impact. During this initial iteration, 38 heuristic violations were detected, with the most frequently violated heuristics being Match between System and Real World and Consistency and Standards. [1]

[1]This chapter has been published as "Heuristic evaluation of a medical imaging platform" in *IFMBE Proceedings of IUPESM World Congress on Medical Physics and Biomedical Engineering 2022,* [131].

3.1 Introduction

The implementation and the broad use of health information technologies have rapidly grown over the course of the last decade. New software products, platforms and informative systems are transforming healthcare in different ways, from the purely diagnostic and care aspect to the implementations of new services for supporting patients and improving the quality of offered assistance [86]. The usability of such technologies is a challenging aspect, strictly related to safety. When a technology is poorly designed, it can lead people to make mistakes while interacting with it. Moreover, if the medical software is installed on hardware medical devices, additional criticalities may arise, such as interference [55] and compatibility issues. In healthcare, this can be especially serious, leading to consequences for the patient.

Human interaction and usability evaluation, together with proper risk management [15], must be taken into consideration in the designing and developing stages of new systems and products in the medical field. Moreover, the international standard IEC 62366-1:2015 (Application to usability engineering to medical device) establishes a process for manufacturers to evaluate the usability of medical devices [2]. The study and the evaluation of medical software must be performed during its design and development stages to address the safety issues related to the identified usability problem, so that edits can be introduced in time. Medical device manufacturers are becoming increasingly aware of this aspect, with investments in this field, even in the early stage of development. Besides, a proper usability evaluation can identify severe bugs in the early developing stages, optimizing the operating time and lowering the possibility of future related adverse events [54]. In this study the usability evaluation of the Multimodal Biomedical Imaging Platform All-in-One is presented. The study has been conducted during its design and development. The software is a complex Biomedical Imaging Platform All-in-One conceived to be used in diagnosis and planning, allowing visualization, exploration, measurement, comparison of multiple studies and planning of surgical procedures in compliance to the Digital Imaging and Communications in Medicine (DICOM) standard.

There are many methods to evaluate system usability [38] such as heuristic evaluation [101], task analysis [8], usability test and cognitive walktrough [111]. Currently, no single method identifies all (or even most) potential problems [147]. That is the reason why during the development of a new medical software different steps and methods to evaluate usability could

be required. In this work HE is applied to Biomedical Imaging Platform All-in-One as part of the iterative analysis mentioned in 2.2.4.

HE refers to a method based on having expert evaluators who analytically examine the usability-related aspects of a user interface [100]. Because the evaluator's abilities are primarily reliant on HE, it is more subjective than typical user testing evaluation [93]. The fundamental benefit of HE is its low cost; it does not require expensive laboratory equipment to capture user interactions, expensive field studies, or difficult-to-process results from large-scale interviews [64, 100]. In a brief amount of time, an evaluator can find a wide range of usability issues and possible defects in a complicated system [93]. As a result of these factors, HE has been increasingly popular in recent years [26, 123]. HE has been used to analyze a wide range of medical technologies in the healthcare field, including infusion pumps [81], telemedicine system [127] and user manual [6, 25].HE has also been used as a support to evaluate and compare medical devices in hospital procurement. The aim of this work is to design and apply heuristic evaluation to analyze the Human-Computer-Interface of the Multimodal Biomedical Imaging Platform All-in-One.

3.2 Materials and Methods

Heuristic evaluation is a so-called analytical approach, i.e. it utilises a small set of experts to evaluate a user interface based on a set of heuristic principles regarding the interface design [68]. This technique was developed by Jakob Nielsen [101], and it has been used extensively in the assessment of user interfaces. HE is carried out by usability experts who examine an interface for potential violations of a set of usability standards (referred to as "heuristics"). These perceived violations could involve a variety of interface elements such as windows, messages, icons, navigation and interaction. Typically, a relevant subset of heuristics is selected. The selection of heuristics is based on the type of system and interface to be evaluated. Zhang [156] tailored these design principles into 14 Usability Heuristics to facilitate the heuristic analyses of medical devices. Following the evaluation of the heuristics, the potential violations are assigned a severity score. In general, HE is more subjective than traditional user testing evaluation since it is heavily dependent upon the evaluators [93]. In this work, a reference sheet has been developed and provided to expert evaluators, to normalize and better ana-

Table 3.1: Severity scale adopted in this work [131]

	Description	Mitigation
0	Not a usability problem	No fix required
1	Cosmetic problem only	Need no to be fixed unless extra time is available
2	Minor Usability problem.	Fixing this should be given low priority.
3	Major usability problem	Fixing this is an important and high priority.
4	Usability catastrophe	Fixing this is imperative

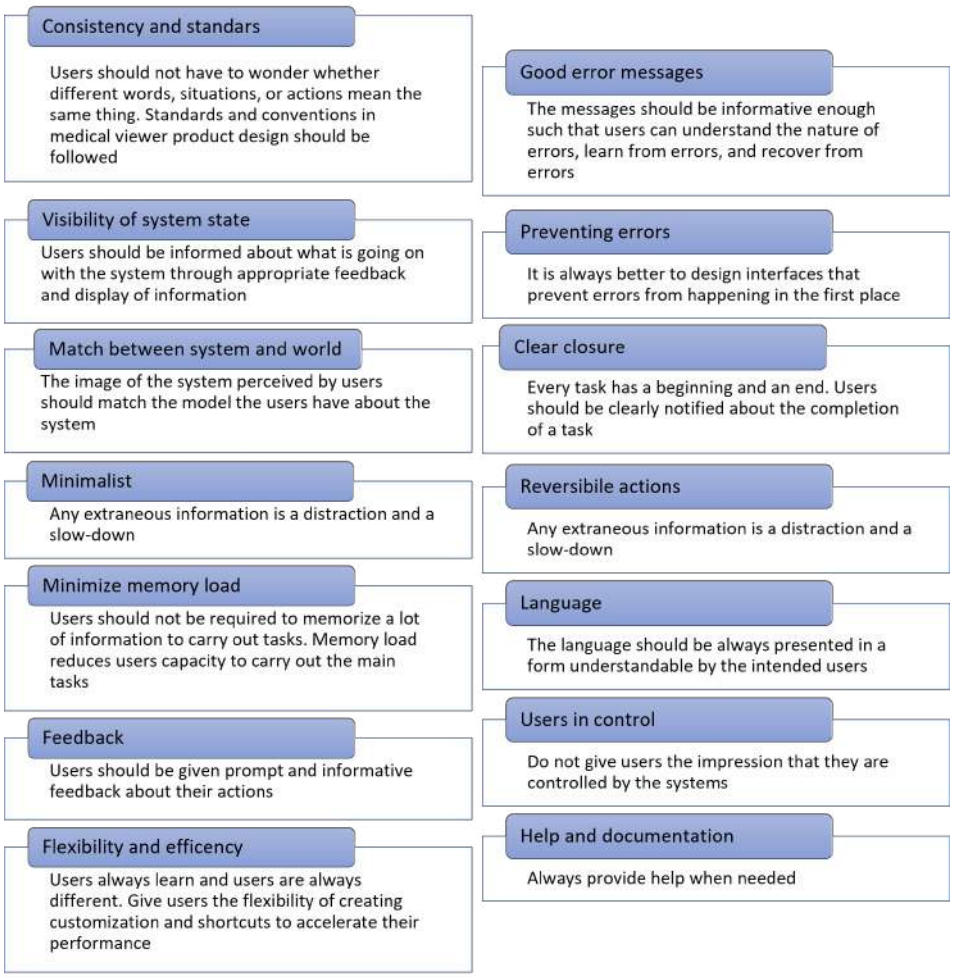
lyze evaluation results and to mitigate subject dependencies of this type of usability assessment. The reference sheet consists of four parts:

- heuristics list and description,
- severity rating scale,
- task list,
- a table to collect results.

Zhang's heuristics [156] have been adopted for the heuristic evaluation of the image platform. These consist of a set of 14 heuristics: consistency and standards, visibility of system state, match between system and world, minimalism, minimization of memory load, informative feedback, flexibility and efficiency, good error messages, preventing errors, clear closure, reversible actions, use of understandable language, users in control and help and documentation (Figure 3.1). A detailed description of each heuristic has been added to help usability evaluators to directly apply the heuristic on the medical imaging platform.

The aim is to identify design issues that have the potential to result in safety and usability problems. For each identified safety and usability problem, a severity rating must be assigned to help identify the high-priority issues. Table 3.1 shows the severity rating scale that has been developed in this work. Indeed, each heuristic analysis must include the development of a rating scale that most appropriately categorizes the types of risks encountered for the evaluated technology. Afterward, the reference sheet provides a list of tasks that guides each evaluator to explore all the software parts and functionalities. Tasks are intentionally very general to avoid guiding or

Figure 3.1: Heuristic [131]

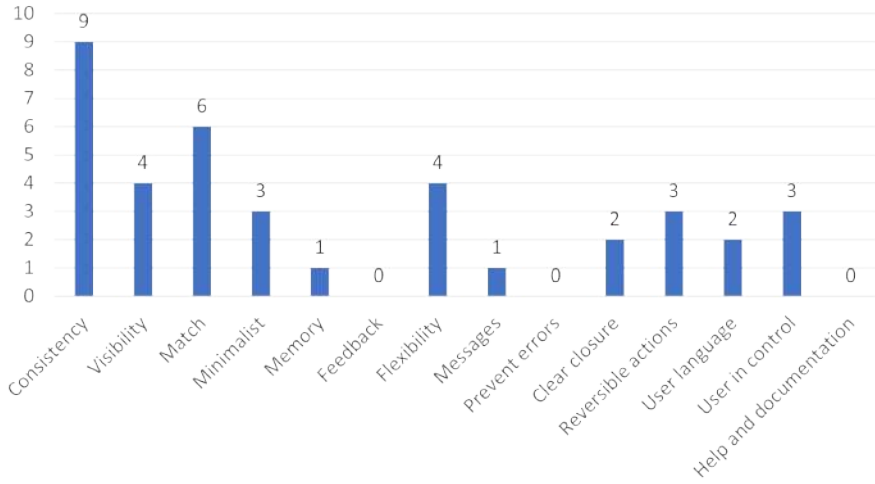


helping user through the specific functionalities. They are divided into five sections that represent the main capabilities of the software: DICOM visualization, measures, comparison, planning and multi-slices. For each section, more specific tasks are suggested as for example *"flip 2D image horizontally and vertically"* in DICOM visualization section. In the last section of the reference sheet, a template to record findings is given, to ensure each evaluator provides adequate detail as part of their analysis. Space for assessors is included in the form to provide four key information for each usability problem:

- Where the violation is
- Description of the problem and of the potential consequences it could cause
- The heuristic that has been violated
- The severity rate associated to the issues.

The above-cited template has been given to the usability experts in advance so that they could have time to get familiar with them before the session. Three usability specialists conducted the heuristic evaluation, having each evaluator analyze the interface on his own. This approach is mandatory to ensure that each evaluator provides impartial and unbiased feedback. The user has been suggested to go through at least two rotations of the task list to become familiar with the software. Firstly collecting any initial impressions, then identifying violations and usability problems. The results of the evaluation have been recorded as written reports by filling up the given template. Written reports have the advantage of presenting a formal record of the evaluation, but on the other hand, they require to be read and aggregated by an evaluation manager. HE has been extended with a debriefing session to provide some design advice. The evaluators, together with the evaluation manager and the representatives of the design team, joined the debriefing. The session was focused on discussing possible redesigns to address the major usability problems and whatever could emerge as a general problematic aspect. The debriefing has also been a good opportunity for discussing the positive sides of the design since HE does not address this important issue.

Figure 3.2: Frequencies of heuristics violations

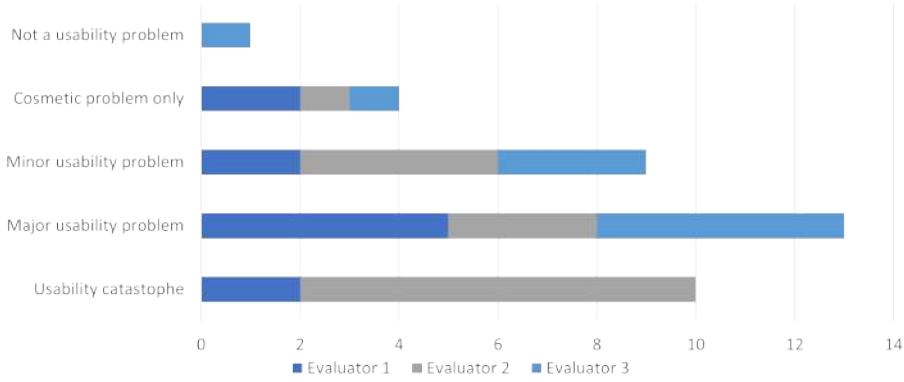


3.3 Results

The results of experts' analyses are presented in table 3.2 and figure 3.2. From 14 heuristics, 11 were violated and the experts pointed out 38 violations in total. According to violation frequencies, the most violated heuristics is consistency, with nine violations. Lack of consistency was identified in button icons and text, as well as medical tools naming. One example of button icon inconsistency has been noted in the rotation and restoration icons, which are not easily distinguishable. The labeling of medical tools for adjusting the level and window width of histograms was incorrect. Moreover, while the software refers to some of its functionalities as brightness and contrast, assessors pointed out that WW (Window Width) and WL (Window Level) should be used. Regarding the match between the system and the world, six heuristics violations were found. The most serious regards image flip and rotation. In the planning environment, instrument dimensions are not correctly scaled based on CT volume dimension. This has been reported as a severe violation of the match between the system and the real world. Clear closure has been violated two times, the most critical violation is the absence of messages and pop-ups at program shutdown.

The severity rate distribution of these issues is reported in figure 3.3. 13 issues have been scored has *major usability problem* and 10 as *usability*

Figure 3.3: Issues severity rating [131]



catastrophe.

3.4 Conclusion and Future Work

The aim of this study is to evaluate through heuristic evaluation a first prototype of a medical imaging software, designed for visualization, measurement, planning and studies comparison, in compliance to the DICOM standard. The Biomedical Imaging Platform All-in-One owns all the main characteristics of the most popular commercial software products on the market, but it has been developed with particular attention to HCI. Medical imaging platforms usually have a lot of functionalities and very specific and technical tools which can lead to poor Graphical Interface and cumbersome user experience. Furthermore, as medical software, this can be dangerous for patient safety. Investing in usability and user experience design may advance safety, productivity and user acceptance in the field. A reference sheet for conducting a heuristic evaluation of the Biomedical Imaging Platform All-in-One has been developed. The evaluation follows Zhang's heuristics and it has been conducted by three evaluators. The most violated heuristics are *Consistency and Standard* and *Match between System and Real World*. This reveals two key aspects of this category of software solutions. The imaging platforms must follow a rigid consistency and standard in order to present a fast learning curve for users. Matching between the system and the world is a crucial aspect whose violation can be a severe risk. A wrong dimen-

Table 3.2: Heuristic violations reported from evaluators [131]

Violation	Heuristic Violated	Rate
Rotation and restore not easily distinguishable	Consistency	4
Layout: *Set Default* doesn't match button functionality	Consistency	2
Overlays: *Set Default* doesn't match button functionality	Consistency	2
Layout: 3D position buttons, confusing icon.	Consistency	2
Double arrow button, not clear functionality	Consistency	3
Button panel refers to MPR as 2D	Consistency	3
Multi slices: no patient name reference over views	Consistency	4
Anonymize: button icon not clear	Consistency	4
Filters: button icon not clear	Consistency	3
3D position button is disabled. No explanation is given	Visibility	2
Planning: not clear user to click *Done* before adding a new vector	Visibility	1
Loading: sub-folder look empty	Visibility	3
Loading: impossible to open Dicom tree or sub-folders	Visibility	3
Inability to localize what image you are looking at	Match	3
Planning: 3D object not scaled	Match	4
Flip in button and hotkeys 'H' and 'V' reversed.	Match	3
"Interpolation" in the context menu causes a misunderstanding	Match	2
No visual advice that series are locked	Match	4
Planning: change the surgical tip of the added vector, nothing happens	Match	3
Vectors list: deleted item retain their numbers	Minimalist	2
Measures list: deleted item retain their numbers	Minimalist	2
Interpolation function confusing	Minimalist	1
Planning: button functionalities not clear	Memory	4
Impossible to change measure color	Flexibility	0
User should set preferred action for mouse buttons	Flexibility	1
User should select zoom or scroll functionalities for mouse wheel	Flexibility	1
Cropping should allow the user to define a box and rotate it	Flexibility	3
Error messages not very clear	Messages	3
Close button causes a shutdown without a message	Closure	4
Some icons used for closing all button and closing single environment	Closure	3
Impossible to reopen tools bar from working window	Undo	2
Impossible to set bookmarks in notable slices	Undo	2
Impossible to undo a single action, possible to reset all actions	Undo	2
Brightness and contrast used, instead of WW WL	Language	3
Maxillofacial misspelled	Language	3
Cannot move labels out of the way. Overlap of multiple measures	User	4
Crosshairs in ellipse cover up too much anatomy	User	4
Planning: Overlaps of labels and anatomy	User	4

sional relationship or unclear orientation of the images can lead to serious risks in diagnosis and planning. One limitation of heuristic analysis is that the assessment is performed on the technology in isolation, without considering the users, processes, or environments the device will be used in. As a result, some usability issues may only come to light once the technology is considered in the context of the system of use. This is especially true for software to be used in a hospital environment and in pre-surgery and surgery environment. The developed Biomedical Imaging Platform All-in-One is thought to be used in a healthcare environment, so it stands to reason that it is going to be interfaced with external informative systems, such as Electronic Health Records (EHRs), Computer-Aided Management Systems (CAFMs) for healthcare facilities [59, 60] and, if needed, with other generic workplace management systems [56]. To help overcome these challenges, a usability test must also be designed and submitted to potential users, such as orthopedics, surgeons, and radiologists.

Chapter 4

Usability test of the Biomedical Imaging Platform All-in-One

After the heuristic evaluation, the work moved on to the usability test. On this subject, we identified a gap in the literature regarding the formalization of the implementation and administration of usability tests for Software as Medical Device (SaMD), such as the multimodal Biomedical Imaging Platform All-in-One. This deficiency obstructs the comparison of results between different tests for the same class of SaMDs. The objective of this study is to provide a reproducible usability testing framework for SaMDs by establishing a standardized protocol that ensures repeatability and facilitates comparison. The usability testing protocol was applied to the Biomedical Imaging Platform All-in-One for validation. The results of the usability testing protocol applied to the case-study software demonstrate good values of the software's effectiveness and efficiency, along with user satisfaction. Moreover, the outcomes confirm the protocol's robustness, applicability, and reproducibility, enabling reliable usability assessment and comparative analysis of medical software. [1]

[1] A pilot test of the usability testing was conducted and presented in "Design and validation of a usability test for a 3D biomedical image viewer" at the *Eighth National Congress of Bioengineering*. This chapter has been published as "Enabling reliable usability assessment and comparative analysis of medical software: a comprehensive framework for multimodal biomedical imaging platforms" in *Health Technology* [31].

4.1 Introduction

In healthcare systems, new technologies and Medical Device (MD)s have become more significant during the latter part of the last century. [108]. The increasing use of MDs contributed to the improvement of health care and its quality, leading to new requirements regarding their characteristics and safety. In this sense, the usability of MDs gained interest, making the adoption of standards such as IEC 62366-1:2015 [2] essential for new products released in the market. As a consequence, usability is now part of the whole risk management process and is systematically considered and evaluated in the design, construction and implementation of MDs. The content of standards and scientific literature dealing with usability shows that the concept of usability is widespread [41].

Among standards addressing usability, UNI EN ISO 26800:2011 describes ergonomic principles for interfaces to improve safety, performance and usability by analysing the target audience, environment, goals, and expected results [1]. EN ISO 9241-11:2018 provides a framework for understanding the concept of usability and applying it to situations where people use interactive systems, other types of systems (including built environments), products (including industrial and consumer products), and services (including technical and personal services). The standard describes usability as "the extent to which a system, product, or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [3]. Finally, EN ISO 25066:2019 addresses various usability testing approaches such as inspection and heuristics review of the interface, tests and post-market surveys [4]. The term usability is, therefore, very broad, and additional features to the list of ISO parameters are included in the scientific literature, such as ease of use, learnability, flexibility, attitude, and memorability [118], considered very important in promoting positive outcomes for both healthcare professionals and patients [41].

Given the complexity of MD usability issues across sectors, disciplines such as Human Factors Engineering (HFE) actually study the topic. HFE addresses the user interface in addition to the design of tools, machines, and systems, by taking into account human capabilities, limitations, and characteristics [57]. The goal is to ensure safe, comfortable, and effective use. Ergonomics, usability engineering, and user-centred design are considered synonyms [44]. The elements that influence user experience, especially in the use of medical software, need to be studied extensively because the way

people perceive and use informatics tools can affect how well they understand the data and the outcomes of the analysis [58, 88]. Because of this, it is becoming more crucial to research the usability of medical software because a poorly made or designed interface might be challenging to use and lead to mistakes when using it. [90]. Problems affecting medical device interfaces can also be the cause of recalls due to software interface errors. In this sense, interface usability testing is a highly-effective methodology for identifying usage errors and barriers, as well as a practical method for improving the efficacy and efficiency of products, services and systems [48].

Usability testing can be performed during the development stage to reduce errors and optimize the design. Furthermore, testing can be repeated after the product has been distributed, allowing for a prompt identification of problems, and highlighting critical issues [35]. Conducting usability evaluations can face numerous obstacles, such as the lack of adequate and validated guidelines [115] and clear indications of the environment, parameters, data type, and team characteristics. Moreover, the costs associated with evaluating the usability of MDs can be a major limitation. The description of the current context related to usability testing shows that there is an urgent need to define a systematic method to design a cost-effective usability validation procedure.

Zhang et al. [156] modified the existing heuristic evaluation method for software usability evaluation, applying it to medical devices, and using it to evaluate the patient safety of the device by identifying and evaluating usability problems. Shin and Lee [121] proposed a method to design and implement a time-cost effective test procedure for a comprehensive usability validation test by selecting the scenario with the lowest time-cost, starting from an activity diagram based on uFMEA (Use Failure Mode and Effects Analysis). Quality Function Deployment (QFD) is a structured approach to defining customer needs or requirements and translating them into specific plans to produce products to meet those needs. Despite being more suited to different kinds of analysis [87], it may also be used as a tool to develop usability evaluation models. However, using the QFD methodology has limitations, being not a procedure that searches for the optimal solution, but rather a technique designed to match designers' and users' needs in designing a product. Moreover, the QFD methodology is insufficient to understand the correlations among the physical design factors of a product, which is indeed a crucial outcome of usability tests [65].

The literature review reveals that, at the moment, all usability tests for (SaMD) are designed in compliance with international standards or with few changes to the protocol, as described by Zhang et al. [156]. The review also reveals a lack of formalization in the implementation and administration of such usability tests, preventing the comparison of results from different tests for the same class of SaMD. The scope of this study is to provide a reproducible usability testing protocol for medical software to ensure repeatability and comparisons of similar SaMD for visualizing medical images and data.

The developed protocol was applied to a Multimodal Biomedical Imaging Platform All-in-One, designed by Imaginalis S.r.l. (Sesto Fiorentino, Italy) with different users tested in a real-life scenario. To ensure the reproducibility of the testing protocol, custom software was designed and developed for recording and calculating test results (see Section 4.2.5).

4.2 Materials and Methods

The GUI of the tested software allows users to view medical data, images, and other relevant information. The designed protocol admits only two modalities for mirroring the screen during the test: cabled (via High-Definition Multimedia Interface (HDMI) connection) or wireless. These modalities are available on all possible devices on which the tested SaMD can be installed regardless of its type: workstations, laptops, and tablets. These modalities are mandatory and cannot be changed to reduce variability and ensure the reproducibility of the testing protocol. Therefore, the machine on which the SaMD is installed must have an Operative System (OS) that supports display mirroring and at least one HDMI port or a stable WiFi connection.

According to the IEC 62366-1:2015 [2], the goal of usability testing is reducing risks. However, for this specific scope, the fastest access to all basic functionalities and overall user satisfaction was identified as the main goal beyond the evaluation of the risks. Moreover, the usability test may also reveal missing useful functionalities in the current release of the software ([92]) which were not detected with analytical approaches.

4.2.1 The Multimodal Biomedical Imaging Platform

The platform under testing is intended for pre-, post-, and intra-operative usage in the human and veterinary fields. Despite the imaging platform sup-

ports the visualization of any DICOM image, it was mainly developed to satisfy the needs of orthopaedics, and, therefore, is mainly focused on Computed Tomography (CT), fluoroscopy, and radiography. However, at the moment of preparing the first iteration of the usability testing procedure, the CT modality was the one with the highest level of readiness, leading us to limit the tasks in the administered tests, as described in this work, to CT acquisitions only. The proposed software provides a three-dimensional (3D) representation of the CT volumetric data, including Multi-Planar Reformation (MPR) and 3D volumetric reconstruction views (see Fig. 4.1). The basic controls (zoom, pan, rotation, and scroll) enable the series navigation. The MPR functionalities include image enhancement, colour inversion, slab thickness control, annotation and measurement, histogram window control, and intensity presets. 3D functionalities cover different volume crop and rendering modes, auto-play, transfer functions control, and preset saving. Moreover, advanced functionalities such as compare, surgical planning, and multislice viewing are supported.

Figure 4.1: Multimodal Biomedical Imaging Platform All-in-One [31]

The set of proposed functionalities is based on the interview of a large group of orthopedics, radiologists, and veterinaries. Additionally, the imaging software systems already present in the market were analysed. The GUI was developed according to Gestalt visual design principles. Moreover, the law that predicts the time taken to acquire a target (Fitts) and the human

choice-reaction time (Hick-Human) were taken into account ([32]).

After completing the first proposal of the GUI design, the method of heuristic evaluation proposed by Zhang was applied [131]. The result of the evaluation was analysed and the requested modifications and fixes were implemented. The heuristic evaluation process was repeated until only minor issues remained. At that point, the development was suspended and the current version of the software was used for usability testing.

4.2.2 Participants

The complexity of the software and the environment in which it is meant to be utilized determine the optimal number of participants. According to research [36, 135, 141], between three and twenty people can yield trustworthy results, with five to ten being a good start. Generally speaking, a higher number of testers is needed for more difficult, risky initiatives, but fewer participants are needed to test more creative ideas. Following these considerations, twelve volunteers were recruited to compose the testing population.

4.2.3 Environment

According to the above-mentioned requirements, only two modalities are allowed for mirroring the screen during the administration of the usability test: HDMI cable and wireless. In both modalities, the testing occurred in person.

HDMI testing

The environment for the HDMI test consisted of two adjacent rooms: one room designated as the test room and a second one as the observation room (see figure 4.2).

The chosen rooms were adjacent to allow an HDMI cable to be passed through. That enabled the duplication of the test machine screen on the monitor of the observer. As a precaution, the test machine was not connected to the Internet, while superfluous operative system processes were suspended to avoid compromising software productivity.

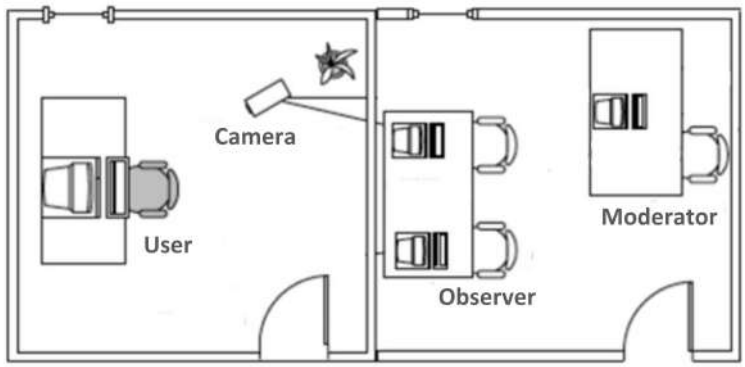


Figure 4.2: Test Environment Scheme [31]

Wireless testing

For wireless testing, no external cable was needed for mirroring, as a Google Meet session was set up on the same machine where the tested software was installed to share the screen. In this scenario, the superfluous operative system processes were not suspended, to evaluate their impact on the performance of the SaMD during the test.

4.2.4 Test Conductors

The testing procedures require a minimum set of two people for conducting the test:

- **Moderator**: in charge of managing the progress of the test; responsible not only for administering the tasks, but also for observing the user's facial expressions, resolving any problems, and answering the possible questions arising during the session;
- **Observer**: responsible for reporting the user's performance of the tasks, tracking down the time taken to perform each task, and leaving comments on eventual issues and user's difficulties. In specific cases, when reporting the task performance while taking notes may result complicated, the presence of more than one Observer can be useful.

Optionally, a third person (namely the **Recorder**) can be involved. The Recorder observes and analyses the footage coming from the camera, which

frames the user from the entire scene's perspective during the test. In the absence of this third person, the footage recorded with the external camera can still be analysed after the conclusion of the test.

