



UNIVERSITAT DE BARCELONA

Final Degree Project

Biomedical Engineering Degree

**“Automatic segmentation of regions
of interest in vaginal brachytherapy”**

Barcelona, 22 de Gener de 2023

Author: Lluís Pellicer Bolet

Director/s: Adrià Casamitjana

Tutor: Aida Niñerola

ABSTRACT

The postoperative endometrial carcinoma treatment often includes radiotherapy (external radiotherapy and/or vaginal brachytherapy) to prevent the reappearance of the tumour. This project aims to improve the efficiency of the vaginal brachytherapy treatment by developing an automatic segmentation algorithm capable of delineating both the clinical target volume and the organs at risk, reducing the time required by experts to exert such task.

In this project, we develop an AI-based framework that uses a V-Net architecture at its core. To train and evaluate the model, we use retrospective CT images and corresponding manual delineations from patients treated in Hospital Clinic.

The creation of the algorithm was achieved successfully, resulting in a completely functional creator of automatic segmentations. About its performance, the results were found satisfactory in the cases of the vagina, the rectum and the bladder, having acceptable discrepancies in the dosimetry output. On the other hand, the bowel and the sigma models would require further improvements since the segmentations obtained didn't match the ground truth.

Overall, the project represents a step forward in the application of artificial intelligence algorithms to radiotherapy related processes.

Keywords: Vaginal brachytherapy, segmentation algorithm, automatically, regions of interest, deep learning, V-Net, organs at risk.

ACKNOWLEDGEMENTS

I would like to thank my tutor Aida Niñerola for accepting me as her pupil in this project, for always being supportive and for providing wise advises through all the process. I would also like to express my most sincere gratitude to my director Adrià Casamitjana, since the project would not have been possible without his invaluable help and for having the patience to guide me through the intricate parts of the project. The colleagues in the Biomedical Imaging Group of Hospital Clínic were also a valuable source of knowledge when the study required an opinion from a professional point of view.

Special thanks to my family and friends (especially Santiago Romani) for their loving support and accompanying me through these last months.

INDEX

ABSTRACT	2
ACKNOWLEDGEMENTS	3
LIST OF FIGURES	6
LIST OF TABLES	7
GLOSSARY	8
1. INTRODUCTION	9
1.1. PROJECT CONTEXT AND JUSTIFICATION	9
1.2. OBJECTIVES.....	10
1.3. STRUCTURE AND METHODOLOGY	10
1.4. SCOPE AND LIMITATIONS OF THE PROJECT	11
2. BACKGROUND	12
2.1. GENERAL CONCEPTS	12
2.1.1. ENDOMETRIAL CARCINOMA TREATMENT.....	12
2.1.2. AI IN MEDICINE	14
2.1.3. 3D SEGMENTATION.....	15
2.2. STATE OF THE ART	16
2.3. STATE OF THE SITUATION	19
3. MARKET ANALYSIS	21
3.1. MARKET EVOLUTION	21
3.2. CURRENT MARKET SECTOR.....	22
3.2.1. 3D SEGMENTATION.....	22
3.2.2. VAGINAL BRACHYTHERAPY.....	23
3.3. POTENTIAL CLIENTS.....	23
4. CONCEPT ENGINEERING	25
4.1. STUDY OF SOLUTIONS	25
4.1.1. DATA ORIGIN.....	25
4.1.2. MODEL	27
4.1.3. TRAINING FRAMEWORK	29
4.1.4. EVALUATION METRICS	31
4.2. SOLUTION PROPOSAL	33

5. DETAIL ENGINEERING	35
5.1. DATA ACQUISITION	35
5.2. DATA PREPARATION.....	36
5.2.1. ORGANIZATION AND FORMAT CONVERSION	36
5.2.2. DATA FUSION AND STANDARDIZATION.....	37
5.2.3. CREATION OF A TEMPLATE	38
5.3. ALGORITHM DEVELOPMENT.....	39
5.3.1. INITIALIZATION.....	39
5.3.2. SEPARATED OAR TRAINING	42
5.3.3. FINAL PRODUCT	43
5.4. RESULTS	44
5.4.1. INITIALIZATION.....	44
5.4.2. SEPARATED OAR MODELS	46
5.4.3. FINAL PRODUCT	47
5.5. DISCUSSION.....	49
6. EXECUTION CHRONOGRAM	51
6.1. WORK BREAKDOWN STRUCTURE	51
6.2. PERT-CPM DIAGRAM.....	51
6.3. GANTT DIAGRAM.....	52
7. TECHNICAL FEASIBILITY	53
8. ECONOMIC FEASIBILITY	55
9. REGULATIONS AND LEGAL ASPECTS	56
10. CONCLUSIONS AND FUTURE WORK	57
11. BIBLIOGRAFY	58

LIST OF FIGURES

Figure 1. CT images of the pelvic region (top) with their corresponding segmentations of the ROIs (bottom) – (axial, coronal and sagittal planes of the body, respectively)	13
Figure 2. Types of segmentation (source: [6]).....	15
Figure 3. U-Net architecture (source: [34]).....	27
Figure 4. Representation of Convolution and Pooling processes (source: [29]).....	28
Figure 5. Representation of the DSC equation	31
Figure 6. Representation of the JC equation	32
Figure 7. Representation of HD equation.....	32
Figure 8. Representation of MSD equation.....	33
Figure 9. Summary of the transformations of the data to adjust it to the model training.....	38
Figure 10. Application of a template.....	39
Figure 11. Visual representation of the training process.	41
Figure 12. Evolution of Training-Validation Dice Loss.....	44
Figure 13. Metrics of the initialization model evaluation in the test set.....	45
Figure 15. Overall performance of the initialization model.	45
Figure 14. Differences between the manual and the automatic segmentation in the initialization model.....	45
Figure 16. Metrics of the individual OAR models evaluation in the test set.....	46
Figure 17. Visual representation of the performance of the individual OAR models. (1. Bowel, 2. Rectum, 3. Bladder, 4. Sigma).....	46
Figure 18. Metrics of the final algorithm evaluation.....	47
Figure 19. Comparison of the final predicted segmentations with respect to the manual contouring. (Sagittal, coronal and axial planes, respectively)	48
Figure 20. Relative dosimetry differences of the automatic segmentations with respect to the manual ones.	48
Figure 21. WBS of the project.....	51
Figure 22. PERT-CPM diagram.....	52
Figure 23. GANTT diagram of the project. [41]	52

LIST OF TABLES

Table 1. Articles about the implementation of DL in 3D segmentation in cancer treatment of abdominal regions.....	17
Table 2. Results of the studies and clinical acceptance	17
Table 3. 3D automatic segmentation products in the market	22
Table 4. Comparison between an external and an internal source of data for the model	26
Table 5. Comparison between the three explored models.....	28
Table 6. Anonymization labels employed in the database.	35
Table 7. Intensities assigned to the different organs in the final Mask file.	37
Table 8. Transformation of the file names according to the official nomenclature.....	38
Table 9. Description of the transforms applied to the image to be interpreted by the model.....	40
Table 10. New transforms for separated OAR training.	42
Table 11. SWOT analysis of the project.....	53
Table 12. Economic viability of the project.....	55

GLOSSARY

VBT: vaginal brachytherapy

CT: computed tomography

ROI: region of interest

DL: deep learning

CTV: clinical target volume

OAR: organs at risk

EC: endometrial carcinoma

EBRT: external-beam radiation therapy

BT: brachytherapy

MR: Magnetic resonance

AI: artificial intelligence

ML: machine learning

CNN: convolutional neural network

DSC: dice similarity coefficient

HD: hausdorff distance

DVH: dose volume histogram

DICOM: digital imaging and communications in medicine

NIFTI: neuroimaging informatics technology initiative

MRN: medical record number

1. INTRODUCTION

1.1. PROJECT CONTEXT AND JUSTIFICATION

Endometrial carcinoma is the most common gynaecological cancer among women. As other types of cancer, it consists of an uncontrolled proliferation of malignant cells, in this case originated in the internal coating of the body of the uterus, known as endometrium [1]. This disease is often treated by surgery, where the uterus is removed, mostly eradicating the cancer. Nevertheless, this treatment is usually combined with postoperative radiotherapy, which aims to prevent the recurrence of the carcinoma.

This prophylactic procedure, called vaginal brachytherapy (VBT), consists of inserting an applicator through which will enter the radiation source into the vagina. In order to compute the dose that has to be applied to the patient, which varies depending on the thickness of the vagina and the disposition and dose absorption of the nearby organs, computed tomography (CT) images must be acquired. These images will be used to delineate the structure of the regions of interest (ROI), a method called segmentation.

This segmentation is done manually, which turns into a considerable amount of work for the specialist that has to exert the task. Each CT exploration contains between 150 and 300 2D slices that will be reconstructed in 3D volumes afterwards, so the specialist has to delineate the ROIs for all these images to obtain results for only one patient. This technique takes hours, consuming a lot of time of the specialists that could be dedicating to other tasks. Moreover, since humans are not exact machines, the results of the segmentation can lead to intra- and inter-rater variability caused by tiredness, differences of experience between specialists or simple errors in the interpretation.