4.2.5 Equipment

A list of tools needed for carrying out a usability test for both HDMI and wireless modalities for touchscreen and mouse/keyboard configuration was defined in A.1. The settings for the different modalities/configurations are quite similar and are described in Tables A.1 and A.2.

Custom *Stopwatch* software was developed to ease the observer's tasks. The software is developed in C/C++ and the source code has been made publicly available at `https://github.com/eletiri93/Stopwatch`. The software enables the observer to record the amount of time spent on each task during the usability test while also noting any noteworthy user behavior: the GUI shows the current task, a stopwatch, and a space for taking notes. On the right side of the screen, a table summarises the recorded times and notes. After the completion of each task, the observer is able to export a CSV (Comma Separated Values) file containing the recorded times and notes.

4.2.6 Exploratory Tasks and Specific Scenarios

Based on the experience acquired during the heuristics evaluation, the list of 55 exploratory tasks (see A.6) was produced. The objective was to ensure that participants did not become overly fatigued while covering all of the most crucial capabilities within realistic time constraints. The tasks had to be clear, short, and as independent as possible (e.g., the failure of one task should not compromise the success of the following tasks). Moreover, four specific scenarios were developed with the help of an external radiologist consultant (see A.7). The exploratory tasks and specific scenarios can be modified to allow the testing protocol to be tailored to the specific application and the provided functionalities of the tested software.

4.2.7 Test Evaluation

Tests were evaluated in terms of effectiveness, efficiency, and user satisfaction. The first two parameters were evaluated with the aid of a purposely

developed *Stopwatch* software, notes taken by the Observer, and camera footage. User satisfaction was evaluated by using a post-test questionnaire administered at the end of each test session.

More specifically, the effectiveness evaluates the participant's capacity to finish each suggested task, independent of the amount of time required. It is evaluated through the following score system:

- Score: 1. Failure. The user fails to complete the task, despite some suggestions
- Score: 2. Partial success. The task is partially completed or completed after suggestions
- Score: 3. Complete success. The task is completed without any difficulties or suggestions

On the other hand, efficiency measures how fast each participant completes each assignment and is evaluated by timing the performance of each task. The above-mentioned *Stopwatch* software was used to record timestamps. The recorded value corresponds to the time interval between the end of the reading of the task by the moderator, and the moment when the user asserts the completion of the task.

Finally, scores given to each of the statements proposed in the post-test questionnaire were analyzed to assess user satisfaction. The agreement scale was used, where **5** represented the *fully agree* and **1** represented *strongly disagree* options. All statements were designed to have consistent meanings (e.g., "The software is intuitive to use", "I had no problem using the basic features")

For comparison purposes, it is important to evaluate only the functionalities common to all the compared SaMD. This enables a direct comparison of effectiveness and efficiency, task by task, and scenario by scenario. It is also crucial for assessing the user satisfaction evaluation: if the tasks and scenarios differ between the assessed SaMD, user satisfaction may be higher for the simpler test suite and lower for the specific and innovative features tested. However, the methods for comparing similar SaMD using the proposed testing protocol are beyond the scope of this paper.

4.2.8 Statistical Analysis

Efficiency, effectiveness, and user satisfaction variables' distribution were tested using the Shapiro-Wilk test. The variables were statistically described as mean ± standard deviation (SD) for normally distributed quantitative data; median and interquartile range (IQR) for non-normally distributed data; and frequency count and percentage for qualitative data. The difference between the groups of participants and the hardware modalities was evaluated using the independent *t-test* for normally distributed quantitative variables and the Mann-Whitney test for non-normally distributed quantitative variables. The relationship between efficiency, effectiveness, and user satisfaction was tested with the Spearman test. When a significant difference was detected, Cohen's d (for normal distribution) or Cliff's delta (for non-normal distribution) was calculated as a measure of the difference. The significance level for all tests was set to 0.05 ($p < 0.05$).

4.2.9 Pilot Testing

Three weeks before the test administrations, the *Pilot Test* in touchscreen mode was carried out to validate the proposed test method, as well as the above-described environment, equipment, and tasks. Four persons were involved in the pilot test administration: a moderator, two observers, and a user. The observers prepared all the necessary environment and equipment, while the moderator made sure that all the procedures were followed correctly. The pilot test was very useful, as it uncovered some task-related issues, such as the duration of some of them or the used lexicon.

4.2.10 Testing Procedure

After the administration of the pilot test and its further analysis, the final testing procedure was set up. Before the participant arrived, the moderator and the observers verified the instrumentation, making sure that nothing was missing or abnormally working by using a dedicated checklist (see Appendix A.2). A single testing session performed by one user took about 1.5 hours and consisted of the following steps:

1. Introduction to the test, including the software description and the desirable goals (about 5 mins). See A.3

2. Signing of the recording agreement (about 2 mins). See A.5

3. Compilation of the pre-test questionnaire (about 5 mins). See A.4
4. Execution of exploratory tasks (45-50 mins). See A.6
5. Specific scenarios (10-15 mins). See A.7
6. Compilation of the post-test questionnaire, including user's feedback (10-15 mins). See A.8

Figure 4.3 illustrates the entire workflow of the testing protocol as described in this section.

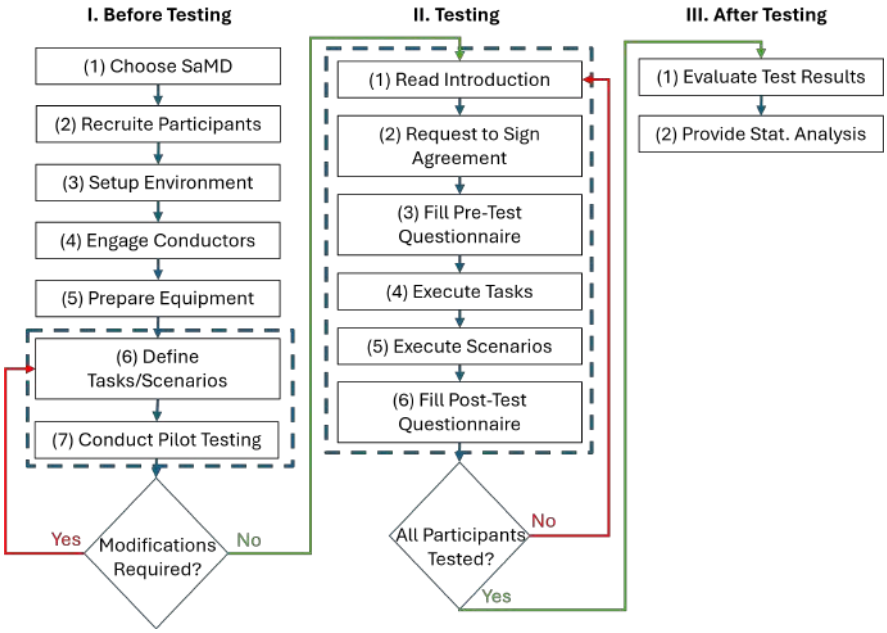
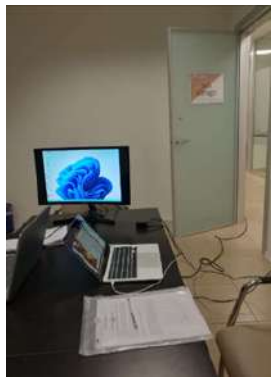


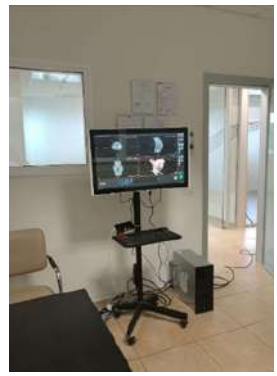
Figure 4.3: The flow chart of the proposed testing protocol illustrates the sequence of steps to perform before testing (I), during testing (II), and after testing (III). Dashed-border boxes represent the steps that could be repeated more than once [31]

4.3 Results

Tests were conducted from October 12, 2022, to November 24, 2022. Figure 4.4 illustrates the environment and equipment configuration.



((a)) The observation room: laptop for the moderator; laptop for the observer; printed task list and supportive documentation; external monitor, connected via an HDMI cable to the test workstation [31]



((b)) The test room for touchscreen modality: touchscreen display placed on the support with height regulation; workstation with tested software installed, connected through an HDMI cable to an external display; smartphone, placed on support with angular regulation.

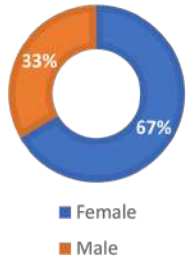
((c)) The test room for mouse/keyboard modality: traditional display, mouse and keyboard placed on a large desk; comfortable chair; workstation with tested software installed, connected through an HDMI cable with an external display; smartphone, placed on support with angular regulation; camera placed on a tripod.

Figure 4.4: Test Environment & Equipment Setup [31]

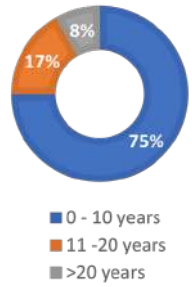
According to the observations, done in Section 4.2.2, twelve users were recruited for the testing. Six of the participants had a clinical background, and six had an engineering background (see the detailed distribution of the sample of users in figure 4.5). Seven tests were conducted via HDMI connection and five using Google Meet. Ten tests were administrated in touchscreen modality, while the remaining two were administrated using a mouse/keyboard.

The statistical analysis of the participants did not reveal any significant differences between the groups of engineers and clinicians. Only two explorative tasks showed a significant difference in efficiency (task 6 with

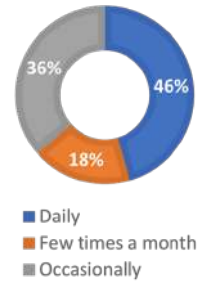
Cliff's $\delta = -0.6944$, and task 33 with Cliff's $\delta = -1$). One task showed a significant difference in the efficiency between the two modalities (HDMI vs. wireless - test 7 with Cliff's $\delta = -0.8857$), and one showed a significant difference, again on the efficiency only, between mouse/keyboard vs. touchscreen (task 19 with Cliff's $\delta = -1$). Consequently, the results of all participants were combined for evaluation. Data collected during the pilot test was not included in the analysis, as the final version of the protocol was edited afterward due to the subsequent considerations.



((a)) Gender

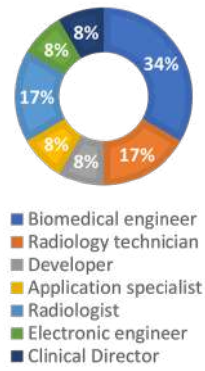


((b)) Experience



((c)) Usage

((d)) Approach to new tech



((e)) Professions

Figure 4.5: Pretest questionnaire results. Distribution of test samples in terms of gender (a), experience (b), frequency of use of medical software (c), approach to new technologies (d), and professions (e) [31]

Figure 4.6 shows the statistical description of the effectiveness as frequency count and percentage for each explorative task, while figure 4.7 shows

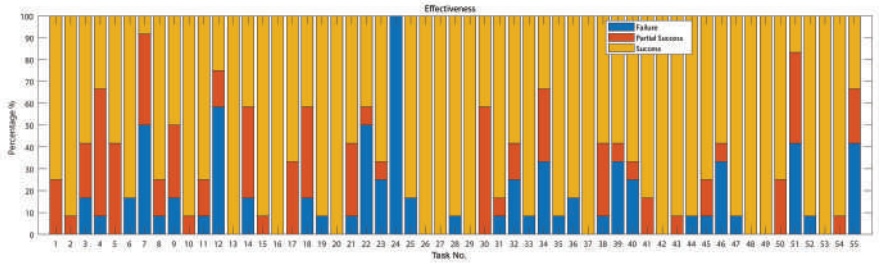


Figure 4.6: Effectiveness for explorative tasks, evaluated as frequency count and percentage obtained for each task [31]

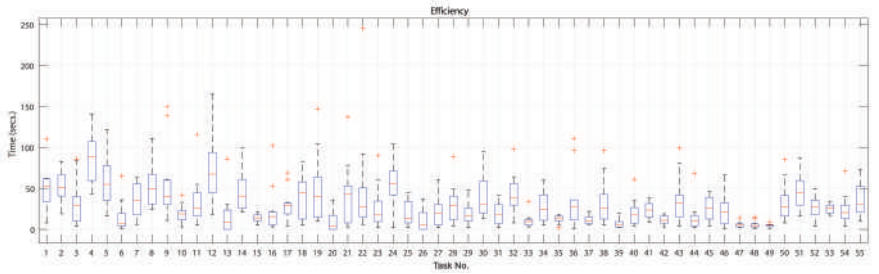


Figure 4.7: Efficiency evaluation for explorative tasks. In the graph, the box represents the IQR, spanning from the 25^{th} to the 75^{th} percentile. Higher IQRs suggest significant variability in the time taken to complete the task. Smaller IQRs with medians (red line inside the box) close to the bottom indicate tasks completed very quickly. Outliers are denoted as (+) [31]

the statistical description of the efficiency for each task in terms of median and IQR. figure 4.8 presents effectiveness and efficiency for the specific scenarios.

Figure 4.9 reports the statistical description of user satisfaction as frequency count and percentage for each assessing question.

The statistical analysis of efficiency and effectiveness indicated a statistically significant negative monotonic relationship ($\rho = -0.658495$, see figure 4.10 and Fig. 4.11), while no statistically significant relationship with satisfaction was found. Two outliers were detected.

At the end of testing, user feedback was collected to improve the existing functionalities in future software updates.

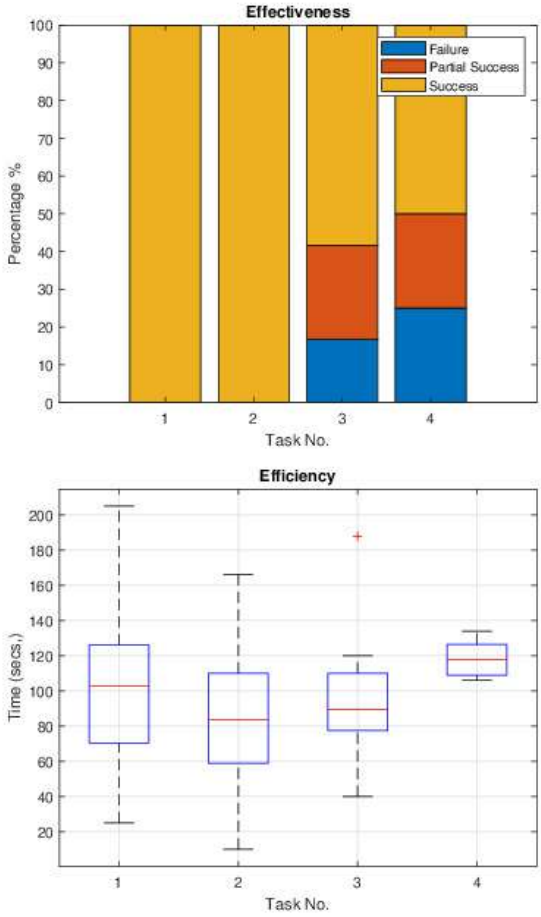


Figure 4.8: Effectiveness and efficiency evaluation for specific scenarios. Effectiveness: frequency count and percentage. Efficiency: median, IQR, range, and outliers [31]

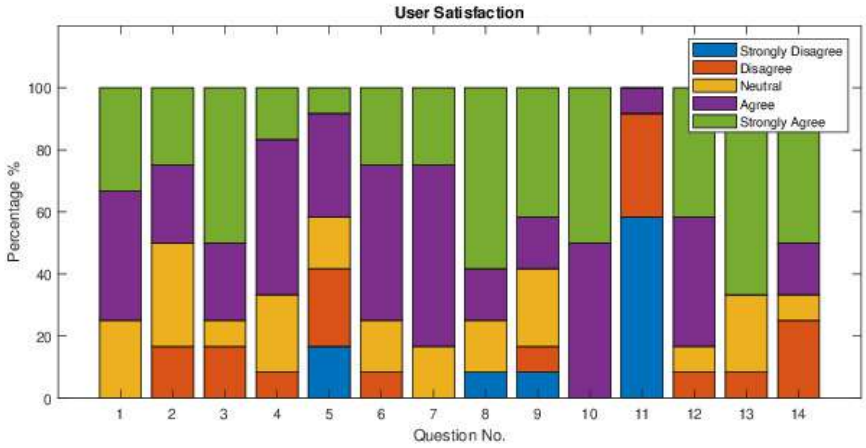


Figure 4.9: Post-test questionnaire. Responses in a range from 1 to 5, were evaluated as frequency count and percentage obtained for each question [31]

4.4 Discussion

A number of considerations were made at the end of the first usability testing iteration.

As demonstrated in Section 4.3, the analysis of the two environmental testing modes (HDMI and wireless) revealed no significant variations. This suggests that any of the two modes can be chosen for testing, based on hardware specifications and the preferences of the test conductors. Furthermore, the results highlight that the suspension of superfluous operating system processes on the test machine has no impact on the performance of the software and the test suite, and it is therefore discretionary.

Remote testing may allow a larger group of participants while obtaining similar results as those performed in-person [7, 23], but it shows some limitations and drawbacks. The main challenge of remote testing consists of the participants being in charge of the whole environment and equipment set-up, which may lead to an unwanted and unexpected set of problems. Another issue that might be faced is the possible different time zones for the participants and the test conductors. Due to these reasons, the developed usability protocol takes into consideration only in-person testing, minimizing the sources of variability that remote testing could introduce, thus improving

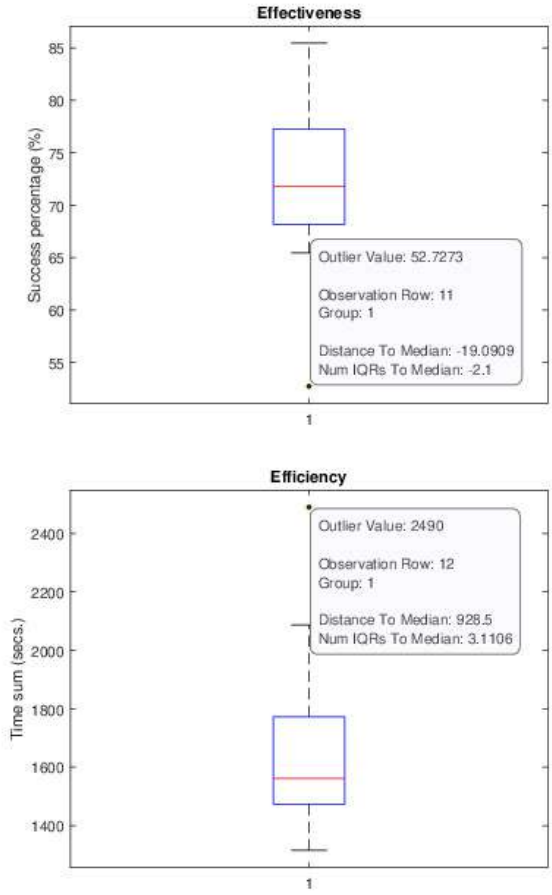


Figure 4.10: Top to bottom: effectiveness in terms of success percentage; efficiency in terms of overall time spent. Two outliers were detected [31]

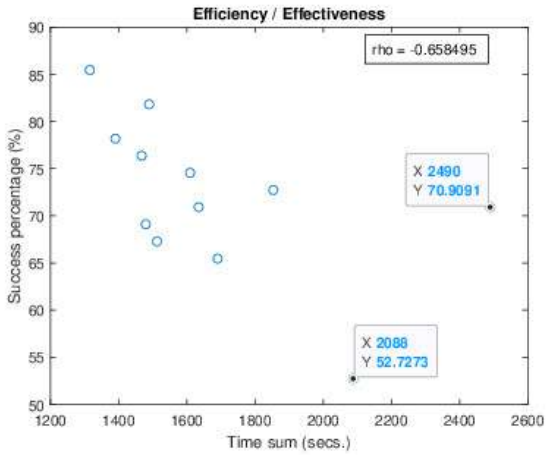


Figure 4.11: Correlation between effectiveness and efficiency. Each point on the scatter plot represents the relationship between effectiveness (y-coordinate) and efficiency (x-coordinate), providing a visual representation of the distribution and spread of data and confirming the presence of two outliers [31]

its reproducibility and consistency.

It is worth noting, that the test description and the terminology used can greatly influence the success (or failure) of the test results. It is indeed very important to carefully formulate the tests and the tasks, asking, when possible, for the help of an external specialized end user for implementing the appropriate terminology and lexicon.

The results show that none of the participants from the selected testing group chose to learn a new technology by reading the technical documentation.

Overall, the results for user satisfaction showed that the tested software was deemed intuitive to use and that the majority of users believed that it could improve their work. Users did not encounter any difficulties using basic features in the mouse/keyboard or in the touchscreen configuration. However, the vast majority of them reported the need for initial support in the software usage. The touchscreen was preferred by the majority of involved users.

The results also highlight that statistically significant differences emerged only on the efficiency of four exploratory tasks: engineers were faster in performing tasks 6 and 33, task 7 was completed faster with HDMI compared to the wireless setting, and task 19 was completed more quickly with mouse than with touchscreen. This aspect underlines that, together with the HDMI modality and the mouse configuration, the academic background/profession is the only personal feature among the analyzed five users (Figure 4.5) that can slightly influence the performance of the test in terms of time. Finding no significant differences in the effectiveness points out that the successful completion of each task is not related to any personal background or administration modality. Therefore, the effectiveness outcomes of the tested software reveal that some specific tasks (e.g., task no. 7, 12, 22, 24) actually shed light on possible weak points of the GUI. Moreover, the statistical analysis of efficiency and effectiveness indicated a statistically significant negative monotonic relationship, which suggests that as more time passes during task execution, the probability of successfully completing the task decreases.

4.5 Conclusion

This work presents a usability study on medical imaging software, focusing on the Multimodal Biomedical Imaging Platform All-in-One by Imaginalis

S.r.l., compliant with the DICOM standard. The article provides a formal protocol for repeatable analysis, allowing for comparison of tests on similar SaMD for the visualization of medical images and data.

The protocol adheres to international standards, covering participant selection, testing environment setup, minimum number of required testers, as well as their roles and specific assignments, equipment requirements for HDMI and wireless modalities, and hardware interfaces (keyboard/mouse, touchscreen). The usability test comprises exploratory tasks and specific scenarios for evaluating software functions in real-life tasks.

Results were evaluated in terms of effectiveness, efficiency, and user satisfaction.

The exploratory tasks and specific scenarios are the only aspects of the proposed protocol that may need adjustment for different visualization SaMD. This is because different software may have varying functionalities.

Thus, the proposed usability testing framework enables reliable usability assessment and comparative analysis of medical software. Besides, the obtained results can serve as a reference for comparing biomedical imaging platforms under development or ready for release.

Chapter 5

Segmentation and 3D modeling for the Biomedical Imaging Platform All-in-One.

This chapter introduces the topic of the bone segmentation and 3D modeling workflows for diagnosis and planning, to be integrated into the user-centered Biomedical Imaging Platform All-in-One. In particular, it offers an overview of bone segmentation and 3D modeling, discussing their significance in diagnostics and planning. The first section introduces the challenges associated with image segmentation and explores its various medical applications in the context of 3D modeling. The second section narrows the focus to bone segmentation using CT, presenting different categories of segmentation techniques while discussing their strengths and limitations. Finally, the chapter examines CBCT, detailing the key differences between CBCT and conventional CT technology.

[1]

[1]The part of this chapter related to bone CT segmentation has been presented at the IUPESM World Congress on Medical Physics and Biomedical Engineering 2022 "Bones segmentation techniques in computed tomography"

5.1 Segmentation in Medical Imaging

In the field of medical imaging, segmentation refers to the task of identifying and delineating specific anatomical or pathological structures from 2D or 3D medical images, such as Magnetic Resonance Imaging (MRI), CT or ultrasound scans [77]. In practice, the process involves dividing the image into multiple segments, where each segment represents a different object or structure of interest.

This process is crucial for many clinical applications as it is the basis for accurate diagnosis, surgical and treatment planning and monitoring disease progression. Segmentation facilitates the isolation and quantification of anatomical features, enabling clinicians to assess the dimensions, morphology, and attributes of abnormalities, including tumors, thereby allowing for well-informed decision-making. Besides its role in diagnosis, segmentation is crucial in the field of radiotherapy, where accurate targeting of tissues is essential to deliver the optimal dose to the tumor while concurrently reducing exposure to adjacent healthy structures. Other important aspects of segmentation in medical research and education include its capability for quantitative analysis of anatomical variations and as a training tool for medical professionals using interactive 3D models.

Image segmentation can be broadly classified based on the level of the user interaction into manual, semi-automatic and automatic methods [77, 125]. Manual segmentation is done entirely by a human expert, involves detailed outlining of the structures of interest, considered the gold standard as it includes expert knowledge [125]. However it is time-consuming and prone to interobserver and intraobserver variability, as results may vary depending on the observer or even when the same observer does the segmentation multiple times.

Semiautomatic segmentation methods involve some level of user input, which can be before, during or after the segmentation process to fine-tune the results [125]. These methods balance accuracy and efficiency, require less user interaction than manual methods but still leverage expert guidance to improve the segmentation quality. So they are more time efficient than manual segmentation but still need some user input.

On the other hand, automatic segmentation methods do not require any user interaction, so they are the most time-efficient. These methods use various techniques, such as rule-based, statistical models and machine learning algorithms, to segment medical imaging data.

In the general sense, it is possible to divide segmentation into two other categories: semantic segmentation and instance segmentation. In semantic segmentation, each pixel or voxel of the image is assigned a label belonging to its class; in other words, all objects of the same class have the same label, for example, "bone" or "muscle". This is the main difference between instance segmentation, which would differentiate individual instances of the same class and assign a unique label to each structure "bone 1", "bone 2", etc. Both are relevant in medical imaging, where we often want to know if an anatomical structure is present and who it is.

In medical image segmentation, the review of performance is mainly based on the geometric accuracy achieved in delineating the anatomical structures. High geometric accuracy ensures that the segmented regions more accurately depict the true boundaries of the anatomical structures. This fact makes it very important to be employed for reliable diagnosis, treatment planning, and surgical guidance. Metrics such as DC, JI and Hausdorff Distance are commonly used to quantitatively assess the precision and robustness of segmentation algorithms [126].

5.1.1 Usability of Segmentation Framework

While geometric and dimensional accuracy is essential for assessing the quality of medical image segmentation tools, it alone does not meet all the needs of a clinical environment. Therefore, the usability of the software is also a critical factor, as user-friendly interfaces and workflows significantly enhance the efficiency of doctors and technicians. Virzi et al. [140] and Mandolini et al. [91] evaluate the usability of medical segmentation frameworks based on the following factors:

- Automatization degree: the amount of manual interaction required by the user [140].
- Segmentation time: the time required for the segmentation [140].
- 3D visualisation: ability to represent a 3D model realistically [140].
- Supported Operative System (OS): supported operative systems [140].
- Potential extension (plugins): the ability of the software tools to be freely extended by other plugins [91].
- Training time: time required to start using the tool [91].

The most crucial of these criteria was "Automatization degree" [91], since CT scans must be processed rapidly and without regard to the technician's experience level. A major factor in reaching this degree of efficiency is the use of automatic tools. Other crucial elements are "Segmentation time", "Training time" and "Cost" as hospital segmentation technologies need to be easily operated by physicians without requiring a lot of training.

5.2 3D Modeling for Diagnosis and Surgical Planning

For the past few decades, CT and MRI have been the primary sources of imaging data for guiding surgical planning. However, in cases involving complex anatomy, standard two-dimensional (2D) imaging often fails to provide sufficient detail. In response, the adoption of three-dimensional (3D) models in clinical and surgical settings has proven highly beneficial.

Today, 3D modeling is widely used across various surgical specialties for purposes such as medical education, diagnosis and Virtual Surgical Planning (VSP). Personalized 3D models of the targeted anatomical region can be created, allowing for a detailed examination from multiple perspectives. VSP offers several benefits including:

- Assistance in choosing the most suitable surgical approach, such as determining the correct implant [97], [154].
- Reduced need for extensive dissection, soft tissue disruption, and repeated adjustments during a surgical procedure, potentially leading to shorter surgeries and less blood loss [97], [159].
- Virtual surgical practice, which helps surgeons shorten their learning curve [97], [145].
- Enhanced communication between doctors and patients regarding the preoperative plan [97], [144].

With advancements in 3D printing technology and the development of materials that mimic the real properties of tissues, organs, and fluids, surgical planning and simulation have reached a new level of precision in intervention protocols. Using digital data generated by CT scans, healthcare practitioners

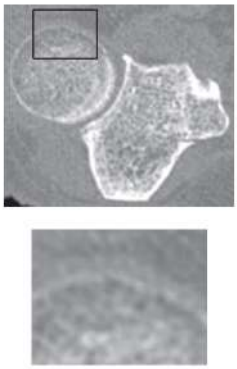
can develop highly accurate and detailed models of both normal and pathological anatomical structures. This capability improves the visualization of intricate structures and enhances surgical techniques but also facilitates the customization of prosthetics, implants, and surgical guides, enabling better alignment of therapies with each patient's unique anatomical characteristics.

Furthermore, the introduction of Augmented Reality (AR) represents a technological advancement in the context of complex surgical procedures. AR is now used in guided implant surgery applications, with both in vitro and in vivo studies demonstrating its effectiveness. In this approach, 3D models in Standard Tessellation Language (STL) format, generated from segmenting images acquired via CT, are uploaded into specialized applications designed for AR systems. This integration allows surgeons to visualize patient-specific anatomy in real-time during procedures, enhancing precision and improving outcomes by providing a more intuitive understanding of intricate anatomical structures.

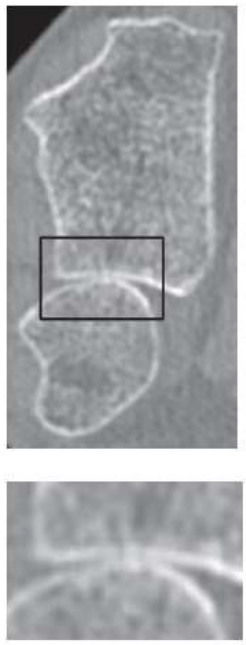
5.3 Bone Segmentation from CT

Since this project is focused on orthopedic applications, the following paragraph provides a brief overview of the aspects of bone in computed tomography before discussing various techniques for bone segmentation. CT imaging helps identify different types of bone tissue because of its excellent contrast, even when anatomical features absorb similarly. The outer layer of bone, called cortical bone, is dense and compact, making it appear very bright on CT scans, especially in the shaft (diaphysis) of the bone. In this area, the cortical layer surrounds the medullary cavity, which holds bone marrow. Near the joints, the cortical layer gets thinner, and the medullary cavity is replaced by cancellous bone. This spongy bone is more porous and less dense than cortical bone, primarily serving to absorb and distribute forces at the joints. The inner part of the bone near the joints looks spongy and textured on CT images. The thin cortical layer also makes it harder to distinguish the bone from surrounding tissues, particularly in older adults or those with osteoporosis, who have lower bone mineral density.

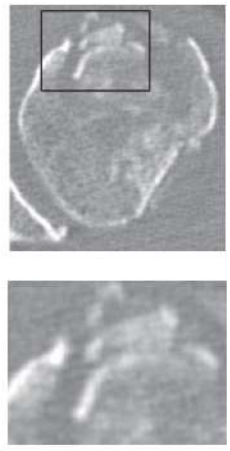
As already mentioned in the previous paragraph, orthopedic planning applications use models created from bone segmentation masks. This process involves classifying data into two groups: voxels belonging to the bone are labeled as foreground, while the rest are labeled as background.



((a)) The weak boundaries of the bone and the textured interior make segmentation difficult.



((b)) There is extremely narrow spacing between adjacent bones.



((c)) Close proximity of bone fragments leads to noise and indistinct boundaries between various tissue types.

Figure 5.1: Bone segmentation presents several common challenges. CT slices of a forearm (Figures a and b) and a fractured humerus (Figure c) illustrate these typical issues

For long bones, segmentation is usually straightforward due to the clear contrast between the bone shaft and soft tissue. However, when dealing with joints such as the knee, elbow, and pelvis with extremities such as feet, hands, and head or with fragments broken from trauma, the precise segmentation became challenging. In particular, it can present various issues:

- Weak bone boundaries: Bones are less dense near joints and in extremities, leading to a very thin cortical layer that has similar intensity to the surrounding tissue, as shown in figure 5.1(a). Segmentation algorithms often struggle to detect these weak edges, which can create gaps in the bone boundary. Additionally, osteoporotic bone can signif-

icantly reduce the contrast between the bone and nearby soft tissue. This may lead to the "leakage problem", where unwanted areas from adjacent regions are included along with the correct bone boundary.

- Varying density: The density of long bones varies significantly along their length. As a result, the intensity histograms for bone and surrounding tissue often overlap [117]. Consequently, using a global intensity estimation for segmentation can lead to inaccuracies.

- Narrow inter-bone spacing: As noted by Sebastian et al. [117], the narrow spacing between adjacent bones at the joints can make the inter-bone tissue appear brighter in CT images, significantly decreasing the contrast at these boundaries. This effect is illustrated in figure 5.1(b). The indistinct edges of the joint components can create the impression that they are in direct contact due to the partial volume effect.

- Fractured bones: A fracture near a joint may involve multiple thin-walled bone fragments that are either very close together or in direct contact. Segmenting this type of data can be very challenging. As shown in figure 5.1(c), it can be difficult to differentiate between cortical bone, cancellous bone, and soft tissue due to increased noise and partial volume effects.

This paragraph introduces methods designed to tackle the problem of bone segmentation in computed tomography, with particular attention to joints, extremities, and fractures. The algorithms presented range from simple and straightforward methods like thresholding to more advanced methods like convolutional neural networks. [2]

Medical image segmentation techniques have been classified as:

1. gray level-based,
2. region-based,
3. edge-based,
4. energy-based,

[2]This part has been published as "Bones segmentation techniques in computed tomography" [134] in Proceedings of IUPESM World Congress on Medical Physics and Biomedical Engineering 2022

5. deep learning-based,

The table 5.1 provides a summary of the examples analyzed in the following paragraphs.

5.3.1 Gray level-based

The simplest technique available for image segmentation is thresholding. The histogram shape, optimization, class separation, and geographical information provided by the co-occurrence matrix are all used to determine the threshold.

Otsu's method is frequently employed for automatic threshold selection. It divides the histogram into two classes: one representing the background and the other representing the foreground. For each potential threshold, the intra-class variance is calculated as the weighted sum of variances of the two classes. The goal is to find the threshold that minimizes this intra-class variance. The threshold that results in the lowest intra-class variance or the highest inter-class variance is selected as the optimal threshold [103]. Similarly to Otsu method, Yen's thresholding analyzes the histogram by dividing it into two classes based on a potential threshold. The technique calculates a criterion function that measures the sum of the Shannon entropy of the two classes. The entropy represents the amount of information contained in each class. The threshold that maximizes the criterion function is chosen as the optimal threshold [151].

Another widely used thresholding technique is adaptive thresholding. In adaptive thresholding, each pixel is compared to an average of the surrounding pixel. Specifically, an approximate moving average of the last pixels seen is calculated while traversing the image. If the value of the current pixel is lower than the average, then it is set to black, otherwise it is set to white. This method works because comparing a pixel to the average of nearby pixels will preserve hard contrast lines and ignore soft gradient changes.

Classic thresholding techniques are widely used in CT images to segment long bones such as the tibia [137] and femur [9]. More complex districts required the development of advanced thresholding techniques. Zhang et al. and Zhou et al. use a 3D adaptive thresholding to segment respectively calcaneus [155] and hip joint [160].

5.3.2 Region-based

Region-based technique splits regions of an image based upon similarity [52]. Upon this group of algorithms, region growing is the most common.

Region growing is used to extract a specific region from a pre-existing image based on some characteristics i.e.: intensity level inhomogeneity or edges in an image. The selection of a seed pixel in this technique necessitates prior knowledge. The seed point or pixel is selected by an operator and then pixels that share a common characteristic property are selected to grow the region [128]. The region-growing technique is usually associated with a set of image-processing operations for the visualization of small, simple and delicate regions such as tumors and lesions. Sometimes region growing can be sensitive to noise which can result in holes of extracted regions that are continuous. It can also lead to the merger of separate regions. Vasilache et.al. [139] combine region growing and wavelet filter to obtain and isolate a model of the pelvic structure.

Region splitting and merging is another technique based on region generation in an image. This technique first splits a given image into multiple sub-images and then merges them again. This process is fast and efficient with less noise generation or altogether immune to noise [46]. This method is based on the production of quad-trees, which have four branches. The branches of the quad-tree represent sub-images [71]. The image region is split into four parts or branches and then merged back together till no partitioning or splitting is possible. The main drawback of region-based algorithms for bone segmentation is the tendency to leak into tissue due to weak boundaries.

5.3.3 Edge-based

Edge-based segmentation deals with identifying and locating boundaries in an image. The edges are sharp discontinuities i.e. intensity values change. Edge detectors, also known as "masks" or "filters," are overlayed on top of an image to detect discontinuities or boundaries [5]. First-order filters (Prewitt, Sobel, Canny), which create thick edges, and second-order filters (Laplacian, zero crossing), which produce finer edges, can be used to determine the change in intensity level values of an image.

Among various edge detectors, the Canny edge detector [24] is the more accurate, reliable and faster edge detector [122], [19].

The canny operator was developed in 1986 by John Canny and is a multi-stage algorithm-based edge detector that helps in the recognition of the wide range of edges and orientations. This mask is adaptable and can be used in a wide range of applications. Its parameters can be tuned to meet specific application

Rathnayaka et al. [110] demonstrate that Canny Edge performs long bones segmentation with better accuracy compared to thresholding. In the contest of extremities bone, Pandey et. al [104] propose a survey on calcaneus bone tumor detection using Canny edge detector.

5.3.4 Energy-based

Active contours model (ACM) [70] is the benchmark energy-based technique in image segmentation. Deformation of the active contour is equivalent to minimizing an energy function, $\varepsilon(x)$, which comprises internal and external spline force as shown below:

$$\varepsilon(x) = \varepsilon_{int}(x) + \varepsilon_{ext}(x) \tag{5.1}$$

The goal of the formulation is to find a parameterized curve that reduces the weighted sum of both spline forces. The internal spline force, $\varepsilon_{int}(x)$, controls the elasticity of contour deformation based on contour tension and rigidity. The external spline force, $\varepsilon_{ext}(x)$, matches the boundary of the deformable model toward the targeted object.

Furthermore, the primary deformable model has been enhanced by the integration of past shape data. Widely used extensions include statistical shape model (SSM) [50] and active shape models (ASM) [30].

The deformable model is extensively used in bone segmentation. In the field of joints and extremity segmentation, Breheler et al. [22] and Wu et al. [149] use modified ASM. Foot and ankle CT images are segmented by Brehler using a Coupling Active Shape Model (CASM) approach. This coupling allows for the monitoring of proximity between the growing shapes, avoiding cases where a single-bone ASM would incorrectly conform to the articular surfaces of nearby bones. Wu et al. use a Registered Active Shape Model (RASM) to perform a better segmentation of pelvic bones. Ballerini et al. [14] present a method to segment each bone complex in the radiogram using a modified active contour approach. Each bone is modelled by an independent contour, while neighboring contours are coupled by an elastic force. The optimization of the contour is done using a genetic algorithm.

Graph cut segmentation is another energy-based optimization technique used for partitioning an image into foreground and background regions. In this method, the image is modeled as a graph, where pixels (or voxels in 3D) are nodes, and edges represent the relationships between neighboring pixels. Each edge is assigned a cost, which reflects the likelihood that two connected pixels belong to the same object. The segmentation is obtained by finding the optimal cut that separates the graph into two disjoint sets: the foreground (bone) and the background (soft tissue or other structures). It is initialized by some scribbles of the user to indicate which pixels are considered surely foreground or background.

In the study by Boykov and Jolly [21], graph cut was applied to segment the femur in 2D and 3D CT data. The authors showed that this method can accurately delineate the femur, even in the presence of weak boundaries and noise. Liu et al. [84] proposed a 3D graph cut-based method for segmenting the pelvic bone from CT images. This approach combined graph cut with statistical shape models to guide the segmentation and improve accuracy. The method was shown to outperform traditional thresholding techniques in regions with poor contrast and was particularly effective in complex anatomical regions such as the sacroiliac joint. Zhang et al. [157] applied graph cut to segment the vertebral body from 3D CT images. In this case, graph cut was used to handle the weak boundaries between adjacent vertebrae and the surrounding soft tissues. The method was shown to be robust in cases of noisy data and was able to accurately segment the vertebrae, even in the presence of osteoporotic bone structures. While graph cut produces globally optimal solutions for a given energy function, it requires good initialization or intensity models to guide the segmentation. Poor initialization can lead to suboptimal results.

5.3.5 Deep learning-based

Recently, deep learning has emerged as a popular research topic in image segmentation [43]. Deep learning uses filters to extract deep features and feed the concatenated feature vector into a dense layer. In particular, CNN has received the most research attention in the image segmentation field.

The first layer is the input layer, which is directly connected to an input image and has the same number of neurons as the input image's pixels. The next set of layers are convolutional layers, which operate as a feature extractor by presenting the results of convolving a given number of filters

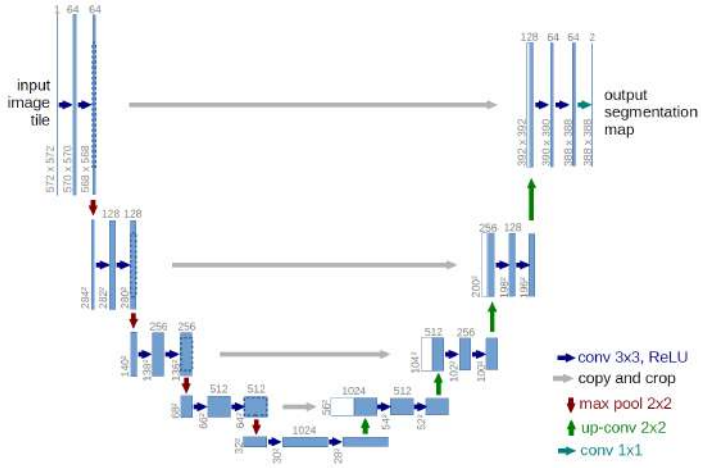


Figure 5.2: U-net architecture

with the input data. The filters, also known as kernels, come in a variety of sizes, which are determined by designers and change based on the kernel size. Each convolution layer's output is referred to as an activation map, which emphasizes the impact of applying a certain filter to the input. To impart non-linearity to the activation maps, convolutional layers are frequently followed by activation layers. Depending on the architecture, the following layer can be a pooling layer, which helps to minimize the dimensionality of the convolution's output. There are several ways for pooling, including maximum pooling and average pooling. During the backpropagation technique in the training phase, the weights of neural connections and kernels are continuously improved.

The U-net architecture is the most often used in medical imaging segmentation. Olaf Ronneberger et al. [113] created the U-net, as shown in figure 5.2, the architecture has two pathways. The contraction path (also known as the encoder) is the first path, and it is used to record the image's context. The encoder is made up of convolutional and max pooling layers stacked on top of each other. The symmetric expanding path (also known as the decoder) is the second way, and it is employed to achieve exact localization via transposed convolutions. As a result, it is an end-to-end fully convolutional network (FCN), which means it only has convolutional layers and no dense layers, allowing it to accept images of any size.

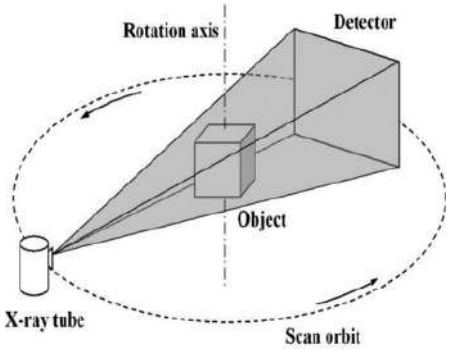


Figure 5.3: Cone Beam Computed Tomography scheme

A large number of studies that used deep learning in segmentation were published. In the field of bones extremities joints and segmentation, Jin et al. [66] and Yao et al. [150] segment rib fractures. Yao et al. use a U-net for bone segmentation and a 3D DenseNet for fracture classification.

Pranate et al. [109] evaluate and compare two types of CNN a residual network (ResNet) and Visual geometry group (VGG) for segmenting and classifying fractures and non-fractures images of the calcaneus. Folle et al. [40] use U-net to segment metacarpal bone.

5.4 Cone Beam Computed Tomography

This industrial research is based on images acquired through a commercial machine (designed, developed, and produced by the company Imaginalis srl) capable of acquiring tomographic images with cone-beam technology.

CBCT is a medical imaging modality that generates 3D representations of the scanned target. It utilizes a cone-shaped X-ray beam to capture 2D images of the target area from different angles around the patient. In contrast to the CT, that uses a fan-shaped X-ray beam that moves in a spiral path, requiring multiple rotations and more complex machinery.

The CBCT device consists of an X-ray source and a detector panel that are placed on opposite sides of the patient.The X-ray generator and sensor do a rotational movement around the patient, taking pictures from different angles. The rotation allows for a complete 360-degree visualization of the area of interest. As the X-ray source and detector rotate, multiple 2D X-ray

projections are acquired at different angles, a scheme is provided in figure 5.3. These projections contain information about the density of tissues along the X-ray path. CBCT generally exposes patients to a lower radiation dose due to its faster process, while CT involves higher radiation levels.

After data acquisition, the collected 2D X-ray projections undergo reconstruction to generate a 3D volume dataset.

In terms of image quality, CBCT provides high- resolution images of bone structures, making it ideal for dental and orthopedic applications. CT excels in providing high-contrast images for both bone and soft tissues; it is used in organ examination and guided procedures. CBCT scanners have a smaller field of view and are typically used to produce detailed images of a specific area, such as the teeth, foot, hand or knee.

See Factor CT3 The See Factor CT3 (Imaginalis, Sesto Fiorentino, Italy) is a multi-modality imaging scanner dedicated to orthopedics, in particular extremities. It can be used in diagnostics, preoperative and intra-operative environments and can acquire CBCT, fluoroscopy (FL) and digital radiography (DR). Equipped with a flat panel detector, the See Factor CT3 utilizes a pulsed x-ray source that projects a cone-shaped beam to illuminate the flat panel. The CBCT mode offers high-resolution 3D imaging and provides detailed anatomical information, especially in hard tissues.

The machine is equipped with a 62.5 cm diameter gantry bore and offers compact design engineering, ideal for integration into operating room workflows. Motorized wheels set the machine in motion with the action of just one operator, hence assuring accessibility to an operating room when needed during a surgical procedure. Further, it has a total width of 90 cm and easily passes through a standard door.

In CBCT mode, the reconstruction volume for each stack is constituted by a cylinder with a diameter of 175 mm in the standard FOV (Field of View) modality and 301 mm in the large FOV; in both modalities, the length of the volume along the Z-axis measures 150 mm. The machine can acquire up to three volume stacks (450 mm). Projections are reconstructed using the Feldkamp, Davis and Kress (FDK) algorithm [37].

Reconstruction algorithm The FDK algorithm is a well-known method for reconstructing 3D images from cone-beam projections. It was introduced by Feldkamp, Davis, and Kress in the 1980s and has since become a standard

method in CBCT imaging. FDK is based on the principles of filtered back projection. It involves two main steps: filtering and back projection. In the FDK algorithm, the acquired cone-beam projections undergo filtering to enhance image quality and reduce artifacts. Various filters, such as ramp or Shepp-Logan filters, can be applied to the projections to remove noise and improve spatial resolution. Filtered projections are then back-projected into a 3D volume space. Back projection is a mathematical process that redistributes the filtered projections back into the volume, reconstructing the original 3D structure.

Cone beam projections are acquired from different angles around the object of interest. Let $p(s, \theta)$ represent the measured cone-beam projections, where s is the two-dimensional position on the detector and θ is the projection angle. The acquired projections undergo filtering to remove noise and artifacts and enhance image quality. The filtering step typically involves applying a Fourier filter in the frequency domain. The filter function $H(w)$ is chosen based on the desired characteristics of the reconstructed image. Filtered projections are back-projected into a 3D volume space to reconstruct the original object. Back projection redistributes the filtered projections back into the volume space.

The back projection operation is mathematically defined as follows:

$$g(x, y, z) = \int_{\theta} \int_{s} p(s, \theta) \delta(x - s\cos(\theta)) \delta(y - s\sin(\theta))\, ds\, d\theta \tag{5.2}$$

where:

- $g(x, y, z)$ is the reconstructed volume.
- $\delta()$ denotes the Dirac delta function.
- (x, y, z) are the coordinates in the reconstructed volume.
- θ is the distance from the X-ray source to the point.

	References	Region of Interest	Techniques
Pixel-based	Van Den Brock et al. (2014) [137]	Tibia	Thresholding
	Anstey et al. (2011) [9]	Femur	Thresholding
	Zhang et al. (2009) [157]	Calcaneous	3D Adaptive Thresholding
	Zhou et al. (2013) [160]	Hip joint	3D Adaptive Thresholding
Region- based	Vasilache et al. (2009) [139]	Pelvic structure	Region Growing
Edge-based	Rathnayaka et al. (2010) [110]	Long bones	Canny Edge
	Pandey et al. (2018) [104]	Calcaneous	Canny Edge
Energy-based	Brehler et al. (2019) [22]	CT foot and ankle	Coupled ASM
	Wu et al. (2013) [149]	Pelvic injuries	Registered ASM
	Boykov et al. (2001) [21]	Femur	Graph-cut
	Liu et al. (2014) [84]	Pelvic	Graph-cut
	Zhang et al. (2010) [157]	Vertebrae	Graph-cut
Deep learning-based	Nikan et al. (2021) [102]	Temporal bones	PWD-3DNet
	Jin et al. (2020) [66], Yao et al. (2021) [150]	Rib fractures	U-net and DenseNet
	Pranata et al. (2019) [109]	Calcaneous fractures	CNN and ResNet
	Folle et al. (2021) [40]	Hand	U-net
	Ding et al. (2019) [34]	Pediatric hand	U-net

Table 5.1: Overview of bone segmentation techniques for different regions of interest [134]

Chapter 6

A novel approach for extremity bone segmentation in cone beam computed tomography

In this chapter, a novel approach for segmenting a single bone in extremities from CBCT images is presented. This task is challenging due to the small and close nature of these bones. The approach proposes an automatic initialization of the graph-cut algorithm based on pixel-based techniques. The segmentation result is post-processed with morphological operations, and then connected components are evaluated to separate the different bones. Finally, the user can select the bone of interest. The approach is compared with a standard graph-cut approach that requires the user's scribbling to segment a single bone, both in terms of accuracy and usability. The automated process simplifies user interaction, allowing easier segmentation of targeted bones. [1]

[1] This chapter has been published as "Extremity bones segmentation in cone beam computed tomography, a novel approach" in *IFMBE procedures* [132].

6.1 Introduction

CBCT an established imaging modality, is discovering new uses in the orthopedic field. The main benefit of CBCT scanners over spiral computed tomography is the capability to obtain 3D images more quickly, accurately, and with less radiation exposure, using smaller imaging devices [13]. That last point has made CBCT an effective imaging technique in emergency departments and surgical rooms, especially for the diagnosis and treatment planning in extremities [63].

Clinical applications of this imaging modality will benefit from the development of more complex image analysis tools. In this context, 3D models of extremity bones are of great interest in orthopedics. Surgeons can use segmented bone models for preoperative planning, helping them visualize anatomical relationships and decide on the best surgical approach for procedures like joint replacements or corrective surgeries In cases of amputation or congenital deformities, single bone segmentation can assist in designing custom prosthetic devices tailored to the specific anatomy of the patient's extremities. For conditions like osteoarthritis or rheumatoid arthritis, detailed bone models can aid in assessing joint damage and progression of the disease. Moreover, bone models of feet can be useful for biomechanical analysis. Accurate extremities bone segmentation from CBCT volumes is crucial for creating these three-dimensional models. A long bone's center section, also known as the shaft or diaphyses, is made up of hard, strongly attenuating, and thick cortical tissue that has higher values on tomographic scans. Segmenting the shaft of a long bone is a reasonably easy task because of the smooth bone borders and strong contrast between the cortical layer and the surrounding low-intensity soft tissue. Global thresholding, followed by fundamental morphological operations, can produce effective outcomes [47]. In contrast, short bones (cube-shaped with roughly equal vertical and horizontal dimensions) are made of spongy bone, covered with a very thin layer of compact bone. The wrist and ankle bones are examples of this kind of hard tissue. As Sebastian et al. [117] listed, there are manifold challenges regarding the segmentation of these kind of hard (shorts bones and epiphyses) tissues from tomography volumes:

Low-contrast and weak bone boundaries: the thickness of the cortical shell is lower than in long bones.

Varying density of cancellous tissues: cancellous bone tissue is vascular and spongy, giving the appearance of textured and inhomogeneous

bone. Thus, bone tissue density characteristics cannot be uniformly described. Hence, simple segmentation methods such as region growing or thresholding unavoidably produce inadequate results [117].

Narrow inter-bone spacing: the articular space between adjacent bones is extremely narrow. The inter-bone zones are diffused and brighter due to the partial volumetric effect, which is a characteristic of the CT modality and makes the bones appear to be in close proximity.

Bone segmentation in the extremities can be recognized as a challenging task. Moreover, extremities such as foot, ankle, hand and wrist, are characterized by a huge number of small, asymmetrical-shaped structures and densities.

Image processing solutions must be designed to facilitate and expedite this work while minimizing manual interaction and inter-operator variability [134]. The application of graph cut for user-friendly single bone segmentation and labeling in CBCT extremities is discussed in this work.

Graph cut is a well-known approach, Boykov et al. [21] applied it to image segmentation for the first time. In the past years, many researchers have applied this method to bone segmentation. Aslan et al. [10] use this approach to segment a single vertebra in CBCT scan. Liu et al. [82] interactively separate bones in CT scans, applying graph cut to segmented images. Pauchard et al. [107] use graph cut to compute a femur finite element model from clinical computed tomography scan for hip fracture prediction.

In a conventional graph cut algorithm, the user makes scribbles on the input image to select the background and object [152]. In this work an automatically seeded algorithm is presented, in which the initialization has been done with a mask obtained through automatic thresholding. This allows to have a larger number of seeds pixels and results in an improved accuracy of segmentation. Moreover, this novel approach reduces user interaction to segment a single bone, as no background scribble from the user is required. Once all the bones were segmented, they were labeled via a connected component algorithm, and then the bones of interest were extracted through a user interaction.

In this chapter, the overall approach is illustrated and some results are reported. Then, these results are compared with the traditional graph-cut approach in the aim of extracting a single bone, in terms of accuracy and usability.

6.2 Methods

The aim of this work is to segment a single bone in CBCT a scan of the extremities, in an accurate and user-friendly manner. To develop and test our method, eight CBCT scans of the extremities were available. In this section, a segmentation algorithm based on graph-cut approach is presented. Graph cut is a segmentation approach that considers both regional and boundary characteristics of the volume, this makes it suitable for segmentation in our target area. In graph-cut segmentation, the volume is represented as a graph [67] [20], a graph G = (V, E) is a generic structure made up of a collection of nodes V that represent the original image's pixels or voxels and a set of arcs (or edges) E that connect the nodes. Two unique terminal nodes, source s and sink t, which stand in for "object" and "background," respectively, in bi-object segmentation are present in addition to nodes set V. As shown in figure 6.1 the graph has two different kinds of edges, represented in yellow and in blue or red. The first form, called n-links, which connects nearby pixels, uses the prefix "n" for "neighbor." The second kind of t-links, where "t" stands for "terminal," joins terminals and pixels.

In graph G, an s/t cut partitions Y into two disjoint subsets S and T, where all object voxels are connected to the object terminal node S and all background voxels are connected to the background terminal node T. The objective is to determine the optimal cut that yields the best segmentation based on given criteria.

Each edge in the graph is assigned a weight, representing a cost associated with cutting that edge. The segmentation is obtained by finding the cut with the minimal total cost, which balances the regional and boundary terms of the energy function. The following energy function, which accounts for both regional and boundary factors, is minimized to determine the best cut:

$$E(A) = \sum_{p \in P} R_p(A_p) + \sum_{(p,q) \in N} B_{p,q}(A_p, A_q) \tag{6.1}$$

A_p is the label of pixel p in an image P. The first term is a regional term and $R_p(A_p)$ is the penalty to assign label A_p to pixel p. The region energy function should reach the smallest value if labels are correctly assigned to all the pixels. The second term is a boundary term $B_{p,q}(A_p, A_q)$, which can be interpreted as a penalty for discontinuity between p and q. $B_{p,q}(A_p, A_q)$ is large when p and q are similar and near to zero when p and q are completely different. In other words, if p and q are similar, then the probability that

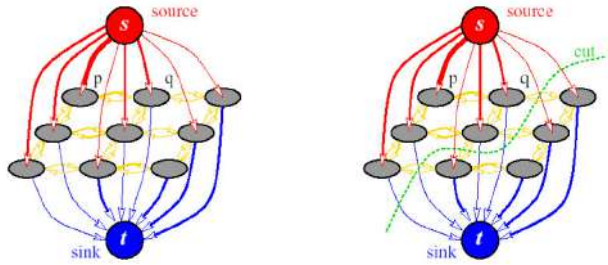


Figure 6.1: Graph cut segmentation algorithm. [132]

they belong to the same object is high. Otherwise, p and q may belong to different objects. Therefore, boundary energy is small if neighboring pixels p and q are different.

N-links (neighbor links) connect (typically) two neighboring nodes and are associated with the boundary terms. In our implementation, their cost is

$$w_{p,q} = f(|I_p - I_q|) \tag{6.2}$$

where I_p and I_q are intensities at pixels p and q and $f(x) = K * exp(-x^2/\sigma^2)$.

T-links (terminal-links) represent regional image properties and are connected to two terminals in the graph. For the smallest cost cut, inexpensive edges are attractive choices. The object and background model determines the weights of the T-links R_p to the object and the background vertex.

$$R_p(obj) = -ln(Pr < I_p|O >)) \tag{6.3}$$

$$R_p(bkg) = -ln(Pr < I_p|B >)) \tag{6.4}$$

where $Pr < I_p|O >$ and $Pr < I_p|B >$ represent likelihoods for object and background and are given by the Gaussian mixture model.