Thus, the main goal of this project is to develop an AI-based algorithm that executes automatic segmentation of the different ROIs: the clinical target volume (CTV), i.e., the proximal third of the vagina wall, and the organs at risk (OAR) that located near the radiotherapy focus. Automating this task will substantially reduce the time required to do dosimetry planning and eliminate the variability caused by human error.

1.2. OBJECTIVES

Main objective

As mentioned before, the main objective of this project is to **improve the efficiency of the vaginal brachytherapy treatment by successfully developing a segmentation algorithm tool capable of delineating the ROIs automatically**, which can be applied to compute the dosimetry in the Radiotherapy Department of Hospital Clínic. In order to achieve that, this project has been divided into different secondary objectives that can be combined sequentially.

Secondary objectives

1. Widen the existing database to train the algorithm.
2. Anonymize and uniformize the data.
3. Develop an AI-based framework for automatic segmentation of CTV and OARs.
4. Assess the accuracy of the model according to quantitative metrics and its impact on dosimetry calculations.
5. Study the feasibility and impact of the framework in real clinical studies.

1.3. STRUCTURE AND METHODOLOGY

This project is a follow-up study from previous works that built an initial structured subset of 220 patients and used them to develop a model that segmented the CTV, excluding the surrounding OARs. In this work, the dataset has been extended including all the new cases of VBT done during last years until the 15th of August of 2023. These patients have been exported from the clinical record in the Radiotherapy Department of the Hospital Clínic using the Oncentra Brachy program.

Once the data has been collected, all images have been anonymized and converted into a readable format for the AI-framework. This requires programming a code in Python language that removes traceable metadata and transforms DICOM images to NIFTI format. Besides, the new acquisitions have been merged with the existing dataset following the same pattern of organization and nomenclature, so the new data has been adapted to the organization system of the previous data.

The AI framework has been trained and evaluated independently to assess its generalizability. A hold-out strategy has been used, dividing the dataset training and testing subsets. To train the model, we further split the training set into training and validation sets to update model parameters and choose optimal hyperparameters, respectively. The trained model effectively predicts the

relevant ROIs in VBT and has been evaluated both from the volumetric (i.e., how well the automatic and manual segmentations match) and dosimetry (i.e., how segmentation differences affect dosimetry) perspectives. For dose calculations, we used the Oncentra Brachy program from the Hospital Clinic Radiotherapy Department.

This process has been carried out by programming with Python but working from the Alfa computer in the Biophysics laboratory of Hospital Clínic. Some useful programs that have been used in this process to work remotely are PuTTY, which is a terminal emulator that allows the programming from an external source, and Filezilla, that allows the exchange of files between two computers. With the aim to ensure the correct development of the project, periodic meetings have been done.

1.4. SCOPE AND LIMITATIONS OF THE PROJECT

The scope of this project includes mainly the points described in the *Objectives* chapter. It is exclusively centred in patients that suffer from endometrial carcinoma that undergo VBT. It wouldn't fit other cases since the morphology of the body after surgery and the radiation method might differ from other therapies. It also works only with CT images, as it is used in clinical practice, and do not generalise to other image modalities. Other images that are currently used in this field are MR images.

The algorithm is set to semantically segment all the different regions of interest of the image, that is to say, to segment each area knowing to what organ it belongs. The number of subjects used in this project is acceptable to implement a trustable trained model, so data augmentation is not necessary. Despite being specially thought to be applied to Hospital Clínic, it could be applied to any hospital with similar technologies.

The computation requirements for training the model have been the main limitation during the development phase of this project. We used the computing cluster from the Biophysics department from the Universitat de Barcelona, which counts on a GPU of 24 GB and a RAM of 8x32 GB. That means that advancements are much more difficult to happen by working from home. Moreover, the training of the algorithm is considerably and computational resources had to be shared with other users, so another limitation of the process would be the limited time to be carried out. About the evaluation, the main limitation is that the manual segmentations used to train the model are poorly segmented, so the results of the automatic segmentation are not expected to be quite accurate. It would be preferable to work with a large database with precisely executed segmentations.

2. BACKGROUND

2.1. GENERAL CONCEPTS

2.1.1. ENDOMETRIAL CARCINOMA TREATMENT

Endometrial carcinoma (EC) refers to a malignancy of the inner epithelial coating of the uterus, called endometrium. It is the most common invasive neoplasm in the female genital tract in developed countries and the sixth most frequent cancer among women. Its incidence and disease-associated mortality are increasing worldwide, representing the 3% of cancer deaths in women [2]. The major affection of this disease occurs in postmenopausal women caused by the rise of estrogen levels, but lifestyle and genetic predisposition can increase the risk factor too [3].

The uterus is composed by the body, where endometrium is found, and the cervix, that connects the uterus with the vagina. EC is produced when there is an excessive estrogenic stimulation that promotes an uncontrolled proliferation of malignant cells in the endometrial lining [4]. Its main symptoms include abnormal bleeding through the vagina and pelvic pain.

The main treatment of endometrial carcinoma is the combination of surgery with postoperative radiotherapy. The surgery, called hysterectomy, consists in the extraction of the whole uterus and often the ovaries and fallopian tubes, eliminating the area where the tumour is located. When the affection has a potential risk of local recurrence, postoperative radiotherapy is applied to minimize such risk. The most likely region where the cancer can reappear after the uterus extraction is the part of the vagina wall that used to lead to the uterus, which will be the target of the radiotherapy. In current clinical practice, two different radiotherapy approaches are found: the whole pelvis external-beam radiation therapy (EBRT) and the intracavity vaginal brachytherapy (VBT). While EBRT is employed to treat cancer in the pelvic region in inoperable patients and cases with an advanced stage of the cancer, VBT is indicated in early stages of the cancer and it is more specific to the target.

The main difference between EBRT and VBT (which is the subject of study of this project) is that in brachytherapy (BT) the radiation is exerted inside the body, in VBT case within the vagina. Thanks to this method, a high dose gradient can be delivered to the target, in this case the top part of the vagina wall, while the radiation that receive the nearby healthy organs is minimized. This is a desired feature since uncontrolled radiation inhibits the proliferation of tumorous cells but also healthy ones [5].

To radiate the target, it is required the use of an applicator to guide the radiation to the target. Only the tip of this applicator radiates the vagina wall to exert the treatment as specifically as possible. This applicator, which can be seen on Figure 1 on the centre of the images, has an adaptative shape composed by cylinders that can vary from 2.5 to 3.5 cm of radius in function of the anatomy of the patient. The delivered dose per patient needs to be planned prior to each radiotherapy session and is computed as a trade-off between the amount of radiation delivered to the target volume, the so-called CTV, and the undesired radiation to the surrounding OARs. This can be achieved using a treatment planning software that receives two volumes as input: a CT image of the patient and the corresponding delineation of CTV and OAR volumes. This delineation must be done in advance.

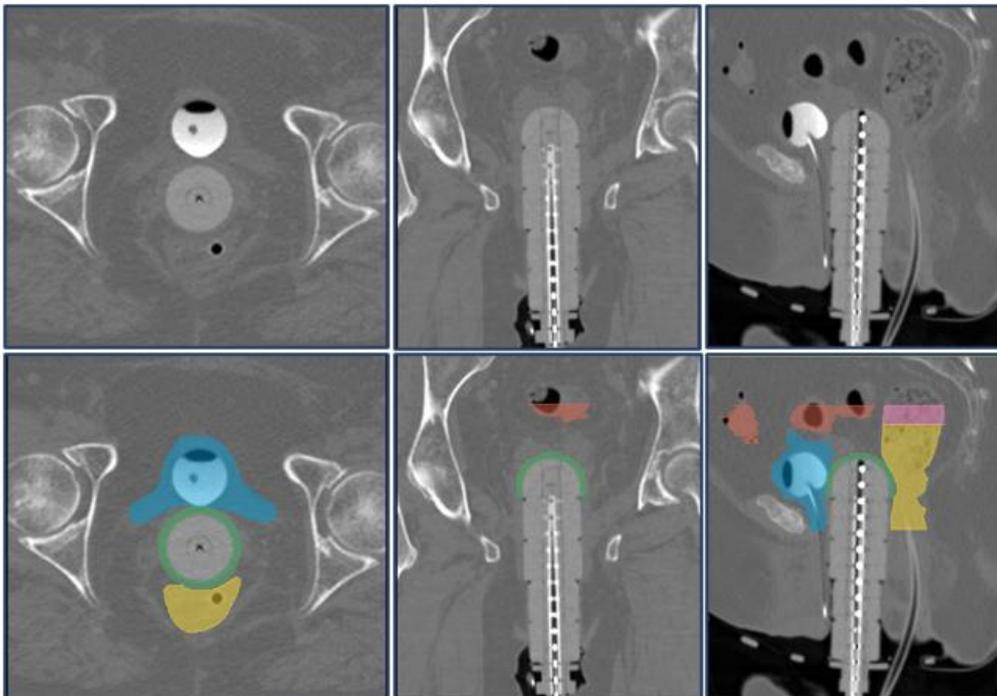


Figure 1. CT images of the pelvic region (top) with their corresponding segmentations of the ROIs (bottom) – (axial, coronal and sagittal planes of the body, respectively)

In order to facilitate the contouring of these organs, two more catheters are inserted into the body: First, the rectal catheter is used to extract the gas of the cavity, and second, the vesical catheter inserts a saline solution into the bladder. Both of these actions allow a better uniformity of the volumes.

Once inserted the catheters and the applicator, the images are acquired with a CT scan that forms a 3D model of the pelvic area composed by hundreds of 2D projections in different axis. This hundreds of images will be the base to delineate the different regions of interest (ROI): the CTV

(green in Figure 1), that refers to the proximal third of the vagina, and the OARs, in this case the bladder (blue), the bowel (red), the rectum (yellow) and the sigma (pink). This delineation process, also called segmentation, is done manually by specialists, which is a challenge since CT images have low contrast resolution and there are a lot of images for each patient. Once the segmentation has been done the dosimetry can be computed, obtaining the quantity of radioactive source (often iridium-192) that has to be inserted into the applicator. The projection of the radioactive source must be performed remotely once the sanitary personnel has left the room to avoid their irradiation. This source will emit radiation in a controlled way during a determined period, and then will be extracted from the body. The treatment is often delivered in more than one session.

All in all, VBT is an effective cancer treatment in low and mid-risk patients that minimizes systemic side effects of radiotherapy thanks to its specificity, preserving organs and their function.

2.1.2. AI IN MEDICINE

Artificial intelligence (AI) concept refers to the emerging vanguard technologies that try to mimic human's ability to think, learn and solve problems. To achieve this goal, computers use specific algorithms that work as instructions to improve performance over time as more data is being processed. AI has advanced to the use of machine learning (ML) with the aim of approaching problem solving.

Machine learning refers to the capacity of a system to learn from a determined database, resulting in the conception of a model that can solve related tasks automatically [6]. Recently, this branch of AI has evolved into deep learning (DL), which derives to the use of artificial neural networks inspired by the human brain to reach accurate conclusions for more complex and non-linear problems without the need of human intervention. DL is conceived to learn and make decisions autonomously, which makes it perfect for image processing and pattern recognition tasks.

DL has become an indispensable tool in the last few years in the medicine field, contributing to the improvement of several areas like robotics, diagnosis, treatments, medical statistics and human biology, among others. Its main application is to decipher complex patterns in large sets of medical data, from patient records to images, in order to analyse and understand physical and biological phenomena [7]. This can be very useful to manage health systems and guide physicians in treatment decisions. Nevertheless, DL must be wielded carefully since its methodology to obtain outputs can be defined as a black box that cannot be interpreted by humans, so it often requires human supervision to assure that the algorithm is not biased.

2.1.3. 3D SEGMENTATION

Image segmentation is an essential technique in image processing and computer visualization that implies the segregation of an image in regions of pixels that share similar characteristics. This allows the recognition, identification and separation of objects of interest, which is fundamental in many medical applications. These applications can go from the diagnosis, identifying and quantifying anatomical structures and pathologies, to the treatment planning and monitoring, assisting surgeries by computer or treating cancer.

Segmentation can be divided into 3 main categories: instance, semantic and panoptic. Instance segmentation is related to counting tasks, detecting each object of a same class. Semantic segmentation, on the contrary, refers to the classification of objects in labels depending on their characteristics, but can't distinguish 2 objects of the same class. Panoptic segmentation represents the combination of the previous two methods, differentiating each object by class and number. Semantic segmentation is the approach that this project has chosen to exert the algorithm.

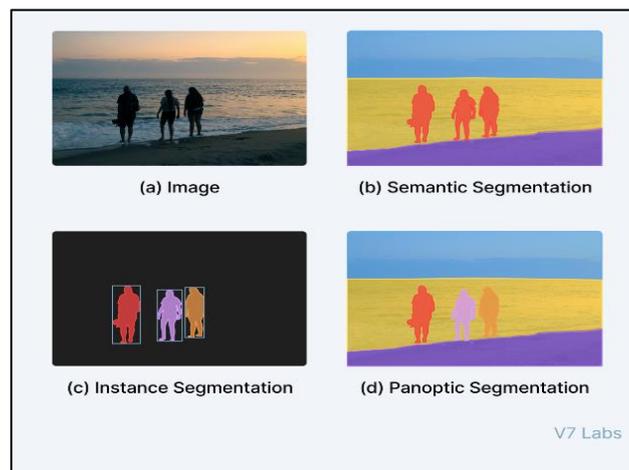


Figure 2. Types of segmentation (source: [48])

The segmentation of medical images has the particularity of being done in 3D, which is very useful to analyse the anatomy and find abnormal structures in the body. Some imaging techniques that can be used to obtain them are CT and MR imaging.

The implementation of DL has probably induced the biggest change in this area by automatizing a process that used to be done manually. Thus, the time required to execute these tasks has been reduced to the minimum, the inter-human variability has decreased and the personalization of the treatment has become an easier goal, adapting to the specific needs of each patient [5]. 3D segmentation has evolved thanks to the parallel advancement in computing power and complexity,

leading to the emergence of new techniques. Overall, DL integrated to 3D segmentation has become a promising tool that keeps refining and spreading in the medical sector, forecasting an increase of relevance in the near future.

2.2. STATE OF THE ART

As mentioned on the previous section, in the last few years medical imaging research has undergone a substantial evolution due to the integration of AI-based algorithms in many subareas of medicine. This upgrade allows to model complex relationships found in the data and the automatization of processes that used to be done by specialists.

Focusing on the cancer treatment planning, which is the subject of study in this project, 3D segmentation with DL is currently being applied to describe anatomically the volumes of organs, tumours, vasculature, and abnormalities present in the patient in a more effective way. Detecting the exact location of the CTV in radiotherapy and the OARs is a valuable information to define more precisely the execution of the procedure, maximizing the effect of radiation in the target and minimizing the side effects in the nearby areas.

Many studies have been published about the implementation of DL models in 3D segmentation tasks related to brachytherapy cancer treatment around the pelvic region. The following tables show an overview of the main characteristics of the most relevant articles that tackle this topic and their results:

Study	Patients	Cancer type	Model	ROI
Wang et al, 2023 [8]	60	Cervix	Modified CNN	Bladder, Rectum, Sigmoid, S. intestine
Olsson et al, 2022 [9]	624	Prostate	MVision	Rectum
Kallis et al, 2023 [10]	40	Cervix	3D U-Net	Bladder, Rectum, Sigmoid
Lempart et al, 2023 [11]	Int-170 Ext-2054	Pelvic	U-Net SENet-154	Bladder, Rectum, Bowel, Femoral head
Duprez et al, 2023 [12]	100	Cervix	nnU-Net	Bladder, Rectum
Zabihollahy et al, 2022 [13]	125	Cervix	3D U-Net	Bladder
Xiao et al, 2022 [14]	313	Cervix	RefineNetPlus 3D	Bladder, Rectum,

Table 1. Articles about the implementation of DL in 3D segmentation in cancer treatment of abdominal regions

As can be seen in Table 1, there are plenty of articles that study the viability of applying DL in the treatment planning stage of cancers located in the abdominal region, such as cervix, prostate and pelvic cancer. In this lower area of the torso, the regions that are mainly segmented are the bladder, the rectum, the bowel, the sigmoid, the small intestine and the femoral head. The difference between this type of studies and other common segmentation studies, such as the ones located in the neck, is that the pelvic region has much more variability between patients than other parts of the body. Despite the difference in location, all the articles use CT images to carry out the study.

There is evidence supporting the idea that the larger a database is the more reliable will be the output of the training of a DL algorithm, but the authors in [8] show that a robust small database could lead to better results than a large dataset with a poor delineation, as the one found in [10]. For a 3D segmentation database to be considered trustworthy it should be previously reviewed and approved by senior radiation oncologists specialised in delineating ROIs.

Looking at the model column of table 1, most of the DL models used are derivations of the 3D U-Net architecture. U-Net is the most common convolutional neural network (CNN) model for medical image segmentation, obtaining contours with high accuracy thanks to the details preserved in the residual connections. Further explanations about its functioning can be found on Section 4.1 (Study of solutions).

Study	Metrics		Dosimetry evaluation	Relative clinical acceptance
	DSC	HD95 (mm)		
Wang et al, 2023 [8]	0.87	1.45	Yes	72%
Olsson et al, 2022 [9]	0.86	5.90	Yes	-
Kallis et al, 2023 [10]	-	-	Yes	30%
Lempart et al, 2023 [11]	0.95	4.86	No	100%
Duprez et al, 2023 [12]	0.85	9.90	Yes	65%
Zabihollahy et al, 2022 [13]	0.85	3.70	No	-
Xiao et al, 2022 [14]	0.95	-	No	100%

Table 2. Results of the studies and clinical acceptance

About the results of these studies, the two main metrics used to test whether the automatic segmentation resembles the manual delineation are the Dice similarity coefficient (DSC) and the Hausdorff distance (HD95). DSC measures the relative overlapping of the predicted voxels with respect to the ground truth. On the other hand, HD determines the largest minimum distance between the ground truth and the prediction. In these studies, the value of the DSC goes from 0.85 to 0.95. It appears that DSC higher than 0.90 can be related to the clinical acceptance. HD, which don't exceed the 10 mm in the examples of Table 2, can be used to detect the presence of outliers since it is very sensitive to them. Values higher than 20 mm could be explained by the presence of a delineation of an incorrect region.