6.2.1 Our idea

These likelihoods for background and object are constructed based on an initialization mask, in which each pixel is marked as a background, object or probable background or object. Accordingly, with the characteristic of

Figure 6.2: Thresholding scheme proposed to initialize graph cut

extremity bones in CBCT, an automatic labeling of volume data is performed. This is achieved using the threshold techniques proposed by Yen and Otsu [151], [103].

The foreground threshold is defined as the value found through the Yen algorithm minus a constant α, set empirically. The background threshold is defined as the value found through the Otsu algorithm plus a constant δ as shown in figure 6.2

All the pixels included in the range between these thresholds are labeled as uncertain pixels to be evaluated from the graph-cut algorithm.

The segmented volume resulting from this algorithm is post-processed with morphological operations. Bones making up extremities are separated via a connected components labeling approach in three dimensions using a 26-connected neighborhood.

Ultimately, a GUI developed in Python allows the user to select the bone of interest.

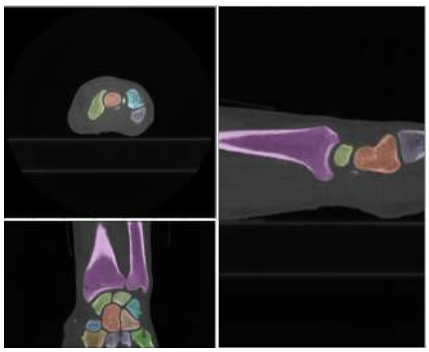
6.3 Results

In this work, a dataset of eight CBCT volumes is evaluated. The dataset consists of four-foot scans and four wrist scans. Acquired with See Factor CT3 the volume are single stacks with dimension 512 X 512 X 512, with a DICOM resolution of 250 μm.

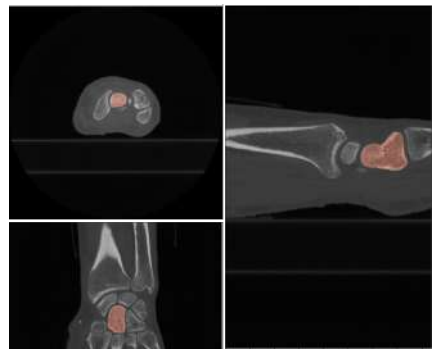
In figures 6.3 and 6.4 the performances of the presented approach on foot and wrist are presented. The algorithm is capable to segment and labels all the bones in these districts, and the user can easily select the bone of interest, quickly switching among adjacent bones.

Table 6.1: Comparison of Dice Coefficient (DC) for bone segmentation using graph cut (GC) with user scribbles vs. thresholding. Results are averaged over the images, means and standard deviation are reported. [132]

Target Bones	GC scribbles (DC ± SD)	GC threshold (DC ± SD)
Trapezium	0.867 ± 0.045	0.896 ± 0.037
Scaphoid	0.890 ± 0.039	0.901 ± 0.030
Lunate	0.823 ± 0.050	0.896 ± 0.025
Navicular	0.882 ± 0.029	0.897 ± 0.019
Talus	0.875 ± 0.042	0.891 ± 0.034
Medial Cuneiform	0.793 ± 0.055	0.820 ± 0.045
Proximal Phalanges	0.784 ± 0.060	0.882 ± 0.028
Navicular	0.896 ± 0.022	0.898 ± 0.021

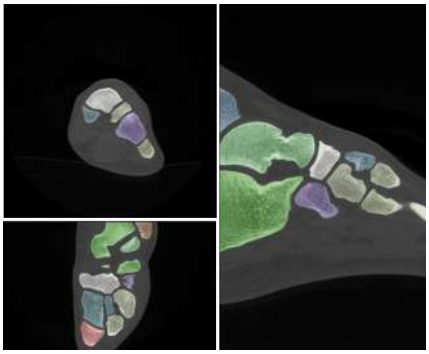


((a)) Graph cut segmentation and connected component labeling

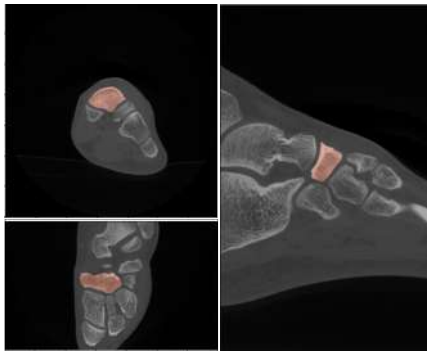


((b)) Capitate bone selection with a click on the corresponding label

Figure 6.3: Wrist bones segmentation obtained with the proposed approach. [132]



((a)) Graph cut segmentation and connected component labeling



((b)) Navicular bone selection with a click on the corresponding label

Figure 6.4: Foot bones segmentation obtained with the proposed approach. [132]

The benchmark for testing our approach performances is the classical Graph Cut approach in which the user makes a scribble to initialize the algorithm and to calculate object and foreground likelihood.

In terms of usability, our approach requires less user interaction and provides more flexibility in segmenting adjacent bones. Because the user selects bones after the entire district is segmented, this allows a faster selection of the desired bone.

In quantitative terms, the metrics of DC are used. A manual segmentation of specific wrist and foot bones is performed using 3D Slicer and then used as the ground truth for metrics evaluation. Table 6.1 reports performance in terms of DC between our approach and classical graph cut, in the segmentation of specific bone targets in wrist and foot. The overall process is performed in 15 minutes for a 3D scan with dimensions and resolution described above.

6.4 Conclusion

CBCT imaging modality is discovering new applications in orthopedics, especially in the extremities. In this work, a novel approach to fast and accurately segment single bone in the extremities is presented. Extremities 3D volume data obtained by the promising imaging techniques of CBCT has

been used. First, the algorithm labels volume according to Yen and Otsu thresholding in order to create a map of pixel used to initialize Graph Cut. Then, the algorithm computes bone segmentation. Connected components are computed to label different bones. Finally, the user selects a single bone of interest (or several ones). A 3D rendering of bone is proposed.

This method was compared with a standard graph-cut approach, where the user must manually label the bone of interest as 'object' and all others as 'background'. Results demonstrate that our approach is more accurate since it's better in terms of the dice coefficient score. Regarding system usability allows less user interaction, resulting more user-friendly.

6.5 Limitation

It's important to acknowledge several limitations of this approach. First, its sensitivity to the type of data volumes being processed is a significant concern, as it may perform well on some CBCT volumes but exhibit poorer performance on others. CBCT scans, unlike conventional CT scans, exhibit variability in pixel intensities and are not calibrated in HU. This lack of standardization can result in similar tissues displaying different pixel values across different scans, potentially affecting the accuracy of the segmentation results. Consequently, the performance of our method may vary depending on the quality and consistency of the input data.

In addition, the complexity of the anatomy within the extremities poses challenges for accurate segmentation. The presence of overlapping structures, variations in bone density, and differences in patient anatomy can complicate the graph cut approach, leading to inaccuracies in bone segmentation, especially in cases with severe deformities or pathology. While our method aims to minimize user interaction, initial thresholding techniques, such as those based on Yen and Otsu, may still require manual adjustments to achieve optimal results.

Another limitation is the computational cost of our method. The graph-cut algorithm, while powerful, can be time-consuming and computationally expensive, particularly when applied to large datasets or high-resolution scans.

Finally, while this study demonstrates promising results, the method's generalizability may be constrained. Machine learning techniques, such as convolutional neural networks, could enhance the generalizability of our re-

sults by enabling the model to learn from a wider variety of data, thus improving segmentation accuracy across diverse anatomical parts.

Chapter 7

Convolutional neural network training strategies in cone-beam computed tomography

Convolutional neural networks (CNN)s have become essential in medical image segmentation. Although CNNs have been very successful in automatic feature extraction in various segmentation tasks, their performance often depends on the availability of large annotated data. Considering CBCT, there is a lack of public dataset, and due to the high volumetric resolution, the computational cost of 3D image processing is high, so finding an efficient training strategy has become critical. This chapter explores three different training methods that have been compared: axial training, Majority Voting (MV) and Multi Planar Training (MPT). In this regard, their performance and computational efficiency in segmenting bones from CBCT images will also be presented. The study identifies MPT as a highly effective and resource-efficient method for enhancing CBCT segmentation while maintaining accuracy.[1]

[1]Part of this work has been published in the article "Deep Learning-Based Workflow for Bone Segmentation and 3D Modeling in CBCT Orthopedic Imaging" in *MDPI Applied Sciences* in 2024 [130].

7.1 Introduction

Convolutional neural networks CNNs have become the gold standard in the field of medical image segmentation. Their ability to automatically learn hierarchical feature representations directly from raw imaging data has made them particularly effective in extracting meaningful patterns from complex, high-dimensional medical images. CNNs have been adopted in various medical imaging tasks, including the segmentation of anatomical structures, organs, and pathological regions in modalities such as MRI, CT, CBCT, and ultrasound. The primary advantage of CNNs lies in their capacity for automatic feature extraction. Architectures such as U-Net and its variants were specifically designed for biomedical segmentation tasks and have become widely used for their effectiveness.

One well-known limitation of CNNs is their need for large data sets. In the context of medical image segmentation, acquiring large volumes of annotated data can be challenging due to the labor-intensive nature of manual annotation by experts. To address this, various techniques such as data augmentation, transfer learning, and synthetic data generation have been employed to mitigate the data scarcity issue. Nevertheless, when faced with limited data, it becomes crucial to explore optimal training strategies to achieve the best neural network performance with the same amount of data.

The first CNNs developed to segment three-dimensional images acquired using MRI or CT were trained on two-dimensional image slices [28]. The majority of these CNNs used axial slices as input [49, 120, 142], due to the high in-plane resolution compared to the slice thickness [80]. CBCT instead offers volumetric high resolution since 3D CBCT images have isotropic voxels, which means that the voxel dimensions are the same in all three spatial dimensions. While axial training simplifies training and reduces computational costs, it fails to capture the full 3D anatomical context.

One method to integrate spatial information between 2D slices without relying on 3D volume-based processing is through a multi-view, multi-learner approach using orthogonal planes like sagittal, axial, and coronal. This approach involves three separate learners, each processing a 2D slice from one of these planes and generating its corresponding segmented slice. The outputs of these learners are then combined using a fusion technique, such as majority voting or learning-based fusion, to produce the final segmentation. This approach is thought to leverage the spatial information of 3D volumes more effectively than purely 2D methods, while being less computationally

intensive than 3D segmentation approaches.

Zhou et al. [161] introduced a method where separate 2D CNNs are trained for each orthogonal plane (axial, sagittal, and coronal), and the final segmentation is determined through a majority voting scheme. This strategy leverages 2D CNNs' simplicity while integrating 3D spatial information more effectively.

Unlike 2D CNNs trained on slices, 3D CNNs process the entire volumetric data in a single pass, allowing the model to capture rich spatial dependencies across all dimensions. For CBCT data, where voxels are isotropic, 3D CNNs are particularly advantageous as they take full advantage of the consistent resolution across all spatial axes. However, 3D CNNs come with increased computational complexity and memory requirements, necessitating careful management of resources during training.

Rather than employing multiple learners to predict segmentation outputs for each 2D plane individually, the multi-view single learner model utilizes a single model that takes all three 2D planes as input. This model learns the input features from the combined planes to generate the corresponding 2D segmentation output.

This work explores a training strategy for CBCT, it consists of training a 2D CNN simultaneously with three different views and is referred to as MPT. This study presents a comprehensive comparative analysis of training strategies for CBCT bone segmentation, using two different neural network architectures: U-Net and SegNet. The objective is to identify the most effective CNN training strategy in terms of segmentation performance and computational efficiency.

7.2 Materials and Methods

7.2.1 Data

Anatomical preparations were scanned using a commercial CBCT device, the See Factor CT3 (Imaginalis, Florence, Italy). The Feldkamp, Davis, and Kress (FDK) algorithm was used to reconstruct the scans [37]. The volumetric data includes $512 \times 512 \times 512$ pixels and has an isotropic resolution of 0.2 mm.

Fifteen CBCT in all, obtained with different acquisition parameters, were considered. These scans focus on orthopedic areas where bone segmentation

poses challenges, such as foot and ankle, hand and wrist, and knees. Each scan was performed on a different anatomical specimen, corresponding to fifteen different patients. The scans were masked using *3D Slicer* ([72]) to obtain ground truth labels. The author thresholded each scan with a value chosen based on the visual result on the image, using the manual threshold functionality in the software, to isolate cancellous tissues. Then the author manually performed segmentation refinements using the *3D Brush* tool in order to complete the mask with trabecular bone and fill holes.

7.2.2 Training Strategies

Three distinct training strategies were assessed and compared.

The first strategy, referred to as *axial* training, is the traditional 2D training method. We employ axial slices of 512×512 pixels to train the developed networks.

The second strategy is the Majority Voting (MV), in which separate convolutional neural networks (CNNs) are trained for each of the three orthogonal slices. The prediction for a voxel is determined by aggregating predictions from all three CNNs, and the voxel is considered part of the foreground if at least two of the CNNs predict it as foreground. This method effectively incorporates spatial information from 3D volumes by leveraging predictions from multiple planes without the computational cost of full 3D processing.

Lastly, an augmented 2D training strategy is evaluated in this work, the (MPT). In this approach, the network is trained with a dataset composed of slices from all three orthogonal planes. In this approach, the network is trained on slices from the axial, sagittal, and coronal planes within each batch, enabling the model to learn features across multiple views concurrently.

7.2.3 Architectures

Two different CNN architectures were developed to compare the effectiveness of the training strategies. Firstly, a U-Net convolutional neural network was developed. This network, implemented as described by Ronnemberg et al. [113] has a U-shaped architecture with a contraction path (downsampling) followed by an expansion path (upsampling). The contraction path captures context and reduces spatial resolution, while the expansion path recovers spatial information. Skip connections, which connect the downsampling and

upsampling paths at corresponding resolutions, enable the network to retain fine-grained details during upsampling.
The second network chosen for comparing training strategies is SegNet [12]. SegNet has an encoder-decoder architecture with a symmetrical contracting and expanding structure. Unlike U-Net, which uses skip connections, SegNet uses pooling indices to perform upsampling. These indices store the locations of max pooling during the downsampling phase and are used to retain fine-grained details during the upsampling. We used the TensorFlow and Keras frameworks in Python to implement the U-Net and SegNet architectures.

The networks under evaluation include six variations: U-Net with axial training, MPT, and MV and SegNet with axial training, MPT and MV training, as shown in fig. 7.1.

A workstation with a GeForce RTX 2070 SUPER (NVIDIA Corporation, Santa Clara, California) was used for the training. We employed an Adam Optimizer, with an initial learning rate of 10^{-3}. The networks were trained for 250 epochs using a batch size of 16. Early stopping and learning rate decay are integrated into the network as callbacks. Moreover, we arranged the dataset according to volumes (each volume represent the scan of a different anatomical preparation) to prevent data leakage, making sure that every slice from a given volume was either in the training set or the test set. In this way, the training and test sets' data independence is preserved, which is helpful for the generalization of the model's performance and an accurate model evaluation. Three volumes were used for testing and ten volumes for training. Moreover, two volumes distinct from the training and test sets were considered to be the validation set.

7.2.4 Evaluation Metrics

To quantitatively assess the performance of our training strategies and facilitate comparison, the predictions of our networks were compared to the ground truth by calculating the following well-established segmentation metrics: the Jaccard index and the Dice coefficient. The Jaccard index evaluates the similarity between two sets by evaluating the ratio of the size of their intersection to the size of their union.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{7.1}$$

The Dice coefficient serves as a metric for quantifying the similarity between two sets. Specifically, it is calculated by dividing the size of the inter-

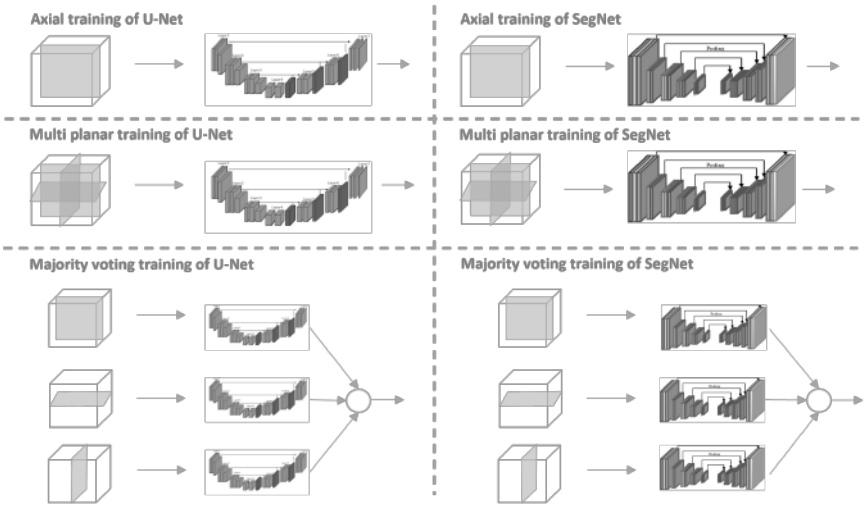


Figure 7.1: Comparison of training strategies: the diagram illustrates axial, multi-planar, and majority voting training methods for both U-Net and SegNet. [130]

section of the sets by the average size of the sets.

$$D(A, B) = \frac{2 \cdot |A \cap B|}{|A| + |B|} \tag{7.2}$$

In the given context, set A comprises pixels labeled as positive (e.g., bone pixels) in the ground truth, while set B encompasses pixels predicted as bone by the CNN being evaluated. Both coefficients have a range from 0 to 1, with 1 being the optimal value.

7.3 Results

To evaluate which training strategies give the best performance, networks trained with different strategies were tested over the same test subset, employing three CBCT volumes. Each network was tested separately using axial, sagittal, and frontal slices. In MV, the results are a combination of the outputs from the three views.

In figure 7.2, segmented images of the human foot are presented, featuring one axial, one sagittal, and one frontal view, alongside binary masks obtained

Table 7.1: Results in terms of Jaccard index and Dice coefficient [130]

		Axial training		Multi-planar training		Majority voting	
		U-Net	SegNet	U-Net	SegNet	U-Net	SegNet
Axial	JI	0.904 ± 0.019	0.917 ± 0.029	0.939 ± 0.010	0.939 ± 0.014	–	–
	DC	0.949 ± 0.011	0.957 ± 0.015	0.969 ± 0.005	0.969 ± 0.008	–	–
Sagittal	JI	0.861 ± 0.014	0.901 ± 0.021	0.936 ± 0.002	0.931 ± 0.020	0.927 ± 0.010[1]	0.933 ± 0.018[1]
	DC	0.899 ± 0.030	0.948 ± 0.011	0.967 ± 0.001	0.964 ± 0.012	0.962 ± 0.004[1]	0.965 ± 0.009[1]
Frontal	JI	0.899 ± 0.006	0.897 ± 0.011	0.933 ± 0.010	0.937 ± 0.013	–	–
	DC	0.946 ± 0.012	0.946 ± 0.006	0.965 ± 0.007	0.968 ± 0.009	–	–

[1]Results obtained through a majority voting scheme which uses axial, sagittal, and frontal views.

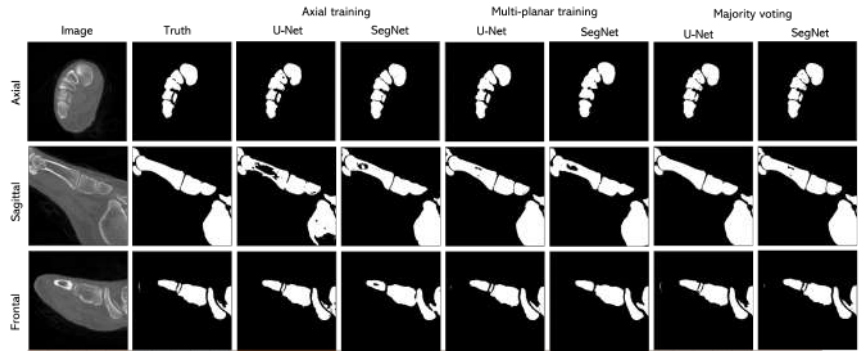
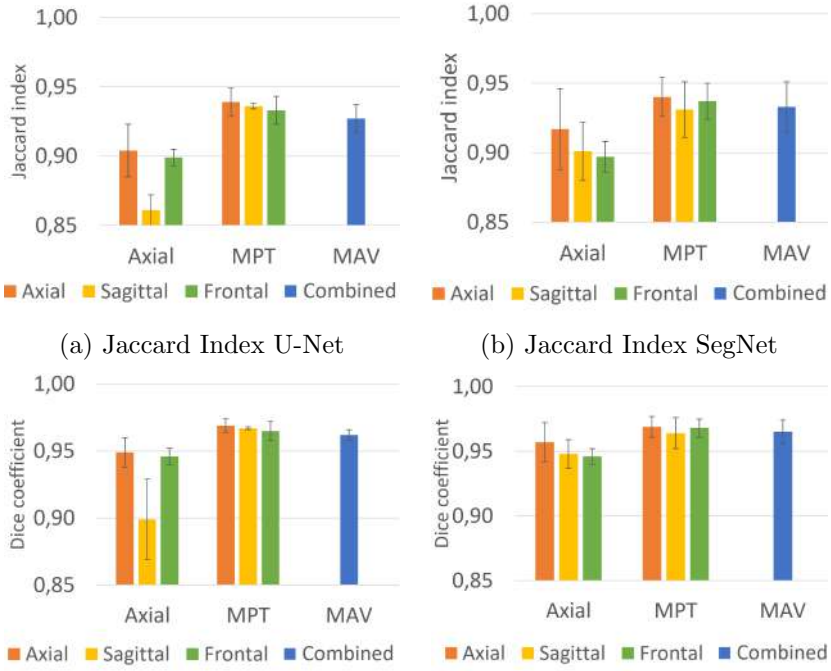


Figure 7.2: Figure shows the qualitative results of the six evaluated networks over the anatomical preparation of a human foot. [130]

from the six networks under evaluation. The proposed networks perform well, with challenges emerging mainly in segmenting frontal and sagittal views using axial-trained networks.

Quantitative evaluation of segmentation performance on experimental CBCT images utilized two metrics: Jaccard index and Dice coefficient. These metrics were computed separately for each volume, and the mean and standard deviation of results across the three test volumes were determined. Table 7.1 displays results in terms of JI and DC. MPT networks exhibited the highest metrics ($JI = 0.939 \pm 0.010$, $DC = 0.969 \pm 0.005$), while *axial* training during sagittal and frontal tests yielded the lowest metrics ($JI = 0.861 \pm 0.014$, $DC = 0.899 \pm 0.030$). In figure 7.3 four charts of the metrics obtained with the two CNNs are presented.

In terms of training and segmentation time, MPT requires more time than *axial* training, while MV requires three times the time requested for *axial* training due to the need to train three distinct networks to perform the voting scheme. *Axial* training took nearly 12 hours, MPT 24 hours, and MV training, involving three networks for the voting scheme, took almost 36 hours. The segmentation of a volume ($512 \times 512 \times 512$ pixels) took about 70 s for the *axial* and MPT and 220 s for the MV training. Although *majority voting* training achieves results as good as MPT, it requires more computational time, as three predictions have to be made to complete the majority voting scheme.



(a) Jaccard Index U-Net

(b) Jaccard Index SegNet

(c) Dice coeff. U-Net

(d) Dice coeff. SegNet

Figure 7.3: Comparison of Dice coefficient and Jaccard index for the two different architectures (U-Net and SegNet) and the three different training strategies under evaluation

7.4 Conclusions

In recent decades, (CNNs) have been widely utilized in medical imaging segmentation tasks. As is commonly acknowledged, a notable challenge when dealing with CNNs is the limited availability of an adequate volume of data. Consequently, it is important to systematically investigate, while holding a constant amount of available data, which training strategies yield superior results in terms of segmentation efficiency and computational costs. In this study, we explore and compare various training strategies along with two distinct neural network architectures for bone segmentation in CBCT images.

The performance of the training strategies evaluated in this study is comparable between the two CNNs , namely U-Net and SegNet. In terms

of Dice coefficient, Jaccard index, and qualitative results, MPT and MV emerge as the more effective training strategies. *Majority voting* results in a good Dice coefficient and Jaccard index for both networks; this finding is in agreement with Minnema et al. [96] they reported that this method has the best performance among those explored in their study. However, our study underlines that MPT, which incorporates all orientations for network training, produces results comparable to MV. Therefore, the MPT is faster to use both during the training and testing phases, requiring the training of a single network and performing predictions using a single network.

This study suggests a training strategy called *multi-planar* training that simultaneously uses all three views of a CBCT scan to train a single network and achieve results comparable to majority voting but with faster processing while maintaining the same volume of data. Our findings highlight that *multi-planar* training is an efficient and effective strategy for enhancing CBCT segmentation.

Chapter 8

Single Bone Modeler

In this chapter, a deep learning-based workflow designed for the segmentation and 3D modeling of bones in CBCT orthopedic imaging is presented, the Single Bone Modeler (SBM). This workflow uses a Convolutional Neural Network (CNN), specifically a U-Net architecture, trained with multi-planar training. The effectiveness of the proposed workflow was evaluated by comparing the generated 3D models against those obtained through other segmentation methods. The accuracy of these models was quantitatively assessed using the Jaccard index, the Dice coefficient, and the Hausdorff distance metrics. The results indicate that the U-Net-based segmentation outperforms other techniques, producing accurate 3D bone models. Eventually, the application of the SBM in veterinary surgery, specifically for the pre-operative planning of acetabular cup insertion is presented. These models, coupled with optical surgical navigation systems, allowed surgeons to visualize and practice the surgical procedure on anatomical replicas, enhancing accuracy in the actual surgery [1]

[1] A preliminary version of the work presented in this chapter has been presented as "Single bone modeler: deep learning bone segmentation for cone-beam ct" in *Proc. of IEEE Int'l Conference of Engineering in Medicine and Biology Society (EMBS), 2024*, [133] and a more complete version published as "Deep learning-based workflow for bone segmentation and 3d modeling in cone-beam ct orthopedic imaginging" *Mdpi Applied Sciense*, 2024 [130]

8.1 Introduction

CBCT is a state-of-the-art 3D medical imaging technology that employs a cone-shaped X-ray beam to produce high-resolution images of anatomical structures. One of the standout attributes of CBCT is its high spatial resolution, enabling the visualization of fine details, particularly in hard tissues. This feature makes CBCT emerge in orthopedics, where it supports implant planning, joint assessment, and the evaluation of traumatic injuries, ultimately enhancing patient outcomes and delivering personalized care [13]. Additionally, compared to traditional CT imaging, CBCT can deliver detailed images at a significantly lower radiation dose, making it a valuable tool in emergency departments and surgical rooms, particularly for diagnosis and pre-surgical planning [63], [45].

Nowadays, CBCT is especially beneficial in maxillofacial and oral surgery. For instance, Bhat et al. [17] proposed a workflow for dental pre-surgical planning using immersive virtual reality and CBCT data. Lalonde et al. [74] utilized 3D-printed models derived from CBCT data to treat a rare type of dens invaginates in a mandibular incisor, highlighting the utility of 3D models in guiding proper treatment. The generation of accurate 3D models relies on precise bone segmentation [138].

Due to the small dimensions of the scanning machine, CBCT technology is particularly feasible, not only for the scanning of the head and maxillofacial district, but also for other extremity anatomical regions such as bone feet, hands and wrist, limbs and joints. However, bone segmentation in these anatomical regions is challenging. Unlike long bones, extremities and joints exhibit weak bone borders, fluctuating densities of cancellous tissue, and small inter-bone spacing. Furthermore, extremities consist of various tiny, asymmetrically formed structures with different densities, such as those in the hand, wrist, and foot. Additionally, unlike conventional CT, CBCT grayscale values are directly associated with X-ray attenuation and lack the standardization provided by HU calibration, introducing nuances and variations in image intensity characteristics that complicate the segmentation process.

The existing literature extensively covers bone segmentation in CBCT, primarily in dental and maxillofacial surgery contexts [136]. Examples of CBCT segmentation for extremities mainly focus on weight-bearing scans of the foot and ankle. However, there is a deficiency of studies related to extremities and joint segmentation from CBCT scans.

U-Net architecture [113] is a widely used CNN architecture for medical imaging segmentation, as previously discussed in this thesis. The U-Net has been applied to conventional CT segmentation of various bony structures [73], [75], [98], [116] [78], [42], [158]. In the context of CBCT data, applications of U-Net have been demonstrated in the maxillofacial region. For instance, Lin et al. [79] used U-Net to accurately segment the mandibular canal in CBCT data, and Zhang et al. employed DBA-Unet for maxillary sinus segmentation [158].

This study proposes a deep-learning-based bone segmentation tool for CBCT orthopedics imaging, the SBM. The aim is to provide an easy-to-use segmentation and 3D modeling workflow for intricate anatomical districts such as ankle, wrist, and joints. This workflow consists of three main steps: bone segmentation, separation, and 3D modeling, designed to be highly intuitive and user-friendly. The main innovation of our work lies in the tool's ability to deliver high-quality results from images of various anatomical district, with minimal user interaction.

8.2 Materials and Methods

This study proposes a deep learning-based workflow using the U-Net architecture to accurately segment extremities and joints in high-resolution CBCT scans: the SBM. As illustrated in fig. 8.1 bones acquired through a commercial CBCT scan undergo segmentation and are separated from the surrounding soft tissues. Binary segmentation is performed using a U-Net. Then bones are separated by the application of the watershed algorithm. Finally, the user through the user interface, can select the bone to be modeled in three dimensions. A simple and intuitive user interface was developed for the SBM. The workflow is optimized to require minimal user intervention, relying primarily on deep learning algorithms that automate most of the process.

8.2.1 Binary Semantic Segmentation

Data Anatomical preparations were scanned using a commercial CBCT device, the See Factor CT3 (Imaginalis, Florence, Italy). The Feldkamp, Davis, and Kress (FDK) algorithm was employed to reconstruct the scans [37]. The volumetric data includes $512 \times 512 \times 512$ pixels and boasts an isotropic resolu-

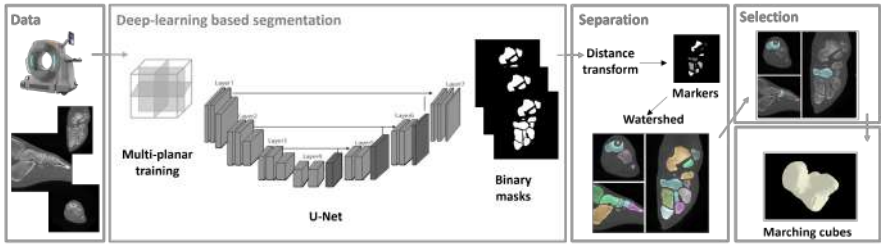


Figure 8.1: Workflow for the Single Bone Modeler: the process begins with acquiring data from a CT scanner, and using a U-Net architecture trained with a multi-planar approach binary masks are generated. These masks undergo separation and selection to isolate and identify individual structures, resulting in detailed segmented models. [130]

tion of 0.2 mm. Scans were performed under varying acquisition parameters (kV and mA).

To train and evaluate the proposed deep-learning models, an in-house annotated dataset of CBCT scans was created. A total of fourteen CBCT scans, acquired with different parameters, were considered. These scans focus on extremities, where bone segmentation is particularly challenging due to the high number of adjacent bones and the similar gray levels between spongy tissues and soft tissues.

To generate ground truth labels, the scans were masked using *3D Slicer* [72]. Initially, we applied a custom threshold to each scan to isolate cancellous tissues. Subsequently, manual segmentation refinements were performed using the *3D Brush* tool to ensure accuracy.

Network architectures The first stage of the proposed workflow involves binary semantic segmentation. A light network architecture with few parameters was implemented to achieve faster inference times, which is critical for ensuring a good user experience in clinical applications. This decision was based on the need for a practical tool that can be easily adopted in routine workflows without requiring extensive computational resources. Consequently, the approach was compared with other methods that share a similar focus on simplicity and efficiency. To achieve this task, a neural network based on the U-Net architecture was implemented, and its performances were compared with another encoder-decoder architecture designed for accurate binary segmentation tasks: SegNet.

The U-Net architecture [113] is a CNN designed for semantic segmentation tasks, where the goal is to assign a label to each pixel in an input image. It has an encoder-decoder path. The encoder path consists of convolutional and pooling layers that progressively downsample the input image. Each convolutional layer is followed by a Rectified Linear Unit (ReLU) activation function. Max pooling operations are applied to reduce the spatial dimensions of the feature maps while increasing the receptive field. The decoder path consists of upsampling and convolutional layers that gradually upsample the feature maps to the original input resolution. Skip connections are introduced between corresponding layers in the encoder and decoder paths to preserve spatial information. These connections directly link layers at the same spatial resolution in the encoder and decoder paths. By concatenating feature maps from the encoder with those in the decoder, skip connections enable the decoder to access high-resolution features from earlier stages of the network. This helps the decoder refine the segmentation masks by incorporating detailed spatial information that may have been lost during downsampling.

SegNet has an encoder-decoder architecture with a symmetrical contracting and expanding structure. Unlike U-Net, which uses skip connections, SegNet uses pooling indices to perform upsampling. These indices store the locations of max pooling during the downsampling phase and are used to retain fine-grained details during the upsampling. The SegNet was implemented as described by Badrinarayanan et al. [12].

Training strategies Based on the result presented in chapter 7, networks are trained with a multi-view single learner model referred to as Multi-Planar Training. In this approach, the network is trained with a dataset composed of slices from all three orthogonal planes (axial, sagittal, and frontal). Each batch, during the training process, contains images from the three views.

A workstation with a GeForce RTX 2070 SUPER (NVIDIA Corporation, Santa Clara, California) was used for the training. We employed an Adam Optimizer, with an initial learning rate of 10^{-3}. The networks were trained for 250 epochs using a batch size of 16. Early stopping and learning rate decay are integrated into the network as callbacks. Before training, a normalization of pixel intensity values to standardize the input images was performed. Data augmentation techniques were applied during training. These techniques included random rotations, flips, shifts, and zooms, as well as the

addition of noise and adjustments to brightness and contrast. Moreover, the dataset was arranged according to volumes to prevent data leakage, making sure that every slice from a given volume was either in the training set or the test set. In this way, the training and test sets' data independence is preserved, which is helpful for the generalization of the model's performance and an accurate model evaluation. Four volumes were used for testing and eight volumes for training. Moreover, two volumes distinct from the training and test sets were considered to be the validation set.

Metrics To quantitatively evaluate the performance of our strategy, we compared network predictions to the ground truth using segmentation metrics described in paragraph 7.2.4: the JI and the DC. We compared our network results with a thresholding method using Otsu threshold and a graph-cut algorithm proposed by Boykov et al. [20], as implemented by Tiribilli et al. [132] and described in section 6.2.

8.2.2 Instance Segmentation

The results of binary segmentation are processed to label each bone separately. This operation is done using an algorithm based on the watershed. The watershed algorithm is a morphological segmentation technique inspired by topographic reliefs. It treats an image as a 3D surface, where pixel intensities represent elevations. The algorithm progressively floods the image from a set of predefined markers, assigning pixels to the nearest basin until all regions are segmented.

Initially, a binary hole-filling operation is applied, effectively closing little void space within the segmented regions. Subsequently, a binary erosion operation is employed to refine the boundaries of the filled mask.

Following the pre-processing stages, a distance transform is computed based on the filled binary mask. The resulting distance map encodes the Euclidean distances from each foreground pixel to the nearest background pixel. Using the computed distance transform, markers for the watershed algorithm are identified through a threshold. The distance transform is noramlized to the maximum value, and all the pixels with a value major of 0.5 are included in the markers. The connected component algorithm is then applied to obtain a labeled array where each region is marked with a unique integer, unmarked pixels (background) are set to zero. Then, regions with dimensions minor than 70 pixels are excluded from the labeled array that represents the markers for the watershed.

The watershed algorithm takes in input the marker array, and the negative of the distance transform that is treated as a topographic surface, where pixel intensity differences define watershed ridges. Starting from the provided markers, the algorithm assigns each unmarked pixel to the closest region by following the path of minimum intensity increase. Boundaries between regions are determined based on intensity gradients, preventing the merging of adjacent basins.

8.2.3 3D Model

To accurately model the bone in 3D in this study, a function implementing the Lewiner algorithm [76] for extracting iso-surfaces from 3D volumetric data was implemented. This algorithm is an enhanced version of the original Marching Cubes algorithm [85], providing faster performance and resolving ambiguities to ensure topologically correct results. Specifically, it uses a refined set of lookup tables to handle all possible configurations of surface intersections within a cube. This approach not only improves the accuracy of surface reconstruction but also ensures the robustness of the generated meshes, making it particularly suitable for complex and high-resolution datasets. The algorithm's implementation allows for efficient processing and accurate depiction of iso-surfaces, contributing to the visualization and analysis of 3D data generated by the segmentation process in our research.

To evaluate results in terms of modeling, a surface-based metric was used to calculate the distance between the obtained mesh and a reference mesh derived from a manual segmentation performed by an expert. The Hausdorff distance is a measure of the extent to which two subsets of a metric space are close to each other. More formally, given two non-empty subsets A and B of a metric space with a distance function d, the Hausdorff distance $d_H(A, B)$ is defined as:

$$d_H(A, B) = \max \left\{ \sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(b, a) \right\} \tag{8.1}$$

Where:

- $\inf_{b \in B} d(a, b)$ measures the shortest distance from a point a in set A to any point in set B.
- $\sup_{a \in A} \inf_{b \in B} d(a, b)$ then considers the farthest distance of these shortest distances for all points in A. This ensures that every point in A is

close to some point in B.

- Similarly, $\sup_{b \in B} \inf_{a \in A} d(b, a)$ ensures that every point in B is close to some point in A.

The maximum of these two quantities is taken to make the distance symmetric and to reflect the greatest extent to which one set can be far from the other.

In brief, the Hausdorff distance $d_H(A, B)$ is the greatest of all the distances from a point in one set to the closest point in the other set, ensuring that both sets are close to each other in a symmetrical sense. Smaller values of the Hausdorff distance indicate better performance.

8.2.4 Graphical User Interface

The GUI of the SBM has been designed with a focus on simplicity and usability, a schematic representation is given in fig.8.2. The GUI presents a clean and organized layout, with essential tools and options easily accessible. The main screen includes a central viewing area for the CBCT images and a sidebar with segmentation tools and options.

The GUI provides real-time feedback, allowing users to see the immediate effects of their actions. Additionally, the GUI guides users through the segmentation and modeling process with tooltips, minimizing the learning curve for new users.

The interface allows the user to load a DICOM folder, and visualize volume in MPR mode. The user can scroll through all the views. Labels are generated fully automatically, and the bone segmentation is superimposed over the MPR views with different colors. By clicking on a specific anatomical part, the user can visualize only the label related to the bone of interest, easily switching between different bones. Once the user selects the bone to be exported, they can save it as a triangular mesh. The GUI is part of the Multimodal Biomedical Imaging Platform All-in-One software [32] developed by Imaginalis S.r.l. (Sesto Fiorentino, Italy), and has been designed following the principles of previous chapters.

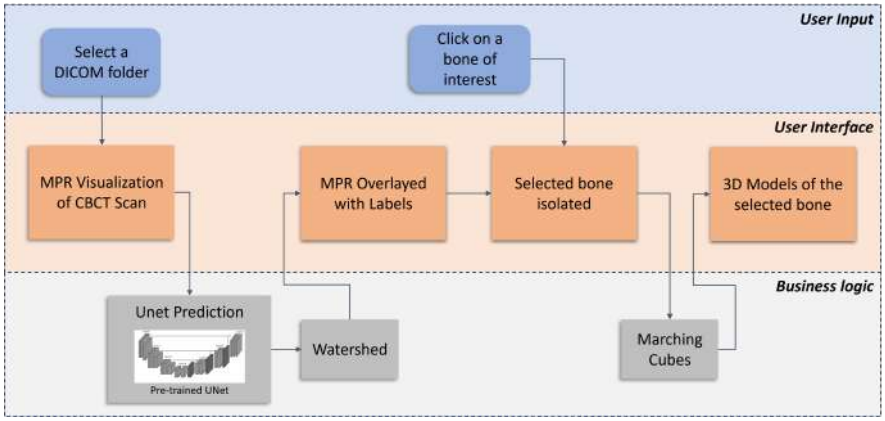


Figure 8.2: This diagram illustrates the workflow of the GUI of the SBM. The user selects a DICOM folder, and the interface allows to visualize the volume in MPR. Users can click on a bone of interest, isolate the selected bone, and generate 3D models. The underlying business logic involves U-Net prediction for initial segmentation, the watershed for labeling, and modified marching cubes for the final 3D model construction.

Table 8.1: Comparisons of performance metrics [133]

	Thresholding	Graph-cut	U-Net MPT [1]	SegNet MPT[1]
JI	0.663 ± 0.089	0.794 ± 0.048	0.941 ± 0.031	0.932 ± 0.018
DC	0.792 ± 0.110	0.885 ± 0.120	0.970 ± 0.015	0.964 ± 0.003

1 $MultiPlanarTraining$

8.3 Results

8.3.1 Binary Segmentation

U-Net, SegNet, conventional Otsu thresholding and graph-cut implemented as described in 6.2 are compared in order to verify the efficacy of the deep learning algorithms in bone segmentation. Metrics are calculated for each of the four volumes under test, and then the mean and standard deviation are computed and reported in table 8.1. The U-Net trained with MPT achieved the best results, with a JI of 0.941 ± 0.031 and a DC of 0.970 ± 0.015.

Qualitative results are shown in fig 8.3. The proposed network, as well as the thresholding and graph-cut approaches, work well for cortical bone. The primary challenges emerge when attempting to segment spongy bone and bone marrow. Due to their lower density, the gray level values of these tissues closely resemble those of soft tissue. As predicted, a thresholding-based approach struggles to segment these tissues effectively. Moreover, this approach is not suitable due to the pixel intensity instability of CBCT. The mentioned problems are partly solved by the graph-cut approach, as it does not rely only on gray-level values. As expected, the best performances are achieved with deep learning approaches. Also, qualitative results confirm the superior performance of the proposed U-Net trained with MPT in handling complex anatomical structures, emphasizing its practical advantages in extremity segmentation.

Results validate the proposed U-Net trained using MPT's superior performance in handling complicated anatomical structures, underscoring its practical utility in extremity binary segmentation. It is worth noting that SegNet also provides good results in terms of evaluated metrics but with a significantly higher number of parameters. This indicates that while both architectures are effective, U-Net, particularly with MPT, offers a more computationally efficient solution without compromising quantitative or quality results.

8.3.2 3D Model

To evaluate the quality of the 3D model derived from the proposed workflow, models were compared using two pixel-based metrics, DC and JI and one distance-based metric the Hausdorff distance. Specifically, the mesh derived from the U-Net with MPT segmentation was compared to those obtained

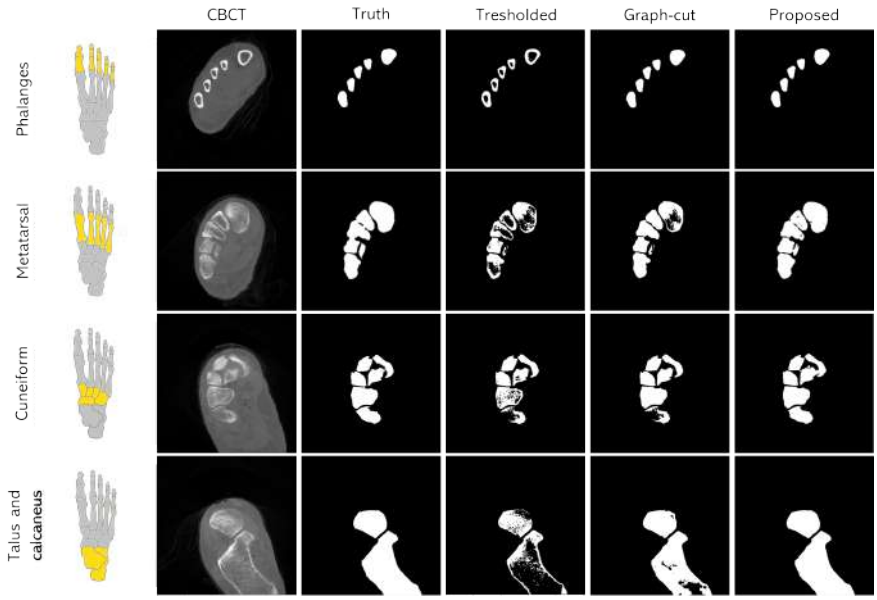


Figure 8.3: Segmentation of bone over an anatomical preparation of a human foot. Comparison between thresholding, graph-cut and the proposed method that achieves best results in terms of JI and DC, U-Net trained with MPT. [133]

	Threshold	Graph Cut	SegNet	U-Net
JI	0.74	0.77	0.83	0.85
DC	0.85	0.87	0.91	0.92

Table 8.2: JI and DC of the 3D model of a human talus, obtained with the proposed workflow [130]

using other methods, including SegNet with MPT, thresholded bone segmentation and with a graph-cut algorithm proposed by Boykov et al. [20] and implemented as described by Tiribilli et al. [132] and described in chapter 6.

Results in terms of JI and DC are presented in table 8.2.

The Hausdorff distance was computed between the mesh produced by each segmentation method and the corresponding ground truth mesh.

The results of this comparative analysis are presented in table 8.3. This table highlights the max, mean and standard deviation of the Hausdorff distances for each segmentation method on the target bone, thereby allowing us to determine which method produces the most accurate and reliable 3D models. By analyzing these distances, the effectiveness of each segmentation technique can be objectively evaluated. The proposed method, U-Net with MPT segmentation, demonstrates the lowest mean in terms of Hausdorff distance. Figure 8.4 depicts a visualization of the Hausdorff distances for the target bone across the evaluated methods. In this context, blue areas on the models indicate minimal differences from the ground truth, while red areas signify substantial discrepancies. The model obtained via threshold-based segmentation shows high Hausdorff distances across the entire surface, indicating poor accuracy. The graph cut segmentation model has high distances in specific regions, reflecting localized segmentation errors. U-Net and SegNet methods exhibit lower Hausdorff distances, with the U-Net model demonstrating the best overall performance, closely aligning with the ground truth. The color bar at the bottom provides a visual reference for the distance values, emphasizing the superiority of the U-Net segmentation approach.

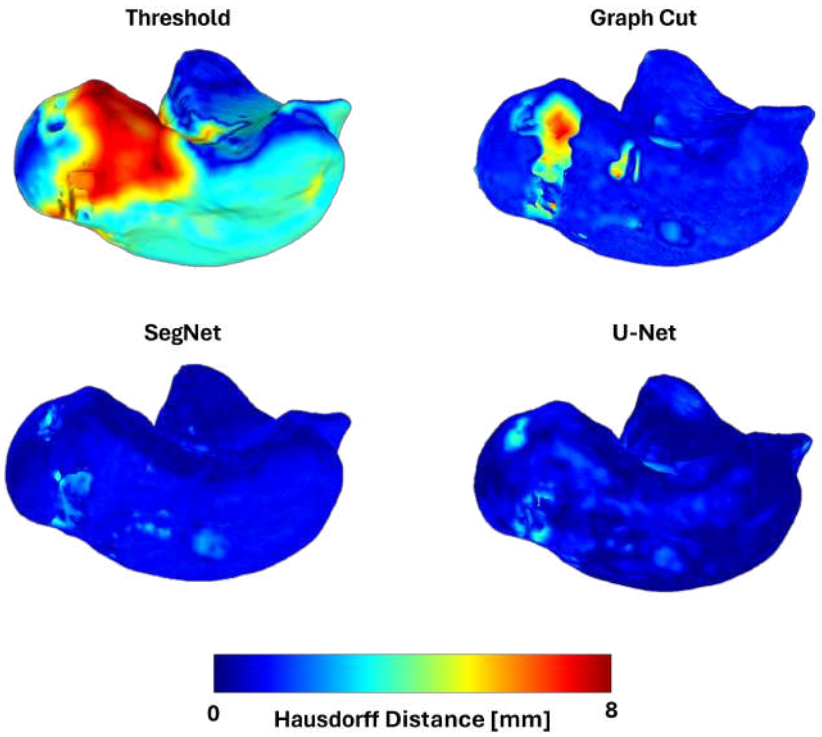


Figure 8.4: Visual representation of the Hausdorff distance. The comparison includes models obtained using different segmentation techniques: threshold segmentation, graph cut segmentation, SegNet segmentation, and U-Net segmentation. [130]

	Threshold	Graph Cut	Seg Net	U Net
Max [mm]	7.9	8.53	2.41	3.81
Mean [mm]	1.48	1.00	0.10	0.067
Std [mm]	2.16	1.96	0.16	0.20

Table 8.3: Max value, mean value and standard deviation of the Hausdorff Distance between reference models under evaluation. The comparison includes models obtained using different segmentation techniques: threshold segmentation, graph cut segmentation, SegNet segmentation, and U-Net segmentation. [130]

8.3.3 Graphical User Interface

In figures 8.5 and 8.6, the user interface is shown, specifically showcasing the segmentation and modeling of a human talus and a human humate. These figures illustrate the effectiveness of the segmentation process and the ability to handle complex anatomical structures. Labels are generated automatically and superimposed on the MPR views. The interface provides a clear and intuitive visualization that aids in the accurate identification and segmentation of these structures. The user can isolate a single bone with a click on the bone of interest and export it as a mesh.

8.4 Discussion

The results of the presented study demonstrate the efficacy of using deep learning techniques, particularly the U-Net architecture, for the segmentation and 3D modeling of bones in CBCT orthopedic imaging.

The performances of the proposed U-Net and a benchmark SegNet architecture for CBCT segmentation of bones were compared. The evaluation on three CBCT volumes revealed that both U-Net and Seg-Net architectures achieve high segmentation accuracy but with distinct differences in computational efficiency and parameter count. U-Net trained with MPT exhibited the highest performance metrics ($JI = 0.939 \pm 0.010$, $DC = 0.969 \pm 0.005$), particularly excelling in handling complex anatomical structures. Although SegNet provided competitive results regarding JI and DC, it required a substantially higher number of parameters than U-Net. This higher parameter count translates to increased computational load and longer training times,

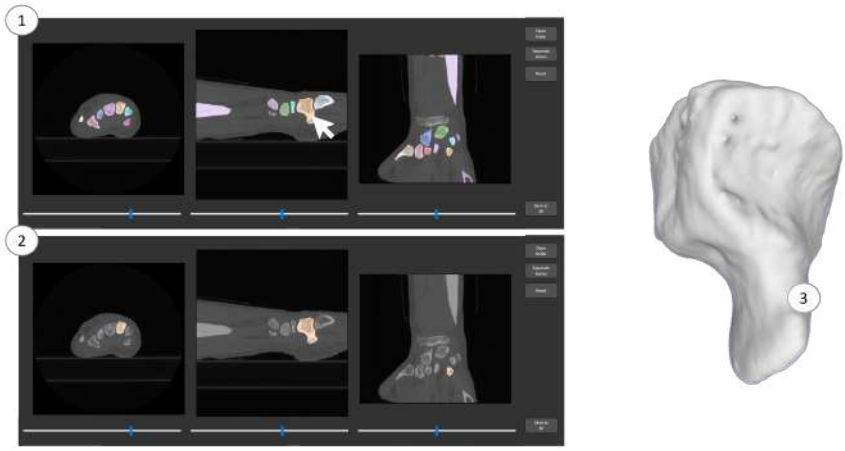


Figure 8.5: Graphical User Interface providing segmentation and modeling of a humate bone of a human wrist [130]

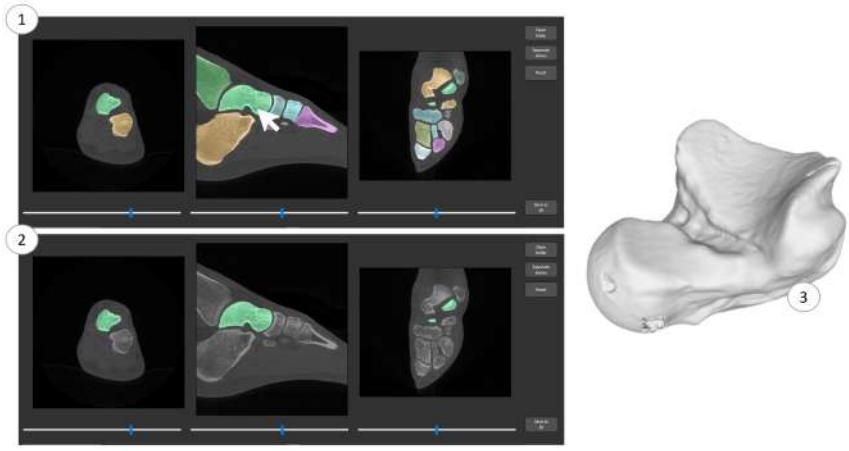


Figure 8.6: Graphical User Interface providing segmentation and modeling of a talus of a human foot [130]

which may be a limiting factor in resource-constrained environments.

The choice to focus on a simple network with fewer parameters was driven by the need to create a tool that balances technical performance with practical usability. While more advanced models may offer incremental improvements in segmentation accuracy, they often come at the cost of increased complexity and longer inference times. This approach demonstrates that a well-designed, simple architecture can provide excellent results with significant advantages in speed and user experience, making it highly suitable for real-world clinical applications.

By comparing the Hausdorff distances JI and DC between the meshes generated by various segmentation methods, it is evident that the U-Net MPT based approach offers superior accuracy. The Hausdorff distance indicates that the U-Net MPT segmentation method yields models with closer alignment to the ground truth, which is critical for applications requiring high precision, such as pre-surgical planning and customized implant design. The workflow's user interface has also proven to be effective in facilitating the segmentation and visualization process. The intuitive design allows users to generate 3D models easily. The ability to accurately segment and model bones such as the talus, wrist, knee, elbow, and shoulder underscores the versatility of the proposed approach.

The method may struggle with highly complex or irregular anatomical structures, such as bones with extensive deformities or fractures. In such cases, the segmentation accuracy could be reduced, potentially requiring manual intervention to correct the segmentation boundaries. The ability to accurately segment and model bones such as the talus, wrist, knee, elbow, and shoulder underscores the versatility of this approach. It allows the generation of an accurate instance segmentation of bones in different anatomical parts without the need to train different neural networks for each specific task.

8.5 Example of a Clinical Application

In this paragraph an example of clinical application of the SBM in veterinary surgery is presented. Specifically, for the pre operative planning of acetabular cup insertion, during total hip replacement in dog. This procedure, aimed at treating severe hip joint conditions such as hip dysplasia and osteoarthritis, is highly complex and traditionally performed blindly, requir-

ing exceptional surgical skill and precision, only a few orthopedic surgeons possess the specialized skills required to perform this complex procedure.

In this procedure, the femoral head (the ball-shaped end of the thigh bone) and part of the neck are carefully excised using surgical instruments. This removes the damaged or diseased components of the hip joint. The acetabulum (socket) of the pelvis is then prepared to receive the acetabular cup component of the total hip replacement. Specialized surgical instruments are used to shape the acetabulum to fit the prosthetic cup securely. The acetabular cup, which is typically made of metal or a combination of metal and plastic, is inserted into the prepared acetabulum. This insertion must be performed with very precise angles. A metal stem with a ball-shaped head (femoral component) is inserted into the femur to replace the excised femoral head and neck. The stem is anchored securely within the femur to provide stability and support. In this kind of surgical procedures, a very careful pre-surgical planning based on the patient's anatomy is required, patient-specific 3D models can be useful to the surgeon. Moreover advanced instrument such as optical navigator, could guide the surgeon during the blind procedure and help to follow the pre-planned trajectory, to be sure to perform the surgical procedure with the required accuracy.

In collaboration with Prof. Filippo Maria Martini, a study was conducted to integrate an optical surgical navigator as an aid to the surgeon during the execution of this complex procedure. In this framework, the segmentation tool presented in this chapter generated patient-specific 3D-printed anatomical models of the femurs and pelvis. That was used to train the surgeon on the use of the optical navigator, performing a pre-planning on patient-specific models.

The dog was scanned with a CBCT machine produced by the company Imaginalis srl. The DICOM dataset was segmented using the SBM, and STL files of the pelvis and the two femurs were exported. The models were 3D printed using a printer with resin technology, as shown in figure 8.7. An optical marker was fixed to the 3D printed model to simulate the surgical procedure assisted by the optical navigator, as shown in figure 8.8(a). Another optical marker was fixed over the surgical instruments to be used in the procedure. The surgeon performs the entire surgical procedure, guided by the navigator system over the 3D-printed models. As shown in figure 8.10, the surgeon can see in real-time movement of the instrument over the 3D model on the MPR views and adjust the insertion angle based on the

Figure 8.7: Patient-specific 3D models of femurs and pelvis obtained through SBM

pre-planned procedure. In figure 8.9 the insertion of the acetabular cup and the metal stem is shown.

The dog was scanned with a CBCT machine produced by the company Imaginalis srl. The DICOM dataset was segmented using the SBM, and STL files of the pelvis and the two femurs were exported. The models were 3D printed using a printer with resin technology, as shown in figure 8.7. An optical marker was fixed to the 3D printed model to simulate the surgical procedure assisted by the optical navigator, as shown in figure 8.8(a). Another optical marker was fixed over the surgical instruments to be used in the procedure. The surgeon performs the entire surgical procedure, guided by the navigator system over the 3D-printed models. As shown in figure 8.10, the surgeon can see in real-time movement of the instrument over the 3D model on the MPR views and adjust the insertion angle based on the pre-planned procedure. In figure 8.9 the insertion of the acetabular cup and the metal stem is shown.