On the evaluation stage, segmentation studies for radiotherapy often include volumetric and dosimetry evaluations. The former aims at evaluating the overlap between manual and automatic delineations, while the latter evaluates the treatment planning deviations when using automatic segmentations instead of manual delineations. In the end, what has to be verified is that the automatic segmentations don't lead to a dose increase, what could cause an overexposure to the OAR, or decrease, leading to an ineffective treatment. For instance, the authors in [10] described two common measures in dosimetry, which are the dose received in 2 cubic centimetres in the OARs (D_{2cc}) and the minimum dose received by the 90% of the CTV (D_{90}). Both measures gave a result of near a 5% of change between the original and the automatic model. The study in [12] used a different parameter, combining the previous metrics mentioned by computing (D_{2cc}/D_{90}), which is called dose volume histogram (DVH) and it is used to determine the differences of radiation that receives each organ. The study in [11], on the other hand, used the mean dose variation (ΔD_{mean}), which takes into account the radiation received in all ROIs. The results showed that there were no significant dosimetry differences, with a variation inferior to 1%.

The percentages in the *relative clinical acceptance* column in Table 2 refer to the number of automatic segmentations that did not require posterior modifications. As can be appreciated, only the studies that didn't examine the dosimetry variations ([11] and [14]) were fully accepted clinically, since they didn't consider the effect that the segmentation could have on the radiotherapy. Other studies as [8] or [12] were partially accepted, with respective values of 65% and 72%. Despite not being accepted as an independent functional tool, these models can be used as a starting point for the segmentation process, demanding only the revision and possible editing of the automatically segmented structures, what entails a significant reduction of the time needed with respect to the traditional method.

In contrast with these other studies, this project focuses on endometrial carcinoma treatment since DL has not been applied yet into 3D segmentation for VBT. Other algorithms cannot be applied to this case given the significant differences in the disposition of the organs caused by the applicator, the rectal and vesical catheters and the lack of uterus.

2.3. STATE OF THE SITUATION

To do so, the project has been developed in collaboration with Hospital Clínic, specifically with the Radiotherapy Department, where approximately one patient per week undergoes vaginal brachytherapy. The images and manual segmentations of these patients are used for model development in the scope of this project.

Since the goal of this radiotherapy-physics group project has a considerable complexity, it has been carried out by 3 different final projects: the first one consisted of data organization and standardization; the second one performed the model training for the CTV segmentation; and the last one, this project, oversees segmenting the OARs and unifying the previous two projects to obtain a final output.

In May of 2023, a master's final project [5] initiated the process by creating an adequate database from a collection of CT images containing the ROIs, so it could be used afterwards for a DL model training. This data was not organized and had not been tested in a preliminary study. CT images from 220 patients were collected from the medical history of the radiotherapy department, covering a range of time of 7 years, from 2014 to 2021. In order to work easily with the acquired data, and coding with Python language, the CT images were transformed from DICOM to NIFTI format. Once the data was transformed, it was standardized to obtain a homogeneous dataset, both in terms of the labels used and the structure.

In June of 2023, a final degree's project [15] continued the project by facing the challenge of evaluating the suitability of DL algorithms to delineate the CTV for dosimetry computations, performing the training with two different CNN. The analysed nets were V-Net and UNETR, which provide some improvements with respect to U-Net despite their similarity. The study showed that both models are reliable, there is no clear better fitting model. UNETR might be more recommended if data augmentation was used. On the other hand, good results can be obtained with V-Net if image filters are applied. The segmentations of the vagina apex were found satisfying.

Finally, this third project, has undertaken the task of segmenting the OAR. This goal is perhaps the most challenging since the CT images have a significant variability due to the anatomical differences among the OARs of the patients, the location and the size of the applicator, the patient's position and the orientation of the CT scan. Additionally, not all the OARs have been completely segmented for all patients, missing some parts. That produces a considerable reduction of the effective dataset. The database update counts on 78 new patients, which were treated between 2021 and July of 2023. That sums up to a preliminary quantity of 298 subjects for the model, which is an acceptable amount to do the training, the validation and the test of the selected CNN, in this case the V-Net. To run the code that manages the data, it is required the use of the *Alfa* computer from the biophysics laboratory of Hospital Clínic, which has a GPU of 24 GB and a RAM of 8x32 GB, allowing the processing of large amounts of information.

3. MARKET ANALYSIS

The rates of uterine cancer are increasing each year by 0.6%, and death rates have risen an average of 1.7% per year too [16], so the use of VBT is forecasted to be established and increased in the upcoming years. Moreover, the implementation of DL in VBT is an emerging sector which still does not comprise much competence. The product that results from this project, that is to say, a 3D algorithm that segments automatically the ROI volumes for vaginal brachytherapy, can be interesting for different fields of medical science.

3.1. MARKET EVOLUTION

Despite the use of BT dates back to the start of the 20th century, it has taken profit of the new technologies that have emerged on the last century, changing from a hardly used technique due to its invasive nature to one of the most relevant therapies for uterine cancer.

The first VBT was performed in 1903, short time after the application of radiation on medicine [17]. From there, many enhancements have been applied, for instance the use of an afterload to minimize the exposure, the use of high activity sources or the optimization of the applicator. Before the computerization of medicine, BT planning could not be based on the anatomy of the patient since the volumes could not be defined, so it was based on the type of applicator. About the prescription of dose, it had to be done manually looking at tables that related exposure with distance, which didn't consider the difference among tissues or the OAR contour. Thanks to the new technologies, currently BT counts on calculus techniques and image processing that help to facilitate most of the planning. For instance, the Monte Carlo simulation allows to estimate the absorbed dose on the different tissues of the body [18].

Due to this computerization, the dosimetry evolved to a process based on the anatomy, with incorporated techniques as the segmentation of the organs to provide a more specific delivery of radiation. Segmentation in medicine has traditionally been done manually by radiologists and technicians, which is a time-consuming process, so some improvements have been tried to be applied on the last few years. The first attempts were related to simple thresholding techniques that tried to discriminate between tissues, but they often required manual corrections. Then the first algorithms appeared, trying to automatize part of the process by analysing characteristics of the image. Some examples can be region-growing algorithms, edge-based methods or clustering algorithms, between others. Nevertheless, they were never applied to BT because its images had too much variability and complexity.

With the introduction of AI, the possibility of automatizing segmentation has become a real possibility using DL networks to carry out the extraction of complex features. The accuracy that show the recent studies described in Section 2.2 (state of the art) foreshadow a revolution in the way segmentation is done, spreading this upgrade in other areas with similar challenges.

3.2. CURRENT MARKET SECTOR

3.2.1. 3D SEGMENTATION

Despite the lack of products on the market directly related to the 3D automatic segmentation applied to VBT, some products with a less specific approach can be found. Some companies with healthcare departments as Philips or Siemens are starting to introduce this kind of software on their stocks, selling the idea that it can be applied to any part of the body. Nevertheless, since the pelvic area has such variability, which is even higher after the uterus removal, the algorithm does not provide accurate results for the patients of interest for this project. The next table shows some of the related products that can be found nowadays in the market:

	Product	Company	Description
1	Pinnacle ³ segmentation with SPICE [19]	Philips Healthcare	Atlas and structures adapt to personalize to the patient with probabilistic segmentation. It does not require much intervention and can be complemented with model-based segmentation.
2	Pinnacle ³ Model based segmentation [20]	Philips Healthcare	Based on a large 3D organ database. The user selects an organ model and drops it into the CT or RM image, automatically segmenting it.
3	Automated segmentation for mPCa [21]	NMMI tools	Self-contouring for metastatic prostate cancer lesions in whole-body PET-CT images.
4	Syngo.MR Onco Engine [22]	Siemens Healthineers	Automatic segmentation of oncologic lesions in MR images for volumetric evaluation.
5	Whole-body dot engine [23]	Siemens Healthineers	Landmark-based automatic segmentation of the anatomical regions of the chest, abdomen and pelvis. Main use for early assessment and follow up of patients.
6	Contour ProtégéAI® [24]	MIM Software Inc.	Automated contouring of normal structures on CT and MR images.

Table 3. 3D automatic segmentation products in the market

As can be seen on Table 3, most of the algorithms that segment different parts of the body have been trained by Atlas or healthy subjects, so it would be difficult for them to contour a postoperative

patient image that has an unknown disposition and shape for the model. Other examples like 3 have been designed to delineate only specific lesions or volumes, what makes them not suitable for other structures. Despite being conceived as an automatic tool, some algorithms as 2 require posterior revisions and modifications to finish the task with precision. Considering these facts, none of the current products satisfy the needs of the segmentation task in VBT framework.

3.2.2. VAGINAL BRACHYTHERAPY

As mentioned previously, the application of DL in segmentation of organs for postoperative endometrial carcinoma treatment has not been done yet. The main methodology currently used for organ contouring is by hand, and it is carried out by technicians and radiologists. Looking at the whole process of the brachytherapy, it is quite clear that this segmentation becomes the bottleneck in terms of the required time, taking between one and three hours to be performed accurately. Moreover, given the repetitive and tedious nature of the technique, it becomes susceptible to vague delineation with poor precision and incomplete segmentations. The variability of the OAR can also lead to human errors and subjective decisions when being implemented by specialists.

On the other hand, Hospital Clínic also considered the possibility of applying the automatic segmentation tool in CT scan offered by Siemens Healthineers company named *DirectORGANS*. Nevertheless, after testing the performance, it was observed that the outcome was not segmenting the ROIs at all. This can be explained by the characteristics of the image. As mentioned on the previous section, when the image is taken in postoperative endometrial carcinoma treatment, there are four main differences with respect to a usual CT image of the pelvic area: Firstly, the lack of uterus, which also promotes a change in the disposition of the surrounding organs; secondly, the insertion of a vesical catheter into the bladder; thirdly, the insertion of a rectal catheter; and finally, the insertion of the radiation applicator through the vagina. All these items modify the anatomy of the patient in a way that an algorithm trained to delineate the normal structure of a body won't be able to recognize the patterns. In conclusion, the Siemens algorithm proposal could not be used in this therapy.

3.3. POTENTIAL CLIENTS

Logically, the main stakeholders of this project will be related to installations with healthcare purposes. Hospitals can be benefitted from the implantation of this technology into their system, improving the accuracy and the efficiency of their VBT planning. It could also open the door to introducing the segmentation step in hospitals that don't count on specialists that can execute this

task and count on the same planning methods for patients. This could help to personalize the treatment, leading to much better results.

The application of the algorithm is suitable for any kind of healthcare installation that perform VBT, from public hospitals in which the government invests in the new technologies, to private ones. It is also appropriate either for specialised oncologic centers as Institut Català d'Oncologia or for general hospitals that have radiotherapy departments as Hospital Clínic. This last one would be the first to apply it since the algorithm is perfectly adapted to its system and would not require further modifications. Nevertheless, it could be applied posteriorly to any hospital worldwide that uses this therapy planning technique.

Even if not considering institutions but individuals as a target for this product, oncologists and other specialists might be interested in reducing the time spent in doing the repetitive task of segmenting, to invest it in spending more time with the patient.

Another sector that could be interested in this proposal are software companies that specialise in medical technology. By acquiring this algorithm, they would possess an upgrade in their product that no other company can offer yet, making their product distinctive with respect to similar competitors. An example of a candidate that could apply this technology is Siemens company, which is the provider of the current equipment found in the radiotherapy department of Hospital Clínic.

In a minor scale, research field could also take profit of this work when developing new DL segmentation algorithms, studying and evaluating the methods that turned out to be successful and mending the mistakes that were made in the process.

4. CONCEPT ENGINEERING

4.1. STUDY OF SOLUTIONS

Since this project is a continuation of two other studies, the approach of the problem has been mainly solved previously. Nevertheless, this section expounds the different paths that have been considered through the progress of the study, from the origin of the data, the type of model or the type of training architecture to the metrics that have been used to evaluate the results.

4.1.1. DATA ORIGIN

The first decision that had to be made during the project planning was what database should be collected in order to train the DL model. This data would consist of anatomical images along with their corresponding manual segmentation of the ROIs that would serve as the ground truth for the model. The main requisites for a database to be acceptable for the project were the following: First, the images would have to show the female pelvic region in CT or MR images, following the acquisition method of the VBT planning. Another relevant trait that should have the images was the lack of uterus characteristic of postoperative endometrial carcinoma patients. The third key point was that the manual segmentations had to contain at least the four OARs (bowel, bladder, rectum and sigma) and the CTV (the proximal third of the vagina wall). Finally, the last characteristic that should have the images of the database, and perhaps the most difficult one to accomplish, was that they should include the applicator of radiation, since its insertion often modifies the disposition of the organs, especially the CTV. Depending on the origin of the data, two main paths were explored:

The first option was to look for public databases of CT or MR images of the pelvic region as an external source of data. The main advantage that it represented was that published databases usually count on a high number of subjects, meaning a bigger database for the training and leading to a more reliable and robust model. Moreover, taking into account that many records of Hospital Clínic have poorly executed segmentations, the idea behind looking for another set of images was in hopes that the segmentation was described more accurately. Finally, considering that public databases are result from a collection of images from different institutions, training the model with generalized data would be beneficial for the implementation of the model in any hospital, avoiding the overfitting to Hospital Clínic's case. On the other hand, since this hypothetical database was supposed to be external, it was very unlikely that it would include all the attributes of VBT images in the desired context.

An example of external database found was the so called CTPelvic1K [25], containing 1184 3D volumes from sub-datasets of the colon, the cervix, the abdomen and the kidney. The images were acquired with CT scans and most of them were sets of fractures and tumours. Only the cervix and abdomen sub-datasets had multiorgan segmentations and none of the images contained the applicator.

The second option was to use the clinical records of the Radiotherapy department in Hospital Clínic. This path had the important advantage of counting on images and segmentations for the training that would be exactly the same type as the ones the model would have to segment afterwards. This includes characteristics as the region (pelvic), the acquisition of the image (CT), the OARs segmented, the applicator and the lack of the uterus. Another benefit of using this data was that the projects that preceded this work [5][15] already collected part of the database and standardized it, making it much easier to manage, and the database would follow similar patterns since it had the same origin.

The main drawback of using these images is that, based on a preliminary analysis of the images, the segmentations had been done poorly in many cases, for instance segmenting only part of the organs or not segmenting some edge slides. As seen in Section 2.2 (*State of the art*), a modest database would be admissible if the segmentations were well delineated, but having fewer subjects with incomplete labels would entail a considerable limitation for the accuracy of the model. Moreover, since the records of the data go back to 2014, the training would be exposed to possible changes in the equipment or methodology through these years that could have affected the way images or segmentations are done.

External data	Internal data
More subjects Possibility of better labels Easy generalization	Specific to postoperative endometrial carcinoma Containing all the OARs Previously collected database Standardized and controlled data
May not be accessible Do not contain all the OARs Do not contain applicator Target different organ or disease	Less subjects Vaguely described labels Possible changes through the years

Table 4. Comparison between an external and an internal source of data for the model

4.1.2. MODEL

In AI, models can be defined as programs that are able to recognize patterns or make decisions from previously unseen data. Due to the complexity of the patterns in the images that this project aims to delineate, the applied model must have an equivalent intricacy. That is why the options come down to CNN architectures, since they have enough capacity to process medical data.

The U-Net architecture is a 2D model that has been proven to be very reliable in segmentation tasks. Its name refers to the U shape of the CNN. As can be seen in Figure 3, on the left part of the U-Net there is a contracting path that does filtering operations at different resolution levels (by means of downsampling operations) to extract features about the context of the input image. On the right part, an expansive path begins to do upsampling to locate the extracted features into feature maps with the same size as the input image [5]. Connecting both paths, the residual connections send the output information of an encoder layer to the parallel input of the decoder layer that is on the same level. This trait is what makes the U-Net stand out with respect other CNN models, since it allows to maintain high resolution details lost during the encoding phase by combining them with the space created in the decoding [26].

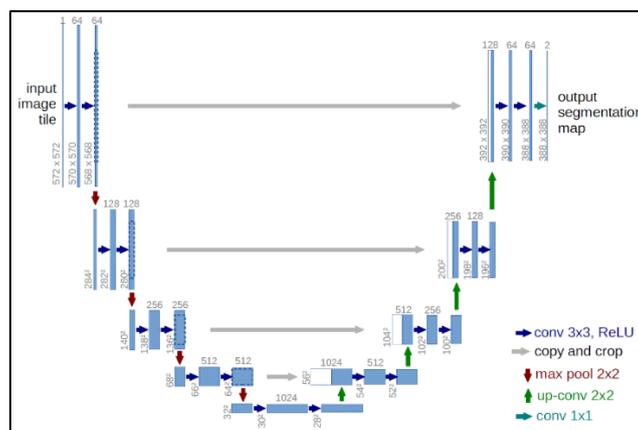


Figure 3. U-Net architecture (source: [27])

The number of output feature maps is equal to the number of segmentation regions (including the background). A final SoftMax function is applied to the output such that each voxel represents the probability to belong to each class. During the training, the filter weights that define the U-Net are updated following a gradient descend fashion according to an objective function between the ground truth labels and the predicted output. This way the model keeps improving as this process is repeated, refining the performance of the weights to assign the correct label to each voxel.

Recently, some upgrades have been done to the U-Net so it can be applicable to 3D images, leading to the creation of the three models described in Table 5. Despite following the same basic scheme explained, each model has its own particularities.