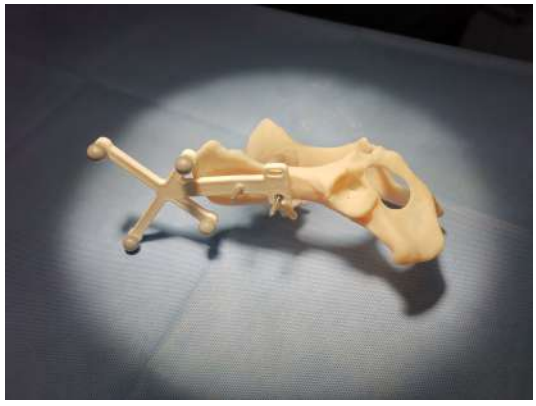
8.6 Conclusion

The integration of accurate segmentation with advanced 3D modeling and a user-friendly interface significantly enhances the practical utility of the entire workflow. The SBM technically improves the bone segmentation process in CBCT images but it is also an advancement in terms of usability. By reducing the need for user intervention, the process is more accessible and

practical for everyday clinical use. This can lead to broader acceptance of the technology in clinical settings. The results from the experiments presented in the previous paragrpah suggest that the SBM tool can be generalized to various anatomical regions and different types of medical imaging, including veterinary. Moreover, demonstrate that the use of these models allow surgeons to visualize and practice the surgical procedure in a detailed and accurate manner, thereby reducing the margin of error during the actual surgery. Additionally, the use of 3D-printed anatomical models in conjunction with optical surgical navigators enhances the surgeon's ability to execute complex procedures such has total hip replacement. This improves the surgical outcomes for patients but also serves as a valuable educational tool for training new surgeons. The ability to replicate patient-specific anatomical structures in fact, enables a hands-on learning experience. In conclusion, the implementation of a streamlined and precise bone segmentation workflow is fundamental to the advancement of surgical techniques and education. It ensures the creation of accurate 3D models that significantly contribute to the success of patient-specific surgical interventions and the development of skilled veterinary surgeons.

In conclusion, the implementation of the SBM workflow ensures the creation of accurate 3D models that could contribute to the success of patient-specific surgical planning. Future work will involve further validation on larger datasets and additional anatomical regions to fully establish the generalizability of this method. Moreover, additional case studies will be investigated to apply the tool for fracture visualization, bone growth assessment, and preoperative planning for other surgical procedures.

((a)) Optical marker fixation



((b)) Patient-specific 3D printed model of the pelvis with marker

Figure 8.8: (a) Optical marker fixation and (b) patient-specific 3D printed model of the pelvis with marker

((a)) Acetabular cup inserted in the 3D printed model of the pelvis

((b)) Metal stem inserted in the 3D printed model

Figure 8.9: 3D models after the surgical simulation

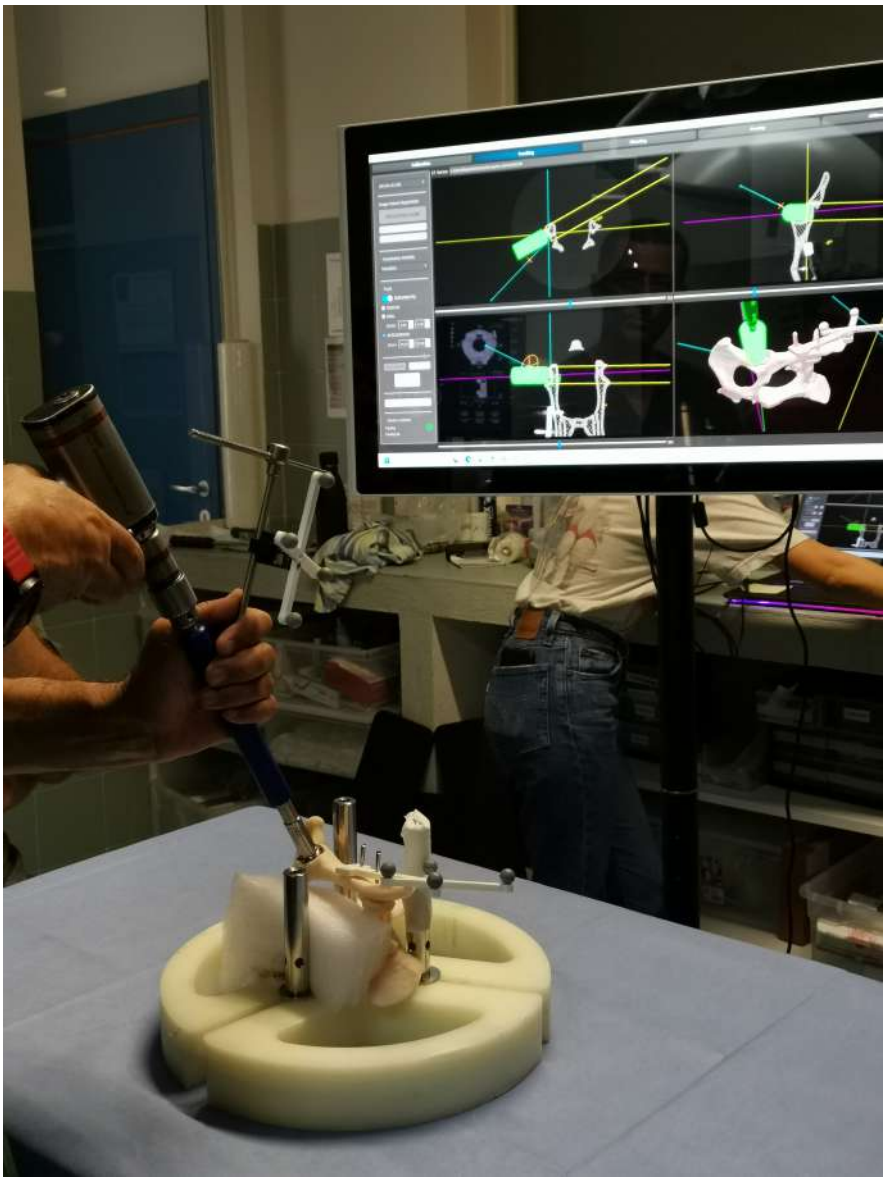


Figure 8.10: Simulation of the optical navigated procedure over the patient-based 3D printed pelvis

Chapter 9

Preoperative planning for fracture reduction

This chapter focuses on the development of a novel workflow for preoperative planning for the reduction of complex bone fractures using Cone Beam Computed Tomography (CBCT) imaging. It introduces a semi-automatic virtual fracture reduction system that uses the Single Bone Modeler (SBM) the segmentation and modeling workflow presented in the previous chapter, and registration techniques to align bone fragments models with a reference template. The proposed method employs ICP algorithm with a new initialization strategy. The subsequent sections detail the workflow's components, including CBCT image acquisition, segmentation, and virtual fracture reduction using 3D mesh processing tools. The results are evaluated through experimental tests involving bovine femur fractures, highlighting the accuracy, efficiency, and usability of the proposed system. [1]

[1]This chapter is part of Dr. Costanza Manfredi's master's thesis in Biomedical Engineering at the University of Pisa, to which the author contributed as a supervisor.

9.1 Introduction

Bone fractures represent the most common form of trauma often requiring hospitalization, imposing a substantial burden on healthcare systems worldwide and, according to a systematic analysis from the Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) 2019, approximately 178 million new fractures occurred globally [97], [148].

Complex fractures require surgical treatment to restore the bone's original anatomy, thereby recovering lost motor function. This procedure is one of the most frequently performed surgeries in orthopedics [97], [153]. Typically, the surgical treatment involves two key steps: first, the bone fragments reduction to restore their correct anatomical positions, and second, stabilizing the bone fragments with fixation devices (as screws and plaques).

Before surgery, the surgeon commonly plans the procedure with the goal of establishing the best possible surgical result and ensuring a functioning limb for the patient [94]. The goal of preoperative planning is to achieve a greater understanding of the fracture, thereby increasing the chance of obtaining an effective reduction and optimising osteosynthesis using the available preoperative data [94].

Relying on 2D radiographs is sometimes challenging and sub-optimal in accurately determining the frature reduction along the three dimensions, and often fail to give surgeons a clear understanding of the fracture pattern, especially in complex cases [97], [154]. As a result, surgical plans may need to be adjusted intra-operatively, requiring implants to be contoured during surgery, which can lead to longer operation times, increased intra-operative bleeding, prolonged exposure to fluoroscopy, and a sub-optimal clinical outcome.

The use of 3D imaging, as CT, avoids the loss of information along the third dimension and allows obtaining a 3D geometrical replica of each fragment, up to the possibility of virtually reducing the entire fracture using proper computer applications. VSP offers several potential benefits:

1. Improved insight into fracture details, including fracture line direction, size, and the number and position of fragments [97], [18].

2. Reduced need for extensive dissection, soft tissue disruption, and repeated adjustments during fixation, potentially leading to shorter surgeries and less blood loss [97], [159].

3. Assistance in choosing the most suitable surgical approach, such as determining the correct screw lengths and plate sizes [97], [154].
4. Automated shaping or pre-bending of fixation hardware and the creation of drilling guides [97], [95].
5. Virtual surgical practice, which helps surgeons shorten their learning curve [97], [145].
6. Enhanced communication between doctors and patients regarding the preoperative plan [97], [144].

The first step in VSP is constructing a 3D patient-specific model that accurately represents the bone and associated fracture fragments [97]. Preoperative planning for fracture surgical treatment involves in general the following stages: 1) creation of patient-specific geometrical models, 2) virtual bone fracture reduction, 3) virtual bone fracture fixation, and 4) analysis of the surgical results to optimize the entire workflow for the interventions in the future.

For surgical planning, the patient-specific geometrical models of the bone fragments are obtained starting from 3D medical images as CT often using state-of-the-art segmentation tool. A step forward is the using of high contrast and high-resolution imaging devices as modern cone beam computed tomography CBCT to improve the fragments real shape acquisition with the possibility to distinguish the cortical parts from those from the trabecular ones.

The second step, consisting in virtual bone fracture reduction, can be considered as a computational puzzle solving in 3D, aims to employ computer algorithms to facilitate the reconstruction of 3D broken objects from the geometry of their fragments, given that the management of the 6 degrees of freedom of each fragment is not trivial. Various approaches have been documented in the literature to address this problem, generally categorized into two main types: (1) boundary matching algorithms, which compare fragment boundaries to find matches [83], and (2) template matching algorithms, which fit fragments into a predefined template used as a reference shape for the broken pieces [83]. These techniques involve matching the bone surfaces of native fragments to an intact template (often obtained from the contralateral limb), proving both effective and computationally feasible [129].

Boundary matching approaches for puzzle solving have their roots in seminal work on algorithms designed to piece together 2D jigsaw puzzles.

However, solving 3D bone fracture "puzzles" presents a significantly greater challenge [83]. Unlike jigsaw puzzles, which involve pieces of similar size and distinctly identifiable shapes, bone fractures involve deformable fragments (bone tissue), potentially indistinct fracture surfaces (such as oblique planes), and may include small, missing, or unusable pieces due to high-energy impacts [83]. Additionally, the available data of the fractures, often derived from segmented CT scans, are characterized by a low spatial resolution

The present work adopts the second approach, focusing on aligning bone fragments into a template of the intact bone. The objective of the present study was to develop clinically effective puzzle-solving methods that would accommodate in vivo complexities and provide accurate reconstructions of highly comminuted articular fractures starting from CBCT images. Because of the high degree of bilateral symmetry in the normal population's lower extremities (less 2%), the intact contralateral limb was used as a template for reconstruction [11].

Automatic fragments reduction with a template is typically achieved using the ICP algorithm [16]. In a notable study by Thaddeus P. Thomas et al. [129], ten tibial plafond fractures were retrospectively analyzed using 3D puzzle-solving methods. Their workflow involved generating geometries of the intact template (controlateral limb) and fragments from pre-operative CT data, segmenting and classifying native surfaces (periosteal and subchondral) based on geometric and bone density information, and reducing the fracture by registering the template to the fracture's base fragment. This process included grossly positioning fragments through initialization steps and fragment-template surfaces registration refinement using the ICP algorithm. However, this initial positioning requires the precise intensity values that characterize bone tissues, making this approach unsuitable for CBCT, which lacks this property.

The objective of this study was to develop new puzzle-solving methods that accommodate in vivo complexities and provide accurate reconstructions also for CBCT acquisitions. We propose a planning workflow that, starting from an accurate bone segmentation, aligns ecah fragment with the contralateral native surfaces using ICP with a new initialization strategy.

The proposed user interface and the effectiveness of the algorithm were tested on a series of bovine bone fracture cases involving four volunteers with medical background.

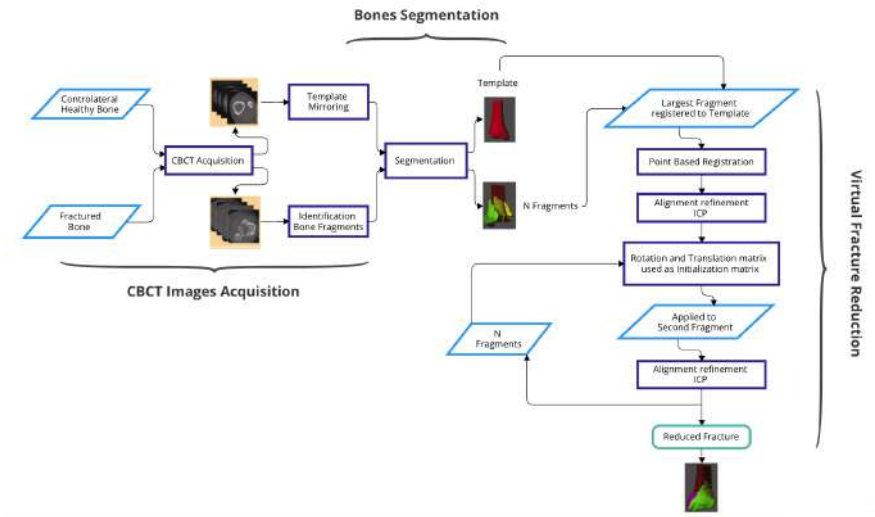


Figure 9.1: A flowchart of the puzzle-solving process is presented.

9.2 Materials and Methods

This section describes the entire workflow (figure9.1) and outlines the new registration strategy for fracture reconstruction:

9.2.1 Workflow

CBCT images acquisition

- CBCT Image acquisition: the workflow starts with the acquisition of a CBCT image of the fractured bone.
- Fragment identification: the number of bone fragments is manually identified. Too small fragments, which cannot be stabilized with the current mechanical approach based on screws and plaques, are not taken into account.
- Reference template acquisition: a CBCT scan of the undamaged bone healthy contralateral side is also required. A mirrored image of this healthy bone will serve as the reference template for reconstruction.

Bones segmentation

- Segmentation, labeling and 3D model reconstruction are required to obtain an accurate geometric model of the bony anatomy from the CBCT data. In this work, both the fractured bone and the healthy bone are segmented using the SBM ([130]) presented in chapter 8.

- The 3D models of the segmented components are exported in STL format. It is important to note that this planning procedure is independent of the segmentation software and can be performed with any software that produces STL-format models of healthy and fractured bone.

Virtual fracture reduction

- The fractured bone fragments are registered to the reference template (mirrored healthy bone) using a fragment-by-fragment approach based on the ICP with a new initialization strategy. The next subsection details our virtual fracture reduction procedure.

9.2.2 Virtual Fracture Reduction

Our virtual fracture reduction is implemented using the functionalities available in the MeshLab tool, a 3D mesh processing software developed by the Visual Computing Lab at ISTI-CNR (Italy) [27]. It is specifically designed for handling large-scale meshes and offers an extensive suite of tools for editing, cleaning, correcting, rendering, and various other manipulations. Our fracture reduction workflow is based on the following steps:

1. The largest fragment is registered to the template. This registration is performed using the "Align Tool" available in MeshLab and it consists in point based registration, asking the user to select 4 corresponding couple of points, followed by an automatic ICP refinement.

2. The total rotation and translation matrix obtained from the largest fragment registration is used as the initialization matrix of the ICP registration of the second fragment (the closer fragment to the largest one).

3. the procedure proceeds with the subsequent closer fragments progressively, applying to the current fragment the previous total registration matrix and optimizing it with the ICP in respect to the template.

.

This procedure requires user's interaction only to initialize the registration of the largest fragment selecting the 4 corresponding couple of points. Then the procedure is completely automatic given that the initialization matrix is taken from the previous fragment registration.

This structured approach ensures accurate and efficient registration of all fragments to the template, leveraging the capabilities of MeshLab for effective fracture reconstruction.

Subsequently, the tool's usability will be tested by individuals external to the project, who will be asked to align a dataset independently using only a brief guide; a questionnaire follows for evaluating and appreciating the experience.

9.2.3 Test Session

Five fracture cases were used to qualitatively evaluate the proposed virtual fracture reduction procedure. Five bovine upper legs are used (weighing approximately in the range of 2.5 - 4.0 kg) of which images were acquired via CBCT before and after femur fracture done with a vacuum press. The number of fragments ranged from 2 to 4. The CBCT acquisitions, before and after the fracture, were performed in a single stack focused on the joint head of interest, with a resolution of 250 μm, using the SeeFactor CT3 by Imaginalis. The image before the fracture has been used in our test session as the template, to avoid the need of the other contralateral leg and to work with a perfect template to objectively measure the final reduction result (often left and right legs are different in geometry).

The segmentation was performed with the SBM ([130], [133]), the automatic segmentation algorithm generated the segmentation masks. The healthly bone was segmented accurately and does not require subsequent refinement. Fragments instead, were subsequently refined through manual optimization with the 3D Slicer open-source software [72].

The effectiveness of our virtual fracture reduction procedure was evaluated by a group of volunteers. Each volunteer was able to use the tool independently, following a brief manual guide and a worm-up phase, without requiring external assistance. They also were trained to employ a manual

Figure 9.2: Five femur fractures were addressed using the puzzle solving method. This figure illustrates a complete example of the virtual fracture reduction workflow, detailing each step in a step-by-step manner to address the entire fracture. The reconstructed (reduced) solutions provide optimal fragment alignment data, facilitating the restoration of the original anatomy. The first column shows the number of fragments and the corresponding segmentation view.

tool for 3D objects management available in MeshLab that allows to change the 6 degrees of freedom using the mouse. The intent of this additional session was to compare the performances of our new ICP-Base reduction procedure with a traditional manual procedures. Each user performed all the reductions using both procedures in a randomized order. We measured the total reduction time and the reduction accuracy of each fragment in terms of HD [146] using the functionality available in MeshLab. Furthermore, at the end of the test, participants were asked to complete a brief structured questionnaire. The questions are listed in the Appendix A.9, with responses on a 1 to 5 points Likert scale.

9.3 Results

There was a range in comminution severity among the five femurs fracture cases that were segmented and reconstructed, figures 9.2, 9.3. The number of fragments ranged from 2 to 4. The automatic fragment initialization algorithm placed the largest fragment close to its anatomic target. Subsequently, the ICP algorithm was able to finalize each fragment's alignment in short time. Larger fragments were quickly aligned instead smaller fragments took more time to align. Much depends on their positioning and the way they are rotated and translated in space relative to the intact model.
The majority of bone fragments had not plastic deformation; but rather fit well to both the intact template (just one fragment of dataset 3), and to adjacent other fragments.

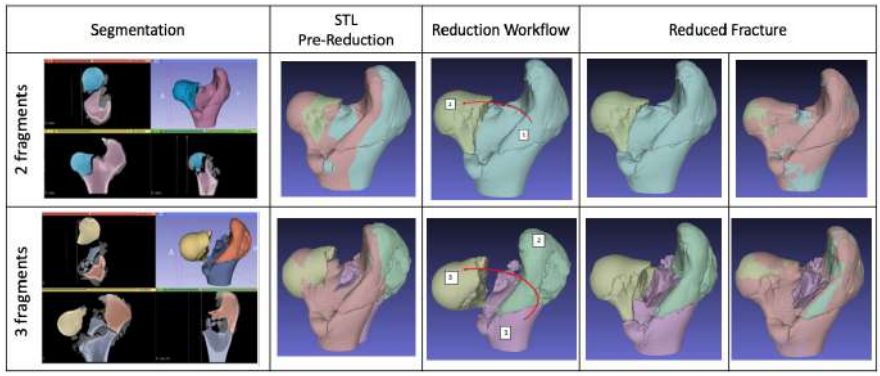


Figure 9.3: Five femur fractures were addressed using the puzzle solving method. This figure presents two examples of the virtual fracture reduction workflow, with cases involving two and three fragments, respectively. Each case includes the pre-reduction state, the chosen workflow, and the post-reduction outcome. The first column displays the number of fragments and the corresponding segmentation view for each case. The reconstructed (reduced) solutions provide optimal fragment alignment data, enabling restoration of the original anatomy.

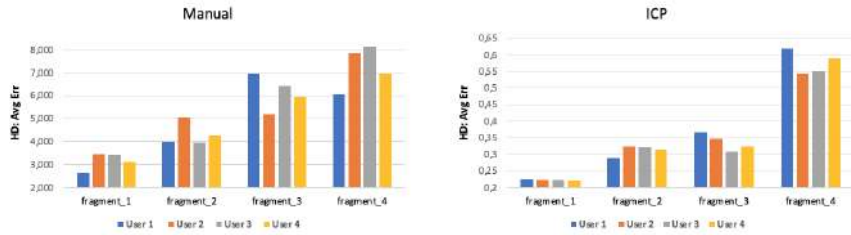


Figure 9.4: Errors made by users when aligning different fragments using the manual method, after the use of the ICP algorithm,

To assess usability, four external users were asked to solve the alignment problem, first using a manual method (mouse aid) and then using the ICP algorithm.
The obtained graphs illustrating their performance are shown below.

An initial comparison was made between the errors made by various users when aligning different fragments using the manual method, followed by using the ICP algorithm, figure 9.4.
Figure 9.5 graphs the mean and standard deviation of the alignment error, measured using the HD metric for each fragment with both methodologies.

At the end, figure9.6, presents the responses to the questionnaire regarding the quality and overall appreciation of the user experience.

9.4 Discussion

Observing the statistic $Mean \pm SD$, the task performed manually appears to be significantly user-dependent and therefore much less precise. The comparison between manual alignment and alignment performed using the ICP algorithm demonstrates the substantial potential of the automated method. The error, as measured by the HD metric, exhibits a difference of an order of magnitude, highlighting the superior accuracy of the automated approach.

It is worth noting that the respective times differ significantly, as the tool executed via the semi-automatic registration workflow takes approximately 1/3 of the time compared to manual manipulation. (To obtain a

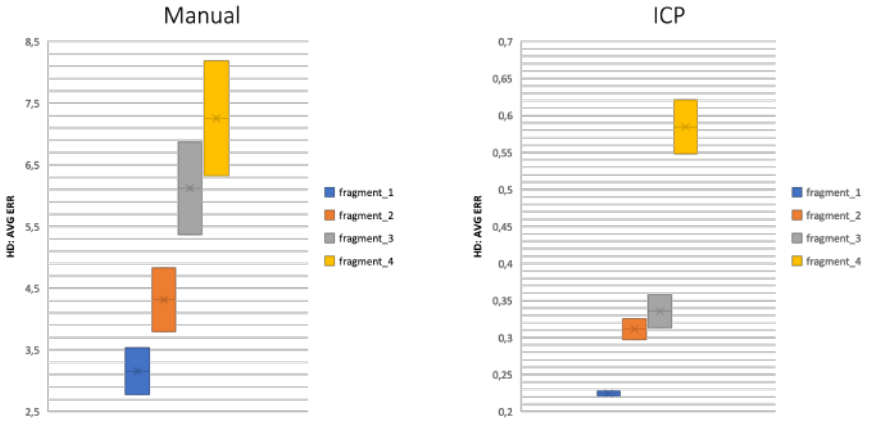


Figure 9.5: Mean ± Standard deviation of the alignment error, measured using the HD metric for each fragment with both methodologies.

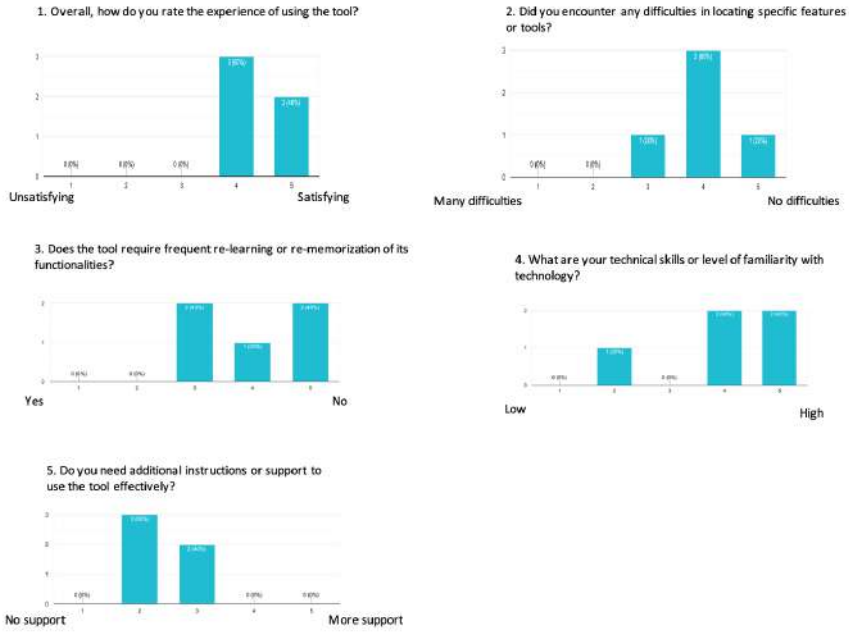


Figure 9.6: Answers to the usability form

comprehensive and definitive statistic, a larger sample would be necessary).

This demonstrates the tool's simplicity and user-friendliness. This step is crucial for assessing the tool's simplicity and user-friendliness in a real-world context, ensuring that even users without extensive knowledge can successfully utilize it.

9.5 Conclusion

The proposed system is capable of virtually reconstructing broken bone fragments for complex bone fracture cases, which is currently an unsolved problem in automatic puzzle-solving algorithms and difficult to achieve using manual methods.

This paper presents several positive points that highlight the significance of the results.

One of the main advantages is the minimal user input required, making the process more intuitive and less prone to human errors. The generation of patient-specific 3D models of healthy bone through SBM results is accurate and usable. Further refinements to accurately separate fragments are requested. Through an initial rough alignment, each fragment is positioned closer to the anatomical target, aiming to improve the speed and stability of the final registration.

The ICP algorithm has proven to be extremely effective in aligning bone fragments, ensuring high precision in solving the puzzle. This has been quantified by calculating distances between the intact model mesh and the meshes of various fracture fragments.

Furthermore, the registration workflow is significantly rapid, optimizing the time required to complete the entire process.

At the same time, there are some negative aspects to consider like the performance of ICP heavily depends on the initial positions of the fragments, which can affect the alignment quality.

Secondly, the current system does not handle interpenetrations between fragments, limiting accuracy in complex situations, such as in less visible areas.

Finally, the time spent on fragment segmentation remains overall quite lengthy, since they need manual refinements. However, this timing can be significantly reduced by improving the SBM to accurately segment various fracture shapes.

Chapter 10

Conclusion

The development of the Biomedical Imaging Platform All-in-One represents an advancement in simplifying imaging visualization and processing in clinical and research settings. This platform has been designed and developed with a focus on orthopedics needs. Extensive usability testing and heuristic evaluations contributed to refining the interface and functionalities, ensuring that the software meets the needs of the user efficiently and effectively. The platform is now near its release, attesting the translation of this industrial research, conducted in Imaginals Srl, Sesto Fiorentino, into a tool ready for real-world application.

Another key outcome is the development of the SBM, an efficient segmentation and modeling tool that facilitates the extraction and visualization of individual bone structures from CBCT data. Despite its high resolution, CBCT introduces artifacts such as beam hardening, shading, and cupping, which introduce noise in grayscale values and complicate accurate bone segmentationâparticularly in anatomically complex regions like joints, hands, feet, and the maxillofacial area. In these cases, poor contrast between bone and soft tissue further hinders segmentation. Additionally, the lack of HU calibration in CBCT images poses further challenges in model creation.

The SBM tool addresses these limitations, demonstrating high accuracy in complex anatomical regions while enhancing segmentation precision and usability. Its integration into the All-in-One platform will further extend the platform's utility, particularly in diagnostic and surgical planning.

10.1 Summary of Contribution

The contributions of this thesis are related to the design of the user-centered Multimodal Biomedical Imaging Platform All-in-One and the study of bone segmentation and modeling workflow, together with the proposal of a workflow for virtual surgical planning for fracture reduction.

Specifically, chapter 2 introduces the design and development of the Multimodal Biomedical Imaging Platform All-in-One [32]. This chapter details the user-centered approach taken to create an intuitive and adaptable platform, addressing complex imaging needs in a streamlined, single interface. The contributions include the initial formulation of platform requirements via collaboration with clinical experts and the proposal of an iterative evaluation framework of its graphical user interface according to usability principles to achieve better accessibility and efficiency in a medical setting. [89].