	3D U-Net [26]	UNETR [28]	V-Net [29]
Encoder-decoder path	Yes	Yes	Yes
Residual connections encoding-decoding	Yes	Yes	Yes
Use of Transformer as encoder	No	Yes	No
Residual connections between layers	No	No	Yes
Use of pooling layers	Yes	Yes	No
Activation functions* [27]	ReLu	GELU	PreLu
Normalization * [30]	Batch	Layer	-

** Hyperparameters can be modified, but these are the recommended choice in the reference articles*

Table 5. Comparison between the three explored models

3D U-Net is the direct evolution of the U-Net adapted to be able to segment volumetric data. It has the same symmetrical paths: the encoding path uses convolution layers, that extract the features of the image by applying filters, and pooling layers, that reduce the size of the image matrix condensing the information by doing averages.

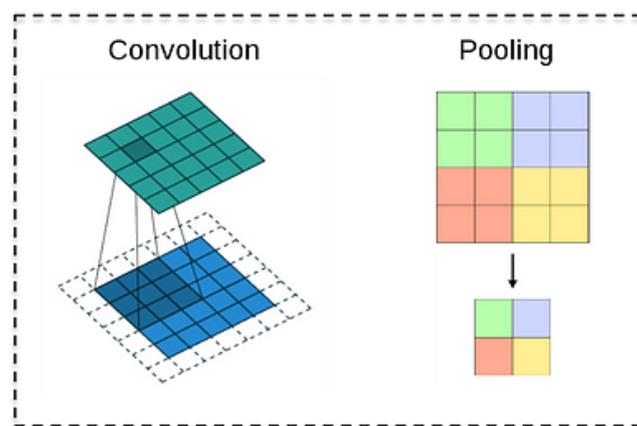


Figure 4. Representation of Convolution and Pooling processes (source: [26])

This way, the image undergoes the process for a determined number of layers following a classic stochastic gradient descent. Once the image is condensed and the features have been extracted,

the decoding path starts to exert de-convolutions and upsampling computations to return the image to its original size. 3D U-Net also uses batch normalization, which is a DL technique that stabilizes the training of a model by normalizing the activation function of the hidden layers, helping the model to be faster [31].

UNETR follows a resembling architecture as 3D U-Net but incorporates some improvements that reportedly help to obtain better results in the segmentations. The main difference with respect to the 3D U-Net is the implementation of transformer blocks in the encoder path. These transformers consist of feedforward neural networks that allow the model to capture long-range dependencies, leading to an optimal learning of the global multiscale context of the image [28]. The 3D images must be previously converted to a 1D sequence so the transformers can operate with it. The decoder path does not incorporate transformers since they don't have the capacity to capture information about the location, only general features. Another difference with respect to 3D U-Net is that instead of using batch normalization UNETR uses layer normalization, which is more recommended for cases in which the batch size is smaller, as in this project [32]. Overall, UNETR properties make the model highly efficient in segmenting small organs and boundary delineations.

V-Net was conceived as a derivation of the U-Net model adapted to work with volumetric data. One of the particularities that differentiates V-Net from 3D U-Net is that, instead of employing pooling layers, V-Net uses convolutional layers with larger strides that substitute the pooling function, resulting on a faster training [26]. Another significant change is that V-Net substitutes the classic stochastic gradient descent by residual connections between layers, in addition to the encoder-decoder residual connections. This modification accelerates the convergence of the training and improves the segmentation results.

4.1.3. TRAINING FRAMEWORK

The quantity, the distribution and the treatment of the database in the training, validation and test sets can affect the posterior performance of the model, so the design of the framework is not a trivial step. Taking into consideration this fact, some options were contemplated to ensure the best possible scenario for the model be trained.

The first improvement that was considered to be applied was the use of data augmentation, which is a technique that creates artificially new data from modified copies of the existing database [33]. These modifications can be performed by geometric or colour transformations, by changing intensities, by mixing images or by applying filters. From the geometric perspective, data

augmentation can modify angles, translations, scale and perform non-linear deformations. From a colour point of view, it can be modified the contrast of the images, the histogram, the quantity of noise or the resolution. This tool would be desirable to generalize the model to fit better in other hospitals. Nevertheless, since the primary goal was to implement the algorithm to Hospital Clínic, where the images are always made the same way with the same CT scan, data augmentation would be redundant and would not provide valuable improvements.

Hyperparameters are parameters whose value is set before the start of the training. In CNN, some examples of hyperparameter can be the number of layers, the type of activation function, the batch size or the optimizer. Most of the hyperparameters are directly recommended by Monai's python package [34], which is the one used for programming DL in healthcare imaging, yet the hyperparameters that have substantial relevance must be studied more carefully.

One of these important hyperparameters is the loss function. The loss function evaluates the difference between the predictions and the labels during the training. The aim is to reduce the loss during the optimization of the model. In segmentation CNN like U-Net, two main loss functions are used: Cross-entropy and Dice. Cross-entropy loss measures the difference between two probability distributions. This means that it calculates the difference between the probability of a predicted pixel to belong to the mask with respect to the ground truth. On the other hand, the Dice loss is a differentiable approximation of the reverse of the DSC (1-DSC), calculating the regions that don't overlap correctly. Some studies [29] state that the Dice loss might work better in cases with class imbalance presence, as in most of medical images.

Changing the focus to the distribution of the database, the images are divided into 3 main sets: the training, the validation and the test sets. The training set is used to allow the model to learn from the segmentations and fit the weights of the CNN to the optimal equation that correctly predicts the disposition of the organs. The validation set is used in the training phase to evaluate if the weights generalize correctly to new data, so it avoids overfitting. Once this process is finished, the test set evaluates the real outcome of the model, assessing its accuracy in unseen data. To separate the data, there are some methods that can be applied:

K-fold cross-validation is one of the most popular techniques for evaluating the model. It is based on dividing the dataset into a specific number K number of subsets that must be previously defined. The data organises using (K-1) subgroups as the training set and the remaining one as the test. On each iteration of the training, the test subset is changed, testing the ability of the model to predict

unseen data to prevent overfitting. A more rudimentary way to achieve this goal is to directly divide the data into fixed groups of train, validation and test. This method allows a faster training since it does not employ iterations, yet it does not have as promising results as K-fold cross-validation.

4.1.4. EVALUATION METRICS

Defining how to measure the performance of a model can be determining for different reasons. First, it has to be chosen a metric that correctly represents what is aimed for, setting if a model gets closer or away from the goal. If the metric doesn't quantify this, it will be challenging to find out if the model is well suited for the task that it should execute. Moreover, selecting a specific metric can be useful to highlight the errors in one section of the algorithm, narrowing the problem to a simpler approach. A similar scenario happens when pointing out the outliers of a metric, to see if the model generalizes correctly or there is overfitting. This way, in segmentation tasks, a large range of metrics can be used to define the accuracy of the model:

Dice Similarity Coefficient (DSC) is a spatial overlap index that measures the relative number of voxels that have been correctly predicted between two sets of binary segmentations. The coefficient can go from 0, indicating no spatial overlap at all between the label set and the prediction set, to 1, meaning a complete coincidence [35]. DSC is defined by the formula described below:

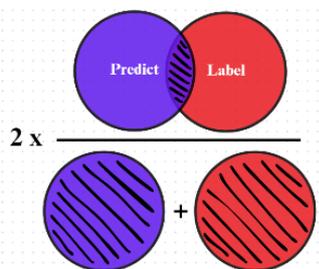
$$DSC = 2 \times \frac{|label \cap prediction|}{|label| + |prediction|}$$


Figure 5. Representation of the DSC equation

The equation multiplies the number of correctly predicted voxels by 2 and then divides it by the number of voxels that sum the label (or ground truth) and the prediction segmentations. Given that DSC has high sensitivity to overlap, it makes the coefficient perfect to evaluate segmentation tasks. That is why, as seen in *State of the art*, it is widely used in this type of studies. Nevertheless, since it doesn't consider contour details or localization accuracy, it is often combined with other metrics that tackle these lacks.

Jaccard Coefficient (JC), similarly to DSC, measures the intersection between the label set and the prediction binary set. The main difference with DSC is that, as can be seen below, JC divides

the intersection by the union of the sets, so it is more sensitive to extreme imbalances between sets while DSC is better in cases where the foreground is much smaller than the background.

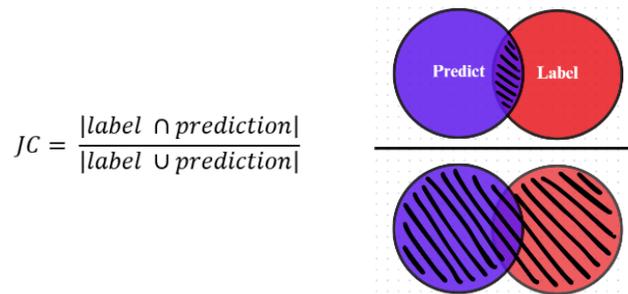


Figure 6. Representation of the JC equation

Hausdorff Distance (HD) measures the largest minimum distance between a point of the label and a point of the prediction contours. In contrast to DSC, this metric doesn't take into account the surface but the delineation of the segmentations. The equation that defines this distance is described below:

$$HD = \max \{ \max_{min} d(lab, pred), \max_{min} d(pred, lab) \}$$

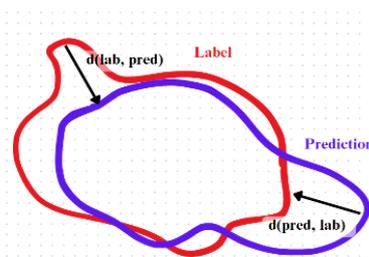


Figure 7. Representation of HD equation

As can be seen on the formula, HD represents the largest distance between the maximum minimum distances of the label points with respect to the prediction and vice versa. HD is often expressed as HD95, which stands for the 95 percentile of HD and it refers to the distance that exceed the 5% of the points. HD and HD95 are effective to detect small outliers of segmentations that would be imperceptible for the DSC, but don't provide valuable information about the overall performance of the model, so they are often used as a complementary metric.

Mean Surface Distance (MSD) represents the average separation between points of the label and the prediction contour. First it calculates the average between the outlier points minimum distances of the labels with respect to the prediction, then it calculates the same but for the prediction with respect to the labels and finally computes the mean between the two numbers obtained.

$$MSD = \frac{1}{2} \left[\frac{1}{N} \sum d_{min}(lab, pred) + \frac{1}{M} \sum d_{min}(pred, lab) \right]$$

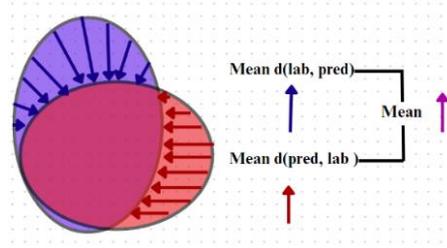


Figure 8. Representation of MSD equation

4.2. SOLUTION PROPOSAL

Taking into account all what has been previously expounded, it was chosen the path that was considered to contribute the most to get good results assuming the less drawbacks as possible.

About the data origin, the internal database of Hospital Clinic was selected, following the steps of the previous projects and rejecting the new proposal of collecting external data considering the lack of ground requisites that it satisfied.

About the chosen option, it could be approached in three different ways: Using the previously collected database, using only new recent data collected exclusively for this work, or adding the new data to the original database. Considering that the years of acquisition of the new data (2021-2023) weren't too distant from the old data (2014-2021), the images didn't differ noticeably, so there was no reason to not join them to obtain a bigger database. This way the third option was selected. Using the internal database, it had to be assumed that the segmentations that were going to be used to train the model didn't have the best quality, so the results were expected to have some limitations.

About the model, there was the starting limitation that the models that should be used had come down to UNETR or V-Net, since they were the ones used in the previous project in which the vagina wall was segmented. However, there was a justification behind the preselection of these two models, being that they are the models that best perform segmentation in medical images currently.

That said, considering the previous project results and some preliminary tests performed on the beginning of this project, it was decided that the model that best adjusted to the needs of our goal was V-Net. Despite UNETR was slightly favoured in [15], it should be noticed that the segmentation targets in both projects have different properties. While the vagina wall is a small organ with few variations, what makes it a good target for UNETR, the OARs often are large organs with a lot of variation.

Changing the subject to the training framework, it was decided that data augmentation was not necessary for this project, since the number of subjects was already considered satisfactory and its application would not entail noticeable better results. About the loss function, Dice loss was considered to provide better results in medical segmentation tasks. Despite being a more robust choice, K-fold cross-validation was discarded due to time limitations. Therefore, the direct separation was selected because of its lack of computational expensiveness.

About the metrics used to assess the performance of the model, a DSC, HD95 and MSD combination was chosen to address the evaluation from different angles without using redundant data. Since JC is so similar to DSC, it was discarded. In order to deal with the contour inaccuracies, HD95 was selected rather than MSD since it is more sensitive to false little volumes that often occur in automatic segmentation tasks. Nevertheless, it was also implemented the MSD in the test evaluation to assure the maximum comprehension of the nature of the errors.

5. DETAIL ENGINEERING

5.1. DATA ACQUISITION

In order to start the project itself, the first step that was carried out was the collection of the data that would serve as the base for the model to learn. To do so, the records of the Radiotherapy Department in Hospital Clínic were revised. There, all the activities performed in the department are stored in chronological order. The target data were the images and segmentations of patients that had undergone vaginal brachytherapy with cylinder applicators from the October 6th of 2021 to August 15th of 2023. In this period of time, 80 patients that fulfilled the desired profile were found. From these patients, only the first session of brachytherapy was selected, since collecting more than one session would lead to a redundant database. Based on a preliminary analysis of the images, 2 of the 80 patients were found to be misclassified, leaving a total number of 78 subjects.

Once the subjects were listed, their data had to be exported from the *Oncentra brachytherapy* program inside the department computer, where the information was stored in DICOM format files. DICOM is used in medicine for storage, transmission and processing of medical images [5]. However, since these DICOM images contain patient details, acquisition data and other information that is considered sensitive content, everything that could link the patient to the data had to be previously erased to respect the privacy rights.

Programming with Python, it was developed a program that read all the slices of all the images and segmentations to extract the sensitive information, from the name and the date of birth of the patient to the medical record number (MRN) and the study ID. These labels were replaced by the ones on the following table:

Before	After
Name and Surnames	Sub-x
MRN	100+x
Date of birth	1990 / 1 / 1
Study ID	000Y

Table 6. Anonymization labels employed in the database.

Before erasing all the identifying data, it was stored in a CSV document with the corresponding transformation to recur it in case it was necessary. Once this process was finished, each patient had a label Sub-x in which the x was a number from 001 to 078, and their corresponding MRN with

a number from 101 to 178. This was done with the aim to posteriorly order the slides of the patients, which were all mixed up in a unique folder. At this point, the database was safe to be extracted from the Hospital Clínic and operate with it.

5.2. DATA PREPARATION

The second process that had to be carried out was the transformation of the data into an organized and standardized format that could be directly introduced to the model to obtain an output. In order to achieve this goal, the process was divided into four subphases: the organization of the DICOM images into patient folders, the transformation to NIFTI format, the standardization of the data, the fusion with the already existing database and the registration to the template.

5.2.1. ORGANIZATION AND FORMAT CONVERSION

Starting with the transformation of format, converting the data from DICOM to NIFTI format was a crucial step to facilitate the manipulation of the data. It is relevant to note that DICOM images are stored in 2D slices, providing a total of 100-300 2D files for each patient. On the other hand, NIFTI is a standard simplified format that operates with 3D files that are defined by matrixes of voxels. This way, the CT images would be transformed from a group of hundreds of slices to a single file. On the other hand, the manual segmentations stored in DICOM-RT struct format, which is used in radiation therapy planning, would be divided from a single DICOM file to a group of 5 NIFTI files in which each one of them would contain mask of one ROI.

To perform this conversion, it had to be modified an existing code originally programmed by [5]. This Python code first organized all the files by patients, copying the CT images and the RT struct into the corresponding patient (sub-x) folder, and ordering the CT slices by number. This way there were 78 folders, where each one contained the CT slices enumerated and the segmentation file of one patient. This process was performed with the *pydicom* library. The second part of the code was in charge of the format conversion. With the library *DicomRTTool*, the DICOM CT slices were transformed to a single file in NIFTI format, named *image.nii*. On the other hand, the RT struct file containing the five segmentations of the different organs (the CTV and the four OARs) was converted into five NIFTI files, one for each segmentation. They were named *mask_ORGAN.nii*, in which it was specified if the mask belonged to the vagina, the bowel, the sigma, the bladder or the rectum in the place of *ORGAN*.

The code also created a csv file called *participants.csv* that stored a list of the patients with the name of their folder and the number of CT slides that originally had the image, in case this information was posteriorly needed.

5.2.2. DATA FUSION AND STANDARDIZATION

With the aim to simplify the training algorithm, the collected database had to be unified and standardized. This process was divided into three parts: Joining the OAR masks into one file, renaming all the database to agree with the adequate nomenclature and joining the new database with the existing database.

Starting with the fusion of the OAR masks files, a Python code was programmed so it would create a new NIFTI file and fill it with the corresponding masks, providing each organ a different number, specified in Table 7. In consequence, the mask would not be binary anymore since each organ mask would have a different intensity. This way the OARs could be distinguished between each other. The colours of the table will refer from now on to the corresponding organs in images of segmentations and plots of the results to allow the maximum comprehension of the figures.

Organ	Bowel	Rectum	Bladder	Sigma	Vagina
Intensity	1	2	3	4	5

Table 7. Intensities assigned to the different organs in the final Mask file.

The files were named *mask_OAR.nii* and were stored in new separated folders, also divided by subjects. Moreover, two versions of the file were created: The first version only included the OARs and was the one used for the first part of the project, given that the vagina had its own autosegmentation model conceived in the previous project [15] and was not required for the training of the OAR model. On the other hand, the second version included the OARs and the CTV and was reserved for the testing of the final algorithm that would eventually be applied into clinics, which had to contain all the ROIs.