Chapter 3 describes the first iteration of the heuristic evaluation of the platform, with the application of Zhang's modified heuristics tailored for medical software usability, as presented in [131]. The contribution of this chapter is identifying specific usability issues for the imaging platform and setting up a systematic framework of iterative evaluation that guided the platform's design improvements. This evaluation informed the subsequent design adjustments to ensure that the platform would meet user expectations.

A formal usability testing protocol developed for SaMD is given in chapter 4. This chapter contributes by developing a standardized, reproducible usability testing protocol designed to ensure consistency across the evaluation of different medical software applications. The protocol validated the effectiveness, efficiency, and user satisfaction of the platform, setting a benchmark for usability testing in medical imaging software to support safe and efficient clinical application, results were published in [89], [31].

The following chapters address the issue of bone segmentation of images obtained with the commercial CBCT machine developed by Imaginalis Srl, the See Factor CT3. In particular, chapter 5 provides an overview of segmentation techniques relevant to medical imaging [134]. Such a chapter fills the gap in relating platform requirements for segmentation, discussing conventional methods, and identifying challenges in segmenting bone structures in complex anatomical regions. These insights provided a ground for designing segmentation tools suited for CBCT and introduced the methods used in the later chapters.

Implementing the algorithm proposed by Boykov et al. [20], in chapter 6 we proposed a novel initialization approach for the graph cut algorithm and we applied a framework to isolate single-segmented bone for extremities in CBCT. This method improved segmentation accuracy as shown by Dice coefficient in contrast to the technique proposed by Boykov et al. and improved usability since it doesn't require user scribbles but only user selection of the bone of interest, as presented in [132].

In chapter 7 and 8, we explore the application of CNN to CBCT bone segmentation. Comparing single-plane with multi-planar training methods, chapter 7 provides a resource-efficient approach, multi-planar training, that retains high accuracy with reduced computational demand. These findings supported the CNN-based segmentation strategy adopted in the subsequent chapter. Chapter 8 presents the development of the SBM, a deep learning tool for efficient, single-click bone segmentation and 3D modeling in CBCT. This chapter contributes a novel, user-friendly workflow for detailed bone modeling that minimizes manual effort, enhancing segmentation precision in complex regions. The SBM is validated against traditional methods, showing improvements in accuracy and usability metrics, making it suitable for integration into the All-in-One platform, as published in [130], [133]. Moreover, an example of a clinical application of the SBM is provided; it demonstrates the application of the bone models for planning and training on the use of an optical surgical navigator in a veterinary surgery procedure of total hip replacement.

In the end, chapter 9 outlines a virtual fracture reduction workflow for preoperative planning, integrating segmentation and modeling techniques with registration methods to align bone fragments accurately. This chapter contributes an innovative, semi-automated system for virtual fracture reduction, validated through experimental tests. The work, which emphasizes the applicability of SBM to the healthy bone and the need for further fragment adjustment, suggests improving the ability of SBM to segment fractured bone.

10.2 Directions for Future Work

In the near future, efforts will focus on integrating the SBM into the Biomedical Imaging Platform All-in-One and conducting a clinical evaluation of the integrated system. As the platform approaches its release, extensive clini-

cal testing in diverse healthcare environments will be essential to validate its adaptability and performance across various specialties, including veterinary orthopedics, thanks to prof Dr. Filippo Maria Martini, and in maxillofacial surgery, thanks to Prof. Massimo Robiony, director of the department of Maxillofacial Surgery of the University of Udine, Italy. Such tests will not only confirm the platform's effectiveness but may also suggest additional features designed for more specific applications.

Future developments will include the integration of the optical navigator presented in paragraph 8.5 in the Multimodal Biomedical Imaging Platform All-in-One.

Continued attention to usability, including real-time adaptations of the GUI based on clinician feedback, could further increase user satisfaction and efficiency. Future iterations might also consider customizable interfaces, enabling users to tailor layouts and tools to fit their specific workflow needs, and enhancing adaptability across various clinical settings.

Additionally, future work could refine SBM's deep learning algorithms to better address different fracture types. By integrating models that respond to various fracture shapes, the platform could enhance its predictive capabilities. Moreover, building upon the virtual fracture reduction framework, future research may develop a fully automated system that guides bone fragment alignment without manual intervention, utilizing advanced registration methods and artificial intelligence to predict optimal fracture reduction patterns. This could reduce planning time in complex cases.

Furthermore, future research could explore the materials in 3D printing of models generated by the SBM. This advancement could enhance preoperative planning and surgical training by providing more accurate patient-specific anatomical models. Improving the accuracy of these printed models could involve incorporating segmentation of both spongy and cortical bone, allowing for more precise reproduction of anatomical structures.

Overall, these directions aim to extend the capabilities and applicability of the All-in-One platform and SBM, ensuring they remain cutting-edge solutions that address emerging needs in medical imaging and surgical planning.

Appendix A

Appendix

This appendix is related to the usability test, previously presented in Chapter 4.

A.1 Test Equipment

Table A.1: Test Room Equipment

		HDMI		Wireless	
Tool	**Description**	Mouse	Touch	Mouse	Touch
Pen	For completing the pre- and post-usability test questionnaires.	✓	✓	✓	✓
Internet Connection	Stable.	✓	✓	✓	✓
Desk & Chair	Desk large enough to accommodate the workstation and the peripherals if needed and the questionnaires. Comfortable chair.	✓	✓*	✓	✓*
Monitor Stand	Optionally, can be used for touchscreen support instead of the *Desk & Chair*.		✓*		✓*
Workstation	For carrying out the test; the software to test must be installed. All system notifications must be disabled. For HDMI tests, it must be equipped with an HDMI connection and capturing software. For the wireless test, it must support Google Meet.	✓	✓	✓	✓
Mouse & Keyboard & Monitor	With non-touchscreen display; minimum 1920x1080 resolution support.	✓		✓	
Touchscreen Monitor	With touchscreen display; minimum 3840x2160 resolution support.		✓		✓
Camera	Records the user from the entire scene's perspective with particular attention to the mouse or touchscreen depending on the test modality.	✓	✓	✓	✓
HDMI Cable	Allows the mirroring of the screen between the observer's monitor and the user's workstation. Allows the observer to see how the user is performing the task.	✓	✓		
Smartphone	Equipped with webcam and microphone, positioned in front of the user to frame his face. Used to communicate with the moderator and by the observer to analyse the user's facial expressions. Must support Google Meet.	✓	✓	✓*	✓*

Table A.2: Observation Room Equipment

		HDMI		**Wireless**	
Tool	**Description**	Mouse	Touch	Mouse	Touch
Pen	For the Observer to mark the tasks already executed, and for the Moderator to note comments. Optional.	✓*	✓*	✓*	✓*
Internet Connection	Stable.	✓	✓	✓	✓
Desk & Chairs	Desk large enough to accommodate the personal computers of the moderator and the observer(s) and all the necessary documentation. Comfortable chairs.	✓	✓	✓	✓
HDMI Cable	Allows duplication of the screen between the observer's monitor and the user's workstation without compromising the software performance. Allows the observer to see how the user is performing the task.	✓	✓		
External Monitor	Connected by the *HDMI Cable* with the user's workstation.	✓	✓		
Moderator Computer	Equipped with microphone or headphones to communicate with the user. Must support video calls, in particular using Google Meet.	✓	✓	✓	✓
Observer Computer	Must have the *Stopwatch* software (see *Equipment* subsection) installed to calculate the time required for each task and to take notes of the progress of the test.	✓	✓	✓	✓

A.2 Checklist

CHECKLIST FOR THE TESTING SESSION

❏ Ensure that all test takers are available on the day and time of the test (user, observer, moderator) and do not have any upcoming commitments

❏ Make sure you have at least two comfortable chairs available to place in the observation room, one for the moderator and one for the observer

❏ Check that the lights in both rooms are adequate

❏ Make sure you have printed the pre- and post-usability test questionnaires and the release form

❏ Print the necessary documents, then the task list

❏ Prepare one pen for the user, who has to fill in the questionnaires, and one pen for the moderator, who takes notes of the progress of the test

❏ Prepare a computer with a program for calculating time and noting the comments

❏ Test the WiFi connection

❏ Check that the software starts correctly on the test workstation

❏ Check that the files mentioned in the task list are named in the same way as those present on the test workstation

❏ Delete the files that contain the configuration of the series at the last testing

❏ Make sure that an observer monitor is present

❏ Check if you have an HDMI cable long enough to connect the machine of the test with the observer's monitor (HDMI modality only)

❏ Check that a capturing Flashback Express software is available on the test workstation (HDMI modality only)

❏ Test the connection between the workstation and observer's monitor (HDMI modality only)

❏ Check that the test workstation is connected to the Internet and has a Google Meet application available (Wireless modality only)

❏ Check if the monitor connected to the test workstation is touch and remove mouse/keyboard if any (Touchscreen configuration only)

❏ Check that the monitor attached to the test workstation is not touch and ensure that there is mouse/keyboard (Mouse/Keyboard configuration only)

❏ Prepare a computer available for the moderator, on which a Google Meet video call can be activated

❏ Check that the devices are charged or have a charger and that outlets are easily accessible and in sufficient numbers

❏ Make sure that there is an application on your smartphone that allows you to start a video calls with Google Meet

❑ Make sure the camera is working properly, then place your smartphone so that it frames the user's face (to be checked on the moderator's computer)

❑ Control the audio of the moderator's smartphones and computers

❑ Mute apps that are not useful for testing purposes on the smartphone

❑ Check for a charged tripod camera

❑ Make sure the camera has enough memory to record the entire session

❑ Check that the camera covers the entire test scene and is placed at the right height

RIGHT BEFORE THE TEST STARTS:

❑ Read the test introduction to the user

❑ Have the user sign for the privacy statement

❑ Have the pre-test questionnaire completed

❑ Activate the Google Meet call on the smartphone positioned in front of the user

❑ Activate a Google Meet call on the moderator's computer

❑ Activate the camera

❑ Activate screen recording on the test machine via the Flashback Express (HDMI modality only)

❑ Activate a Google Meet call on the test workstation (Wireless modality only)

❑ Greet the user, asking him to silence his smartphone, and accompany him/her to the observation room

❑ Check that the smartphones of the participants (observer and moderator) are switched off or in a flight mode

❑ Enable the session recording on the moderator's computer

❑ Start administering tasks

IMMEDIATELY AFTER THE TEST:

❑ Stop a session recording on the moderator's computer

❑ Drop the call on the moderator's computer

❑ Move to the test room and interrupt the call on the smartphone

❑ Stop screen recording on the user's computer (HDMI modality only)

❑ Interrupt the call on the test workstation (Wireless modality only)

❑ Stop camera recording

❑ Have the post test questionnaire completed

❑ Ask the user for feedback on the tested software, if they have any suggestions or comments

❑ Thank the user, offering him/her coffee/water and saying goodbye

❑ Download the various recordings to a shared drive with whoever will analyze the tests

❑ Take the camera to download the recording

A.3 Test Introduction

Hi, my name is ..., and I will guide you through this session today. Now, before you begin, I have some information for you. You probably know why we asked you to perform the test but, nevertheless, let me explain it. We are asking people to try using this software to see if it works as intended. This session today should take about 55 minutes. The first thing I want to clarify is that we are testing the software, not you, so you can't do anything wrong, and you should not worry about making mistakes.

I ask you to think out loud as much as possible during the test. For example, to say what you're looking at, what you are trying to do, what you're thinking. This will be of great help for us. We are doing this to improve the software, so we need to hear your honest opinion. If you have any questions during the test, just ask. I may not be able to answer since we are interested in how people behave on their own. We will use some specific terms during the test. If you don't know the meaning of any term, please ask before starting the task. If you need to take a break at any point of the session, just let me know.

With your permission we will record the session including the screen, our conversation, and the entire scene, but these recordings will be used exclusively by us to improve the software. I ask you to sign a simple permission form that authorise us to record the session. Before we start, I ask you to fill a brief questionnaire.

A.4 Pre-test Questionnaire

DATA COLLECTION QUESTIONNAIRE OF USABILITY TEST PARTICIPANT

PARTICIPANT NAME:

AGE:

GENDER:

- ❑ M
- ❑ F

DEPARTMENT:

CURRENT JOB:

WORK EXPERIENCE IN DIAGNOSTIC IMAGING:

- ❑ 0-10 years
- ❑ 11-20 years
- ❑ $>$ 20 years

USE SIMILAR SOFTWARE:

- ❑ Several times a day
- ❑ Once a day
- ❑ Sometimes a month
- ❑ Very rarely

GENERALLY, HOW DO YOU PREFER TO LEARN TO USE A NEW TECHNOLOGY:

- ❑ Reading the technical documentation
- ❑ With a video-training
- ❑ Testing the technology, supported by the help of experts
- ❑ Having access to technology and the ability to test it individually

A.5 Record Agreement

Disclaimer for consent to registration

Thank you for your participation in our usability test for the software developed by Imaginalis.

This document wants to inform you that we will record your session for the benefit of the members of the research group who cannot be present and to review the information during the analysis.

We will benefit enormously from your feedback and appreciate the time you will devote to us. However, if you do not feel comfortable, you can interrupt the test session at any time.

Please read the following statement and sign it. Thank you.

I consent to the audio/video recording that will be performed (date) _____ during the test session. I authorize the usability testing team to use these recordings for the purposes of the test and retain my right to review and inspect the recordings prior to their use.

Date
Full name (readable)
Sign

A.6 Exploratory Tasks

TASK LIST

It is chosen to submit each participant to the test to 55 exploratory tasks in order to test the many basic features of the software and 4 specific tasks. Both exploratory and specific tasks are carried out on a CT Series because it has more features. It is assumed a duration of the test of 45-50 minutes.

EXPLORATORY TASKS

1. Open the CT CAT that is on the desktop.
2. Open settings and explore the appearance options such as layout and overlays. Change them at will.
3. Test the layout for left-handed user, moving the menu bar from right to left. If you are a left-handed user, leave the menu bar on the left, otherwise return to the initial settings.
4. Navigate views: zoom, rotate, translate, change the window level and window width of intensity.
5. On Axial view, switch to 2D mode; again zoom, rotate, translate.
6. Restore only the Axial view geometry.
7. Switch to 3D mode; restore the entire volume geometry.
8. Take a screenshot, then save it as "Screenshot1" on your desktop. Do not anonymize the screenshot.
9. Anonymize the series and hide all objects on the scene, including axes, but not the volume.
10. Look at the DICOM tags, then close the DICOM tags window. Note: DICOM tags are the information about the considered study.
11. Make the objects visible again on the scene, then open the histogram.
12. Change the current intensity preset using the histogram: set the window width at about 2500 and the window level at about 1000.
13. Save the preset as "Preset1". Note: Preset is how the values are set now.
14. Open the filters and bring the "brightness" parameter to 70, then apply it.
15. Open filters, and, again, bring the "brightness" parameter to 70, then apply it.
16. Reopen the filters and revert the application of last filter.
17. Apply the color inversion; restore the entire volume to its initial state.
18. Open Intensity Preset Menu and select CT Cardiac Preset. Note: Intensity Preset is a 2D visualization preset.
19. Shrink and move the histogram, setup the values in range from -200 to 2500 approximately.
20. Save the Preset as "Preset2".

21. Open Transfer Function Menu and select CT Lung Transfer Function. Note: Transfer Function is a function, in 3D, that allows to visualize different tissue in different way, transferring intensity values into color.
22. Change the color of some point at will.
23. Add a new point.
24. Add a new tissue. Note: add a new window. The colored window is a tissue, so you have to add a second one.
25. Remove a point.
26. Save the preset as "Transfer Func1".
27. Open the coronal view in fullscreen.
28. Try measuring a length. Then measure one angle.
29. Test the measurement with the ellipse.
30. Change the color of the ellipse measure, then delete it.
31. Restore the views layout to the starting one.
32. Scroll the slices of the coronal view, then open the measurement taken list and select the angular one.
33. Open the list of measurements taken, hide the length.
34. Hide the measurements only on the 3D view.
35. Delete the length, then delete all the measurements made.
36. Open the "slab thickness" function. Note: Slab Thickness is a function that allows you to see a number of slices instead of one.
37. Change the thickness up to 46mm. Test the "maximum, averaging and minimum projection" functions. Disable the "Slab thickness".
38. Cut the 3D series using the cube.
39. Hide the cube.
40. Experience the "axial" cutting.
41. Choose the ISO surface rendering, then reset the volume to the initial state.
42. Test the "Play" mode by placing frame rate at 20 fps, then stop the play mode.
43. Upload another CT Series, named CT DOG that is on the desktop.
44. Open the "advanced" menu and select "compare".
45. Explore the volumes at will, then lock them and continue the exploration.
46. Reset the geometry of both volumes to the state when locked.
47. Unlock and reset the geometry of both volumes to the initial state.
48. Exit the "compare" mode.
49. Open the "planning" mode.
50. Try to plan an intervention with the "cannula" tool: place it at will.
51. Open the list of plan made, hide the planning, then delete it.

52. Try the "scalpel" tool, save the planning and exit the planning mode.
53. Select the axial view and open the "multislices" mode.
54. Change the layout, then change the space between the slices.
55. Switch to the sagittal view; then exit the "multislices" mode.

A.7 Specific Scenarios

SPECIFIC TASKS

Scenario 1: Imagine that you are a vet and you have received a patient who has suffered an accident, so you have submitted it to a CT.
1. Identify the bone injury after traumatism, making the appropriate reconstructions.
Reference CT: PHANTOM2 Fracture

Scenario 2: Imagine that you are a vet and you have received a patient who had previously undergone a hemilaminectomy surgery that involved the insertion of a screw into a vertebra.
2. Identify the screw, calculate its inclination between the vertebral bodies (respect to the vertical axis) and check its correct positioning.
Reference CT: PHANTOM1 screw

Scenario 3: Imagine that you are a vet and you receive a patient with a dental abscess of which you have a previous CT available. You submit it to another CT to check the progress of caries.
3. Detect the caries and show the owner if there are any differences between the current state of the tooth and that of a previous CT.
Reference CT: PHANTOM 3 CARIES and PHANTOM 3 CARIES 2

Scenario 4: Imagine that you are a vet and you receive a patient who has previously been submitted to a CT.
4. Make 2D views visible and identify the calcification of the spine.
Reference CT: CT spine

A.8 Post-test Questionnaire

USABILITY POST-TEST QUESTIONNAIRE (mouse mode): Make an x on the score box, indicating with 5=fully agree, 1=strongly disagree.

	1	2	3	4	5	Notes
1. The software is intuitive to use						
2. The software could improve my work						
3. I had no problem using the basic features						
4. I did not find any difficult tasks to perform during the test						
5. I think I completed all the tasks correctly during the test						
6. The features of the software are clear and understandable						
7. The icons are intuitive and clear						
8. Error messages are well-placed and I knew immediately that I was wrong						
9. The software gives feedback that allows me to understand if a task has been completed successfully						
10. I think I'd like to work with this software						
11. I think I would not need any initial support for its use						
12. I felt comfortable using this software						
13. I didn't have any trouble using the touchscreen						
14. I think this mode is more convenient than using the mouse						

A.9 Usability Questionnaire fracture reduction

The questions from the usability survey are listed below:

1. Overall, how do you rate the experience of using the tool?
2. Did you encounter any difficulties in locating specific features or tools?
3. Does the tool require frequent re-learning or re-memorization of its functionalities?
4. Do you need additional instructions or support to use the tool effectively?
5. What are your technical skills or level of familiarity with technology?
6. Do you have any additional feedback or comments regarding the tool's usability?

Appendix B

Publications

This research activity has led to several publications in international journals and conferences. These are summarized below.

International Journals

1. **Eleonora Tiribilli** and Leonardo Bocchi. "Deep learning-based workflow for bone segmentation and 3D modeling in cone-beam CT orthopedic imaging", *Applied Sciences*, vol. 14, no. 17, 2024. [DOI:10.3390/app14177557]
2. Elena Denisova, **Eleonora Tiribilli**, Alessio Luschi, Piergiorgio Francia, Leonardo Manetti, Leonardo Bocchi, and Ernesto Iadanza. "Enabling reliable usability assessment and comparative analysis of medical software: a comprehensive framework for multimodal biomedical imaging platforms", *Health and Technology*, pp. 671-682, 2024.[DOI:10.1007/s12553-024-00859-2]
3. Agnese Simoni, Eleonora Barcali, Cosimo Lorenzetto, **Eleonora Tiribilli**, Vieri Rastrelli, Leonardo Manetti, Cosimo Nardi, Ernesto Iadanza, and Leonardo Bocchi. "Innovative tool for automatic detection of arterial stenosis on cone beam computed tomography", *Applied Sciences*, vol. 13, no. 2, pp. 805-805, 2023. [DOI:10.3390/app13020805]

International Conferences and Workshops

1. **Eleonora Tiribilli**, Ernesto Iadanza, and Leonardo Bocchi. "Bones segmentation techniques in computed tomography, a survey", in *IUPESM World Congress on Medical Physics and Biomedical Engineering 2022*, Springer Nature, in press

2. Elena Denisova, **Eleonora Tiribilli**, Leonardo Manetti, Leonardo Bocchi, and Ernesto Iadanza. "Multimodal biomedical imaging platform all-in-one", in *IUPESM World Congress on Medical Physics and Biomedical Engineering 2022*, Springer Nature, in press
3. **Eleonora Tiribilli**, Ernesto Iadanza, and Leonardo Bocchi. "Heuristic evaluation of a medical imaging platform", in *IUPESM World Congress on Medical Physics and Biomedical Engineering 2022*, Springer Nature, in press.
4. **Eleonora Tiribilli** Ernesto Iadanza, and Leonardo Bocchi . "Extremity bones segmentation in cone beam computed tomography, a novel approach", in *IFMBE Springer Proceedings of MEDICON and CMBEIBH Conference*, 2023.
5. **Eleonora Tiribilli** and Leonardo Bocchi. "Single bone modeler: deep learning bone segmentation for cone-beam CT", in *46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE Explore, in press.

National Conferences

1. Alessio Luschi, Laura Gatti, **Eleonora Tiribilli**, Elena Denisova, Leonardo Manetti, Leonardo Bocchi, and Ernesto Iadanza. "Risk management in a medical device software: evaluation of the usability characteristics of a 3D viewer for reporting", in *GNB Proceedings*, pp. 178-181. Patron editore, 2023. FloRe ID: [insert FloRe ID here]

In preparation

1. Costanza Manfredi, **Eleonora Tiribilli**, Elena Denisova, Vincenzo Ferrari "Preoperative planning for fracture reduction using CBCT volumetric imaging",

Tutoring

1. Costanza Manfredi, "Planning preoperatorio per la riduzione di fratture mediante imaging volumetrico", Master Thesis in Biomedical Engineering at the University of Pisa, 2024

Bibliography

[1] *EN ISO 26800:2011. Ergonomics - General approach, principles and concepts*, 2011.

[2] *IEC/ISO 62366-1:2015, Medical Devices-Part1. Application of usability Engineering to medical devices*, 2015.

[3] *EN ISO 9241-11:2018. Ergonomics of human-system interaction - Part 11: Usability: Definitions and concepts*, 2018.

[4] *EN ISO/IEC 25066:2019. Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Common Industry Format (CIF) for Usability - Evaluation Report*, 2019.

[5] A. Anand, S. S. Tripathy, and R. S. Kumar, "An improved edge detection using morphological Laplacian of Gaussian operator," *2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN)*, pp. 532–536, 2015.

[6] F. O. Andrade, L. N. Nascimento, G. A. Wood, and S. J. Calil, "Applying heuristic evaluation on medical devices user manuals," in *World Congress on Medical Physics and Biomedical Engineering, June 7-12, 2015, Toronto, Canada*, D. A. Jaffray, Ed. Cham: Springer International Publishing, 2015, pp. 1515–1518.

[7] M. S. Andreasen, H. V. Nielsen, S. O. Schrøder, and J. Stage, "What happened to remote usability testing? an empirical study of three methods," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '07. Association for Computing Machinery, 2007, pp. 1405–1414.

[8] J. Annett and N. A. Stanton, *Task analysis*. CRC Press, 2000.

[9] J. Anstey, E. Smith, B. Rasquinha, and J. Rudan, "On the use of laser scans to validate reverse engineering of bony anatomy," *Health Technologies*, 2011.

[10] M. Aslan, A. Ali, D. Chen, B. Arnold, A. Farag, and P. Xiang, "3D vertebrae segmentation using graph cuts with shape prior constraints," in *ICIP International Conference on Image Processing*, 09 2010, pp. 2193–2196.

[11] B. M. Auerbach and C. B. Ruff, "Limb bone bilateral asymmetry: Variability and commonality among modern humans," *Journal of Human Evolution*, vol. 50, no. 2, pp. 203–218, 2006.

[12] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[13] J. Bailey, M. Solan, and E. Moore, "Cone-beam computed tomography in orthopedics," *Orthopaedics and Trauma*, vol. 36, no. 4, pp. 194–201, 2022.

[14] L. Ballerini and L. Bocchi, "Multiple genetic snakes for bone segmentation," in *Applications of Evolutionary Computing*, S. Cagnoni, C. G. Johnson, J. J. R. Cardalda, E. Marchiori, D. W. Corne, J.-A. Meyer, J. Gottlieb, M. Middendorf, A. Guillot, G. R. Raidl, and E. Hart, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 346–356.

[15] F. Bambi, I. Spitaleri, G. Verdolini, S. Gianassi, A. Perri, F. Dori, and E. Iadanza, "Analysis and management of the risks related to the collection, processing and distribution of peripheral blood haematopoietic stem cells," *Blood Transfusion*, vol. 7, no. 1, pp. 3–17, 2009.

[16] P. J. Besl and N. D. McKay, "A method for registration of 3D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[17] S. H. Bhat, K. S. Hareesha, A. T. Kamath, A. Kudva, R. Vineetha, and A. Nair, "A Framework to Enhance the Experience of CBCT Data in real-time Using Immersive Virtual Reality: Impacting Dental Pre-Surgical Planning," *IEEE Access*, vol. 12, no. April, pp. 45 442–45 455, 2024.

[18] M. Boudissa, A. Courvoisier, M. Chabanas, and J. Tonetti, "Computer assisted surgery in preoperative planning of acetabular fracture surgery: state of the art," *Expert review of medical devices*, vol. 15, no. 1, pp. 81–89, 2018.

[19] E. Bourennane, P. Gouton, M. Paindavoine, and F. Truchetet, "Generalization of Canny Deriche filter for detection of noisy exponential edge," *Signal Processing*, vol. 82, no. 10, pp. 1317–1328, 2002.

[20] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient N-D image segmentation," *International Journal of Computer Vision*, vol. 70, no. 2, pp. 109–131, 2006.

[21] Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images," in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 1, 2001, pp. 105–112 vol.1.

[22] M. Brehler, A. Islam, L. Vogelsang, and D. Yang, "Coupled active shape models for automated segmentation and landmark localization in high-resolution CT of the foot and ankle," *Proc SPIE Int Soc Opt Eng.*, vol. 176, 2020.

[23] A. J. B. Brush, M. G. Ames, and J. Davis, "A comparison of synchronous remote and local usability studies for an expert interface," in *CHI EA '04: CHI '04 Extended Abstracts on Human Factors in Computing Systems*, 2004, pp. 1179–1182.

[24] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, 1986.

[25] J. Carneiro, E. Muller, R. Almeida, and A. Almeida, "Medical device manuals analysis using heuristic evaluation," 2018.

[26] A. Cassano-Piché, P. Trbovich, M. Griffin, Y. Ling Lin, and T. Easty, *Human Factors for Health Technology Safety: Evaluating and Improving the use of Health Technology in the Real World*, 2012.

[27] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, G. Ranzuglia *et al.*, "MeshLab: an open-source mesh processing tool." in *Eurographics Italian chapter conference*, vol. 2008. Salerno, Italy, 2008, pp. 129–136.

[28] D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Advances in Neural Information Processing Systems*, vol. 4, 2012, Conference paper, p. 2843â2851.

[29] R. Clark, D. Feldon, J. van Merrienboer, K. Yates, and S. Early, "Cognitive task analysis," in *Handbook of Research on Educational Communications and Technology*, 3rd ed. Mahwah, NJ: Lawrence Erlbaum Associates, 2006, pp. 577–593.

[30] T. Cootes and C. Taylor, "Active shape models- 'Smart Snakes'," *BMVC92*, vol. 13, 2009.

[31] E. Denisova, E. Tiribilli, A. Luschi, P. Francia, L. Manetti, L. Bocchi, and E. Iadanza, "Enabling reliable usability assessment and comparative analysis of medical software: a comprehensive framework for multimodal biomedical imaging platforms," *Health and Technology*, vol. 14, no. 4, pp. 671–682, 2024.

[32] E. Denisova, E. Tiribilli, L. Manetti, L. Bocchi, and E. Iadanza, "Multimodal biomedical imaging platform all-in-one," in *IFMBE Proceedings*, in press.

[33] E. Dijkstra, "The end of computing science?" *Communications of the ACM*, vol. 44, no. 6, p. 92, 2001.

[34] L. Ding, K. Zhao, X. Zhang, X. Wang, and J. Zhang, "A Lightweight U-Net Architecture Multi-Scale Convolutional Network for Pediatric Hand Bone Segmentation in X-Ray Image," *IEEE Access*, vol. 7, pp. 68 436–68 445, 2019.

[35] J. F. Dumas and J. Redish, *A Practical Guide to Usability Testing.* Intellect Books, 1999.

[36] L. Faulkner, "Beyond the five-user assumption: Benefits of increased sample sizes in usability testing," *Behavior Research Methods, Instruments, & Computers*, vol. 35, pp. 379–83, 09 2003.

[37] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Am. A*, vol. 1, no. 6, pp. 612–619, Jun 1984.

[38] A. Fernandez, E. Insfran, and S. Abraho, "Usability evaluation methods for the web: A systematic mapping study," *Information and Software Technology*, vol. 27, pp. 1403–1416, 2021.

[39] P. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement," *Journal of Experimental Psychology*, vol. 47, no. 6, pp. 381–391, 1954.

[40] L. Folle, T. Meinderink, D. Simon, A. M. Liphardt, G. Kronke, G. Schett, A. Kleyer, and A. Maier, "Deep learning methods allow fully automated segmentation of metacarpal bones to quantify volumetric bone mineral density," *Scientific Reports*, vol. 11, 2021.

[41] R. Formicola, C. Amici, M. Mor, L. Bissolotti, and A. Borboni, "Design of medical devices with usability in mind: A theoretical proposal and experimental case study using the LEPRE device," *Designs*, vol. 7, no. 1, 2023.

[42] M. Gheisari, F. Ebrahimzadeh, M. Rahimi, M. Moazzamigodarzi, Y. Liu, P. K. Dutta Pramanik, M. A. Heravi, A. Mehbodniya, M. Ghaderzadeh, M. R. Feylizadeh, and S. Kosari, "Deep learning: Applications, architectures, models, tools, and frameworks: A comprehensive survey," *CAAI Transactions on Intelligence Technology*, vol. 8, no. 3, pp. 581–606, 2023.

[43] E. Goceri and N. Goceri, "Deep learning in medical imaging analysis: recent advances and future trends," 2017, pp. 305–310.

[44] J. Gosbee, "Human factors engineering and patient safety," *Quality & safety in health care*, vol. 11, pp. 352–4, 01 2003.

[45] R. Grassi, E. Guerra, and D. Berritto, "Bone fractures difficult to recognize in emergency: May be cone beam computed tomography (CBCT) the solution?" *Radiologia Medica*, vol. 128, no. 1, pp. 1–5, 2023.

[46] E. Hancer and D. Karaboga, "A comprehensive survey of traditional, merge-split and evolutionary approaches proposed for determination of cluster number," *Swarm Evol. Comput.*, vol. 32, pp. 49–67, 2017.

[47] R. Haralick, S. Sternberg, and X.Zhuang, "Image analysis using mathematical morphology. IEEE trans pattern anal mach intell." *IEEE Trans Pattern Anal Mach Intell.*, vol. 36, no. 9, pp. 532–520, 1987.

[48] C. Hass, *A Practical Guide to Usability Testing.* Springer International Publishing, 2019, pp. 107–124.

[49] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Medical image analysis*, vol. 35, pp. 18–31, 2017.

[50] T. Heimann and H. Meinzer, "Statistical shape models for 3D medical image segmentation: a review." *Medical Imaging Analysis*, vol. 13, 2009.

[51] W. Hick, "A simple stimulus generator," *Quarterly Journal of Experimental Psychology*, vol. 3, pp. 94–95, 1951.

[52] S. Hore, S. Chakraborty, S. Chatterjee, N. Dey, A. S. Ashour, C. Le, and D.-N. Le, "An integrated interactive technique for image segmentation using stack based seeded region growing and thresholding," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 6, p. 2773, 2016.

[53] R. Hyman, "Stimulus information as a determinant of reaction time," *Journal of Experimental Psychology*, vol. 45, pp. 188–196, 1953.

[54] E. Iadanza, L. Baroncelli, A. Manetti, F. Dori, R. Miniati, and G. Gentili, "An rfid smart container to perform drugs administration reducing adverse drug events," in *IFMBE Proceedings*, vol. 37, 2011, pp. 679–682.

[55] E. Iadanza, F. Dori, R. Miniati, and E. Corrado, "Electromagnetic interferences (emi) from active rfid on critical care equipment," in *IFMBE Proceedings*, vol. 29, 2010, pp. 991–994.

[56] E. Iadanza, B. Turillazzi, F. Terzaghi, L. Marzi, A. Giuntini, and R. Sebastian, "The Streamer european project. case study: Careggi hospital in florence," in *IFMBE Proceedings*, vol. 45, 2015, pp. 649–652.

[57] E. Iadanza, *Clinical engineering handbook, Second edition*, 2019.

[58] E. Iadanza, R. Fabbri, A. Luschi, P. Melillo, and F. Simonelli, "A collaborative restful cloud-based tool for management of chromatic pupillometry in a clinical trial," *Health and Technology*, vol. 10, pp. 25–38, 2019.

[59] E. Iadanza and A. Luschi, "An integrated custom decision-support computer aided facility management informative system for healthcare facilities and analysis," *Health and Technology*, vol. 10, no. 1, pp. 135–145, 2019.

[60] E. Iadanza, A. Luschi, R. Gusinu, and F. Terzaghi, "Designing a healthcare computer aided facility management system: A new approach," in *IFMBE Proceedings*, vol. 73, 2020, pp. 407–411.

[61] A. Iannessi, P.-Y. Marcy, O. Clatz, A.-S. Bertrand, and M. Sugimoto, "A review of existing and potential computer user interfaces for modern radiology," *Insights into Imaging*, vol. 9, pp. 599–609, 2018.

[62] Interaction Design Foundation, "Gestalt principles," n.d.

[63] T. Jacques, V. Morel, J. Dartus, S. Badr, X. Demondion, and A. Cotten, "Impact of introducing extremity cone-beam CT in an emergency radiology department: A population-based study," *Orthop. Traumatol. Surg. Res.*, vol. 107, no. 2, p. 102834, 2021.

[64] R. Jeffries, J. R. Miller, C. Wharton, and K. M. Uyeda, "User interface evaluation in the real world: a comparison of four techniques," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1991.

[65] B. Jin, Y. G. Ji, K. Choi, and G. Cho, "Development of a usability evaluation framework with quality function deployment: From customer sensibility to product design," *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 19, pp. 177 – 194, 03 2009.

[66] L. Jin, J. Yang, K. Kuang, B. Ni, Y. Gao, Y. Sun, P. Gao, W. Ma, M. Tan, H. Kang, J. Chen, and M. Li, "Deep-learning-assisted detection and segmentation of rib fractures from CT scans: Development and validation of FracNet," *EBioMedicine*, vol. 62, p. 103106, 2020.

[67] M. Jirik and M. Zelezny, "Image segmentation in medical imaging via graph-cuts," *11th International Conference on Pattern Recognition and Image Analysis: New Information Technologies (PRIA-11-2013). Samara, Conference Proceedings*, 2013.

[68] C. Johnoson, J. T.R., and J. Zhang, "A user-centered framework for redesigning health care interfaces." *Journal of Biomedical Imformatics (38)*, pp. 75–87, 2005.

[69] B. Karsh, M. Weinger, P. Abbott, and R. Wears, "Health information technology: Fallacies and sober realities," *Journal of the American Medical Informatics Association*, vol. 17, no. 6, pp. 617–623, 2010.

[70] M. Kass, A. Witkin, and A. Terzopulus, "Snakes: Active contour models," *Int J Computer Vision.*, vol. 1, pp. 321–331, 1988.

[71] D. Kelkar and S. Gupta, "Improved quadtree method for split merge image segmentation," *2008 First International Conference on Emerging Trends in Engineering and Technology*, pp. 44–47, 2008.

[72] R. Kikinis, S. D. Pieper, and K. G. Vosburgh, "3D slicer as an image computing platform for the quantitative imaging network," *Magnetic Resonance Imaging*, vol. 30, no. 9, pp. 1323–1341, 2012.

[73] A. Klein, J. Warszawski, J. HillengaÃ, and K. H. Maier-Hein, "Automatic bone segmentation in whole-body CT images," *IJCARS*, vol. 14, no. 1, p. 21â29, 2019.

[74] L. LaLonde, M. Askar, and S. Paurazas, "A novel diagnostic and treatment approach to an unusual case of dens invaginatus in a mandibular lateral incisor using CBCT and 3D printing technology," *Dentistry Journal*, vol. 12, no. 4, 2024.

[75] N. Lessmann, B. van Ginneken, P. A. de Jong, and I. Isgum, "Iterative fully convolutional neural networks for automatic vertebra segmentation and identification," *Medical Image Analysis*, vol. 53, p. 142â155, 2019.

[76] T. Lewiner, H. Lopes, A. W. Vieira, and G. Tavares, "Efficient implementation of marching cubes' cases with topological guarantees," *Journal of Graphics Tools*, vol. 8, pp. 1–15, 2003.

[77] J. Li, M. Erdt, F. Janoos, T. chiun Chang, and J. Egger, "Medical image segmentation in oral-maxillofacial surgery," 2021.

[78] D. Liao, C. Shi, and L. Wang, "A complementary integrated transformer network for hyperspectral image classification," *CAAI Transactions on Intelligence Technology*, vol. 8, no. 4, pp. 1288–1307, 2023.

[79] X. Lin, W. Xin, J. Huang, Y. Jing, P. Liu, J. Han, and J. Ji, "Accurate mandibular canal segmentation of dental cbct using a two-stage 3D-UNet based segmentation framework," *BMC Oral Health*, vol. 23, no. 1, 2023.

[80] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. V. D. Laak, B. V. Ginneken, and C. I. SÃ¡nchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[81] K. Liu, F.-y. Chan, C. K. Or, D. T.-f. Sun, W.-s. Lai, and H.-y. So, "Heuristic evaluation and simulated use testing of infusion pumps to inform pump selection," *International Journal of Medical Informatics*, vol. 131, p. 103932, 2019.

[82] L. Liu, D. Raber, D. Nopachai, P. Commean, D. Sinacore, F. Prior, R. Pless, and T. Ju, "Interactive separation of segmented bones in CT volumes using graph cut," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008: 11th International Conference, New York, NY, USA, September 6-10, 2008, Proceedings, Part I 11*. Springer, 2008, pp. 296–304.

[83] P. Liu, N. Hewitt, W. Shadid, and A. Willis, "A system for 3D reconstruction of comminuted tibial plafond bone fractures," *Computerized Medical Imaging and Graphics*, vol. 89, p. 101884, 2021.

[84] Y.-S. Liu, Y. Xia, J. Wu, and J.-H. Lai, "Pelvic bone segmentation using graph cuts and shape constraints," *Pattern Recognition Letters*, vol. 47, pp. 69–76, 2014.

[85] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," *SIGGRAPH Comput. Graph.*, vol. 21, no. 4, pp. 163–169, 1987.

[86] A. Luschi, A. Belardinelli, L. Marzi, F. Frosini, R. Miniati, and E. Iadanza, "Careggi smart hospital: A mobile app for patients, citizens and healthcare staff," in *2014 IEEE-EMBS International Conference on Biomedical and Health Informatics, BHI 2014*, 2014, pp. 125–128.

[87] A. Luschi, M. Monti, and E. Iadanza, "Assisted reproductive technology center design with quality function deployment approach," in *IFMBE Proceedings*, vol. 51, 2015, pp. 1587–1590.

[88] A. Luschi, L. Caltagirone, C. Mondovecchio, R. Miniati, and E. Iadanza, "Assessing the impact of a cis/pacs technology for a cardiology department using qfd methodology," in *IFMBE Proceedings*, vol. 57, 2016, pp. 965–968.

[89] A. Luschi, L. Gatti, E. Tiribilli, E. Denisova, L. Manetti, L. Bocchi, and E. Iadanza, "Design and validation of a usability test for a 3D biomedical image viewer," in *Proceedings of the Eighth National Congress of Bioengineering*. Padova, Italy: GNB Proceedings, June 21 2023.

[90] V. S. Machado Paixao-Cortes, M. Dos Santos da Silva Tanus, W. R. Paixao-Cortes, O. N. De Souza, M. De Borba Campos, and M. S. Silveira, "Usability as the key factor to the design of a web server for the CReF protein structure predictor: The wcref," *Information*, vol. 9, no. 1, 2018.

[91] M. Mandolini, A. Brunzini, G. Facco, A. Mazzoli, A. Forcellese, and A. Gigante, "Comparison of three 3D segmentation software tools for hip surgical planning," *Sensors*, vol. 22, no. 14, pp. 1–16, 2022.

[92] D. Markonis, M. Holzer, F. Baroz, R. L. R. De Castaneda, C. Boyer, G. Langs, and H. Muller, "User-oriented evaluation of a medical image retrieval system for radiologists," *International Journal of Medical Informatics*, vol. 84, no. 10, pp. 774–783, 2015.

[93] M. Matera, M. F. Costabile, F. Garzotto, and P. Paolini, "SUE inspection: An effective method for systematic usability evaluation of hypermedia," *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans.*, vol. 32, no. 1, pp. 93–103, 2002.

[94] C. Mensel, P. H. Gundtoft, and O. Brink, "Preoperative templating in orthopaedic fracture surgery: the past, present and future," *Injury*, vol. 53, pp. S42–S46, 2022.

[95] B. Merema, J. Kraeima, K. Ten Duis, K. Wendt, R. Warta, E. Vos, R. Schepers, M. Witjes, and F. IJpma, "The design, production and clinical application of 3D patient-specific implants with drilling guides for acetabular surgery," *Injury*, vol. 48, no. 11, pp. 2540–2547, 2017.

[96] J. Minnema, J. Wolff, J. Koivisto, F. Lucka, K. J. Batenburg, T. Forouzanfar, and M. van Eijnatten, "Comparison of convolutional neural network training strategies for cone-beam CT image segmentation," *Computer Methods and Programs in Biomedicine*, vol. 207, 2021.

[97] J. Z. Moolenaar, N. Tümer, and S. Checa, "Computer-assisted preoperative planning of bone fracture fixation surgery: A state-of-the-art review," *Frontiers in Bioengineering and Biotechnology*, vol. 10, p. 1037048, 2022.

[98] D. Morita, S. Mazen, S. Tsujiko, Y. Otake, Y. Sato, and T. Numajiri, "Deep-learning-based automatic facial bone segmentation using a two-dimensional u-net," *Int J Oral Maxillofac Surg*, vol. 52, no. 7, pp. 787–792, Oct 2022.

[99] J. Nielsen, "Usability 101: Introduction to usability," https://www.nngroup.com/articles/usability-101-introduction-to-usability/, n.d.

[100] J. Nielsen and R. Mack, "Heuristic evaluation," *Usability Inspection Methods*, 1994.

[101] J. Nielsen and R. Molich, "Heuristic evaluation of user interfaces." *SIGCHI Conference on Human Factors in Computing Systems, Seattle, WA*, 1990.

[102] S. Nikan, K. V. Osch, M. Bartling, D. G. Allen, S. A. Rohani, B. Connors, S. K. Agrawal, and H. M. Ladak, "Pwd-3Dnet: A deep learning-based fully-automated segmentation of multiple structures on temporal bone CT scans," *IEEE Transactions on Image Processing*, vol. 30, pp. 739–753, 2021.

[103] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[104] A. Pandey and S. K. Shrivastava, "A survey paper on calcaneus bone tumor detection using different improved Canny Edge Detector," in *2018 IEEE International Conference on System, Computation, Automation and Networking (ICSCA)*, 2018, pp. 1–5.

[105] V. Patel and D. Kaufman, "Cognitive informatics," in *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*, 5th ed., E. Shortliffe and J. Cimino, Eds., 2021, pp. 115–136.

[106] V. Patel, D. Kaufman, and T. Kannampallil, "Human-computer interaction, usability, and workflow," in *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*, 5th ed., E. Shortliffe and J. Cimino, Eds.

[107] Y. Pauchard, T. Fitze, D. Browarnik, A. Eskandari, I. Pauchard, W. Enns-Bray, H. Palsson, S. Sigurdsson, S. J. Ferguson, T. B. Harris, V. Gudnason, and B. Helgason, "Interactive graph-cut segmentation for fast creation of finite element models from clinical CT data for hip fracture prediction," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 19, no. 16, pp. 1693–1703, 2016.

[108] L. Pecchia, N. Pallikarakis, R. Magjarevic, and E. Iadanza, "Health technology assessment and biomedical engineering: Global trends, gaps and opportunities," *Medical Engineering & Physics*, vol. 72, pp. 19–26, 2019.

[109] Y. D. Pranata, K. C. Wang, J. C. Wang, I. Idram, J. Y. Lai, J. W. Liu, and I. H. Hsieh, "Deep learning and SURF for automated classification and detection of calcaneus fractures in CT images," *Computer Methods and Programs in Biomedicine*, vol. 171, pp. 27–37, 2019.

[110] K. Rathnayaka, T. Sahama, M. A. Schuetz, and B. Schmutz, "Effects of CT image segmentation methods on the accuracy of long bone 3D reconstructions," pp. 226–233, 2011.

[111] J. Rieman, M. Franzke, and D. Redmiles, "Usability evaluation with the cognitive walkthrough," in *Conference companion on Human factors in computing systems*, 1995, pp. 387–388.

[112] D. Ripalda, C. Guevara, and A. Garrido, "Gestalt prototyping framework-evaluation tool," in *Intelligent Human Systems Integration 2021, IHSI 2021*, D. Russo, T. Ahram, W. Karwowski, G. D. Bucchianico, and R. Taiar, Eds. Springer, 2021.

[113] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, Cham, 2015, pp. 234–241.

[114] D. L. Rubin, H. Greenspan, and A. Hoogi, "Biomedical imaging informatics," in *Biomedical Informatics: Computer Applications in Health Care and Biomedicine, 5th Edition*, E. H. Shortliffe and J. J. Cimino, Eds. Springer, 2021.

[115] A. L. Russ and J. J. Saleem, "Ten factors to consider when developing usability scenarios and tasks for health information technology," *Journal of Biomedical Informatics*, vol. 78, pp. 123–133, 2018.

[116] E. Schnider, J. Wolleb, A. Huck, M. Toranelli, G. Rauter, M. Muller-Gerbl, and P. C. Cattin, "Improved distinct bone segmentation in upper-body ct through multi-resolution networks," *International Journal of Computer Assisted Radiology and Surgery*, vol. 18, no. 11, pp. 2091–2099, 2023.

[117] T. Sebastian, H. Tek, J. Crisco, and B. Kimia., "Segmentation of carpal bones from ct images using skeletally coupled deformable models." *Medical Image Analysis*, vol. 7, no. 9, pp. 21–24, 2003.

[118] B. Shackel, "Usability â Context, framework, definition, design and evaluation," *Interacting with Computers*, vol. 21, no. 5-6, pp. 339–346, 12 2009.

[119] A. Sharma, K. Wang, and E. Siegel, "Radiologist digital workspace use and preference: A survey-based study," *Journal of Digital Imaging*, vol. 30, no. 6, pp. 687–694, Dec 2017.

[120] H. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[121] J. Shin and H. Lee, "Optimal usability test procedure generation for medical devices," *Healthcare*, vol. 11, no. 3, 2023.

[122] M. Shin, D. Goldgof, K. Bowyer, and S. Nikiforou, "Comparison of edge detection algorithms using a structure from motion task," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 31, pp. 589–601, 09 2001.

[123] E. H. Shortliffe, J. J. Cimino, and M. F. Chiang, *Biomedical Informatics*, 2021.

[124] W. Sieck, "What is cognitive task analysis?" 2020, accessed: 2024-09-02.

[125] M. P. Starmans, S. R. van der Voort, J. M. C. Tovar, J. F. Veenland, S. Klein, and W. J. Niessen, "Chapter 18-radiomics: Data mining using quantitative medical image features," in *Handbook of Medical Image Computing and Computer Assisted Intervention*, S. K. Zhou, D. Rueckert, and G. Fichtinger, Eds. Academic Press, 2020, pp. 429–456.

[126] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool," *BMC Medical Imaging*, vol. 15, p. 29, 2015.

[127] Z. Tang, T. R. Johnson, R. D. Tindall, and J. Zhang, "Applying heuristic evaluation to improve the usability of a telemedicine system," *Telemedicine Journal & E-Health*, vol. 12, no. 1, pp. 24–34, 2006.

[128] A. M. Taori, A. K. Chaudhari, S. S. Patankar, and J. V. Kulkami, "Segmentation of macula in retinal images using automated seeding region growing technique," in *2016 International Conference on Inventive Computation Technologies (ICICT)*, vol. 2, 2016, pp. 1–5.

[129] T. P. Thomas, D. D. Anderson, A. R. Willis, P. Liu, J. L. Marsh, and T. D. Brown, "Asb clinical biomechanics award paper 2010: Virtual pre-operative

reconstruction planning for comminuted articular fractures," *Clinical Biomechanics*, vol. 26, no. 2, pp. 109–115, 2011.

[130] E. Tiribilli and L. Bocchi, "Deep learning-based workflow for bone segmentation and 3D modeling in Cone-Beam CT orthopedic imaging," *Applied Sciences*, vol. 14, no. 17, 2024.

[131] E. Tiribilli, E. Denisova, A. Luschi, L. Manetti, L. Bocchi, and E. Iadanza, "Heuristic evaluation of a medical imaging platform," in *IFMBE Proceedings*, In Press.

[132] E. Tiribilli and L. Manetti, "Extremity bones segmentation in cone beam computed tomography: A novel approach," in *MEDICON23 and CMBEBIH23*. Springer Nature Switzerland, 2024, pp. 278–284.

[133] E. Tiribilli, L. Manetti, and L. Bocchi, "Single bone modeler: deep learning bone segmentation for cone-beam CT," in *IEEE Explore*, vol. in press, 2024.

[134] E. Tiribilli, L. Manetti, L. Bocchi, and E. Iadanza, "Bones segmentation techniques in computed tomography: a survey," in *IFMBE Proceedings*, 2022.

[135] C. Turner, J. Lewis, and J. Nielsen, *Determining Usability Test Sample Size*. CRC Press, 01 2006, vol. 3, ch. Determining Usability Test Sample Size, pp. 3076–3080.

[136] M. Vaitiekunas, D. Jegelevicius, A. Sakalauskas, and S. Grybauskas, "Automatic method for bone segmentation in cone beam computed tomography data set," *Applied Sciences (Switzerland)*, vol. 10, no. 1, 2020.

[137] J. Van den Broeck, E. Vereecke, R. Wirix-Speetjens, and J. Vander Sloten, "Segmentation accuracy of long bones," *Medical Engineering Physics*, vol. 36, no. 7, pp. 949–953, 2014.

[138] M. van Eijnatten, R. van Dijk, J. Dobbe, G. Streekstra, J. Koivisto, and J. Wolff, "CT image segmentation methods for bone used in medical additive manufacturing," *Medical Engineering and Physics*, vol. 51, pp. 6–16, 2018.

[139] S. Vasilache, K. Ward, C. Cockrell, J. Ha, and K. Najarian, "Unified wavelet and gaussian filtering for segmentation of CT images; Application in segmentation of bone in pelvic CT images," 2009.

[140] A. Virzi, C. O. Muller, J. B. Marret, E. Mille, L. Berteloot, D. Grevent, N. Boddaert, P. Gori, S. Sarnacki, and I. Bloch, "Comprehensive review of 3D segmentation software tools for mri usable for pelvic surgery planning," *Journal of Digital Imaging*, vol. 33, no. 1, pp. 99–110, February 2020.

[141] R. A. Virzi, "Refining the test phase of usability evaluation: How many subjects is enough?" *Human Factors*, vol. 34, pp. 457–468, 1992.

[142] R. Vivanti, A. Ephrat, L. Joskowicz, N. Lev-Cohain, O. A. Karaaslan, and J. Sosna, "Automatic liver tumor segmentation in follow-up CT scans: Preliminary method and results," *Lecture Notes in Computer Science*, vol. 9467, p. 54â61, 2015.

[143] J. Wallyn, N. Anton, S. Akram, and T. F. Vandamme, "Biomedical imaging: Principles, technologies, clinical aspects, contrast agents, limitations and future trends in nanomedicines," *Pharmaceutical Research*, vol. 36, p. 78, 2019.

[144] D. Wang, N. Li, M. Luo, and Y.-k. Chen, "One visualization simulation operation system for distal femoral fracture," *Medicine*, vol. 96, no. 32, p. 7770, 2017.

[145] D. Wang, K. Zhang, M. Qiang, X. Jia, and Y. Chen, "Computer-assisted preoperative planning improves the learning curve of pfna-ii in the treatment of intertrochanteric femoral fractures," *BMC Musculoskeletal Disorders*, vol. 21, pp. 1–11, 2020.

[146] Wikipedia contributors, "Hausdorff distance — Wikipedia, the free encyclopedia," 2024, [Online; accessed 11-July-2024].

[147] M. W. Wronikowska, J. Malycha, L. J. Morgan, V. Westgate, T. Petrinic, J. D. Young, and P. J. Watkinson, "Systematic review of applied usability metrics within usability evaluation methods for hospital electronic healthcare record systems," *Journal of Evaluation in Clinical Practice*, vol. 27, pp. 1403–1416, 2021.

[148] A.-M. Wu, C. Bisignano, S. L. James, G. G. Abady, A. Abedi, E. Abu-Gharbieh, R. K. Alhassan, V. Alipour, J. Arabloo, M. Asaad *et al.*, "Global, regional, and national burden of bone fractures in 204 countries and territories, 1990–2019: a systematic analysis from the global burden of disease study 2019," *The Lancet Healthy Longevity*, vol. 2, no. 9, pp. 580–592, 2021.

[149] J. Wu, A. Belle, R. H. Hargraves, C. Cockrell, Y. Tang, and K. Najarian, "Bone segmentation and 3D visualization of CT images for traumatic pelvic injuries," *International Journal of Imaging Systems and Technology*, vol. 24, no. 1, pp. 29–38, 2014.

[150] L. Yao, X. Guan, X. Song, Y. Tan, C. Wang, C. Jin, M. Chen, H. Wang, and M. Zhang, "Rib fracture detection system based on deep learning," *Scientific Reports*, vol. 11, pp. 1–10, 2021.

[151] J. Yen, F. Chang, and S. Chang, "A new criterion for automatic multilevel thresholding," *IEEE Transactions on Image Processing*, vol. 4, no. 3, pp. 370–378, 1995.

[152] F. Yi and I. Moon, "Image segmentation: A survey of graph-cut methods," in *2012 International Conference on Systems and Informatics (ICSAI2012)*, 2012, pp. 1936–1941.

[153] Y. Yoshii, T. Ogawa, A. Shigi, K. Oka, T. Murase, and T. Ishii, "Three-dimensional evaluations of preoperative planning reproducibility for the osteosynthesis of distal radius fractures," *Journal of Orthopaedic Surgery and Research*, vol. 16, pp. 1–9, 2021.

[154] Y. Yoshii, Y. Totoki, W.-l. Tung, K. Akita, and T. Ishii, "A comparison of radiographic outcomes between 3D preoperative planning and conventional planning in the osteosynthesis of distal radius fractures," *The Journal of Hand Surgery (Asian-Pacific Volume)*, vol. 24, no. 03, pp. 303–310, 2019.

[155] J. Zhang, C.-H. Yan, C.-K. Chui, and S.-H. Ong, "Fast segmentation of bone in CT images using 3D adaptive thresholding," *Computers in Biology and Medicine*, vol. 40, no. 2, pp. 231–236, 2010.

[156] J. Zhang, T. R. Johnson, V. L. Patel, D. L. Paige, and T. Kubose, "Using usability heuristics to evaluate patient safety of medical devices," *Journal of Biomedical Informatics*, vol. 36, no. 1, pp. 23–30, 2003.

[157] X. Zhang, A. Wahle, and M. Sonka, "Graph-based vertebra segmentation in CT images," *IEEE Transactions on Medical Imaging*, vol. 29, no. 3, pp. 732–744, 2010.

[158] Y. Zhang, K. Qian, Z. Zhu, H. Yu, and B. Zhang, "Dba-unet: a double u-shaped boundary attention network for maxillary sinus anatomical structure segmentation in CBCT images," *Signal, Image and Video Processing*, vol. 17, no. 5, p. 2251 â 2257, 2023.

[159] Y. Zheng, J. Chen, S. Yang, X. Ke, D. Xu, G. Wang, X. Cai, and X. Liu, "Application of computerized virtual preoperative planning procedures in comminuted posterior wall acetabular fractures surgery," *Journal of Orthopaedic Surgery and Research*, vol. 17, no. 1, p. 51, 2022.

[160] S. Zhou, Y. Cheng, Y. Wang, K. Dong, C. Guo, J. Bai, and S. Tamura, "Segmentation of the hip joint in CT volumes using adaptive thresholding classification and normal direction correction," *Journal of the Chinese Institute of Engineers*, 2013.

[161] X. Zhou, R. Takayama, S. Wang, T. Hara, and H. Fujita, "Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method," *Medical Physics*, vol. 44, no. 10, pp. 5221–5233, 2017.