At this point of the project, files needed to train the model had the right format to start the process. First, the CT file, that would serve as the input of the model. Secondly, the manual segmentations file, that would serve as the labels or ground truth for the training to evaluate if the model was improving. Nevertheless, these files had to be named according to the adequate nomenclature and adjusted to the already existing database created in the previous projects. The resulting nomenclature is exposed on Table 8, which indicates the number of the subject, the number of the

VBT session and the type of image. The number of the subjects had to be modified to agree with the already existing database that had 220 patients ($y = x+220$).

Data	Original name	Official nomenclature
Image	<i>image.nii</i>	<i>Sub-y_ses-0_CT.nii</i>
Label	<i>mask_OAR.nii</i>	<i>Sub-y_ses-0_desc-oar_dseg.nii</i>

Table 8. Transformation of the file names according to the official nomenclature

Once completed all the transformation steps summarized on Figure 9, the new database was joined with the previously collected database, forming a final database of 298 subjects.

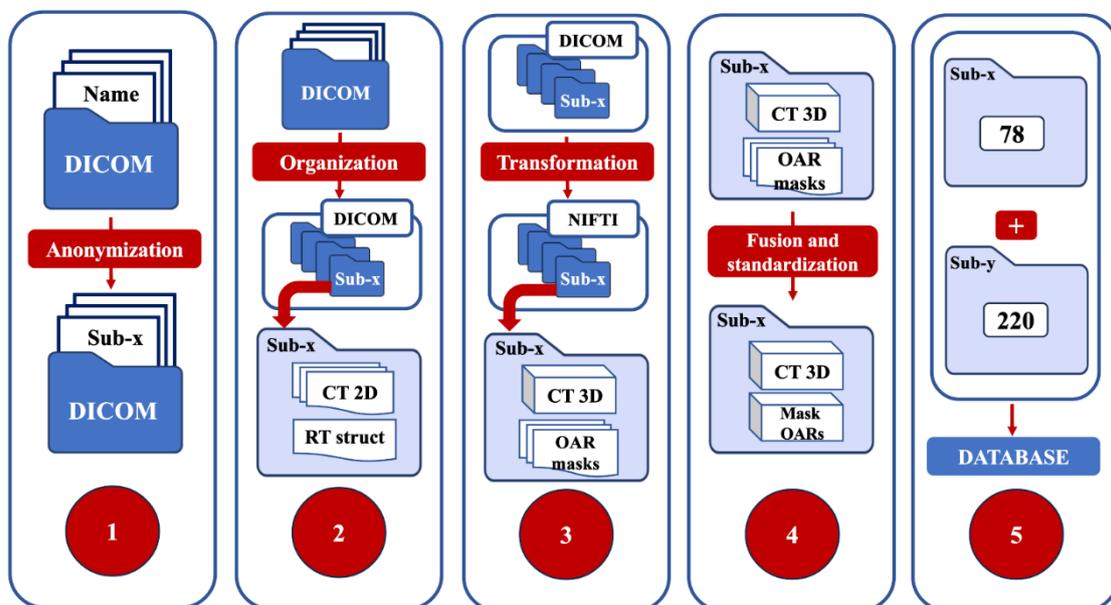


Figure 9. Summary of the transformations of the data to adjust it to the model training.

5.2.3. CREATION OF A TEMPLATE

Before starting the training, a final transformation was performed to facilitate the learning process of the model. Instead of inserting to the model the complete CT image, only the area with clinical relevance was kept. By applying a template, the variability would be reduced and the performance of the model would improve without losing dosimetry accuracy.

The clinical relevance of the dosimetry comprises a ratio of 2 centimetres above and below the applicator. The code was programmed to crop the image 2 centimetres above the tip of the applicator, and 2 centimetres below the end of the radiating area described by the semicircle of the

top. In practice this was achieved by rigid alignment of each image to a given template where the cropping was defined.

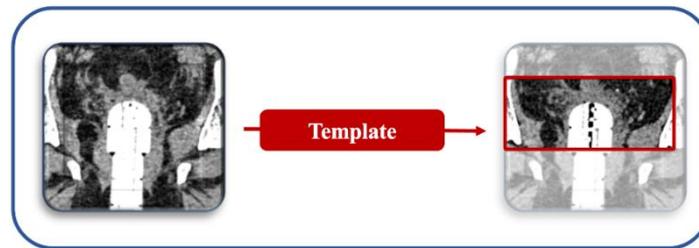


Figure 10. Application of a template

5.3. ALGORITHM DEVELOPMENT

To create the models that would be used in the algorithm, it was employed the Alfa computer in the biophysics laboratory of Hospital Clínic, which had the required power to train the algorithms. With the aim to facilitate the use of this computer to future users, it was developed an instruction manual that could guide new researchers into the protocols and the functioning of Alfa.

The programming of the algorithm was conceived on the basis of the decisions that had been previously made. The model architecture would be V-Net and the library *Monai* [34] would set the programming framework.

From a preliminary analysis of the database, all the patients that didn't have all the OAR regions segmented were excluded. Due to the variability and complexity of the OARs, the preliminary tries of training lead to the misclassification of certain areas. To facilitate the training, it was divided into two codes: The initialization, in which all the OARs would be segmented together to locate approximately the regions, and the training of each organ separately, which would provide better outcomes for the contouring details.

5.3.1. INITIALIZATION

As has been already mentioned, the preparation of the initialization model was conceived as a standard training of all the OAR regions. The structure of the Python program was composed by the division of the data, the transforms applied to the images, the setting of the model and the training scheme.

First, the database was split using *train_test_split()* function. It was determined that the train-validation set would be composed by 80% of the database and 20% would form the test set. Then, the 80% of the train-validation was split again into 80% train set and 20% validation set. Therefore, the final distribution of the database was 64% train, 16% validation and 20% test. The subjects were randomly distributed into the three groups to ensure the maximum heterogeneity, and the final result was saved.

Following with the code, the next step was to process the images by applying transforms. Using *Monaí's* library, a series of transforms were connected to ensure the optimal interpretation of the data. Their function is detailed on Table 9:

Transform	Function
<i>LoadImaged()</i>	Loads the image, the label and the vagina segmentation into a dictionary to make them operable.
<i>EnsureChannelFirstd()</i>	Solves an ordering issue related with programming characteristics of some libraries. It brings the channels before the dimensions.
<i>Orientationd()</i>	Orients the image into specific axes to make all the set images follow the same coordinates.
<i>DivisiblePadd()</i>	Since V-Net loses resolution when the image pixels are not divisible by 16, this function creates a Pad of the image to make the spatial size become multiple of 16 in all the axis.
<i>ScaleIntensityRanged()</i>	The intensity of the image is normalized to highlight the differences between tissues.
<i>RandSpatialCropd()</i>	Crops the image in random windows (but with constant size) to avoid the image to be too computationally expensive to be processed. By doing it in random windows it ensures that the model learns all the parts of the image.
<i>AsDiscreted()</i>	Prepares the labels to be trained or evaluated, binarizing or codyfing them.

Table 9. Description of the transforms applied to the image to be interpreted by the model.

For the setting of model, it had to be defined that V-Net was the architecture employed and that the images had 3 dimensions. Besides, it was defined that there were 2 input channels (the image and the segmentations of the vagina) and 5 output channels (the 4 OAR and the background). As hyperparameters, the loss function was defined as Dice loss as specified on the section 4 (*Concept Engineering*). On the other hand, Adam's optimizer was selected since it was the one recommended in *Monaí's* platform from an empiric study [36].

Finally, a training scheme that had been described in an independent code was brought up. Besides the variables described on the previous paragraph, 2 more variables related with the epochs were introduced. An epoch can be defined as a complete cycle of training of all the database. The first variable created was the maximum epochs that could undergo the training phase, which was set to 5000. The second was the validation interval, that defined how many epochs had to proceed to execute another validation phase, which was set to 5. The validation wasn't done in each epoch since it would be very time consuming. Once established the variables, the code would create a training and a validation csv files to store the results of the epochs. The vagina segmentation was loaded too to avoid the model to segment on that area. This way the OAR algorithm would incorporate the knowledge to never overlap the CTV region.

To begin the training with the first epoch, the model would load all the images of the training set and the weights would start to be shaped from random values. On each iteration, that is to say, on each subject segmented, the weights of the previous iteration would be optimized to approach the best fit to the data. At first, the modifications of the weights on each iteration would be broad, but as the weights got closer to the perfect fit, the variations would get more subtle and refined.

This way, the model would undergo 5 epochs, and then would evaluate the resulting weights in the unseen validation set of images to see if the weights generalize correctly with the unseen data. This step would avoid the overfitting of the weights to the training data. If the weights performed better than the last stored weights, the *best_metric* variable would be replaced by the new weights. If not, the *best_metric* would remain the same. If the best metric hadn't changed in a lot of epochs, that was a sign that the weights were starting to overfit to the training data and the model wasn't generalizable, so the loop was stopped.

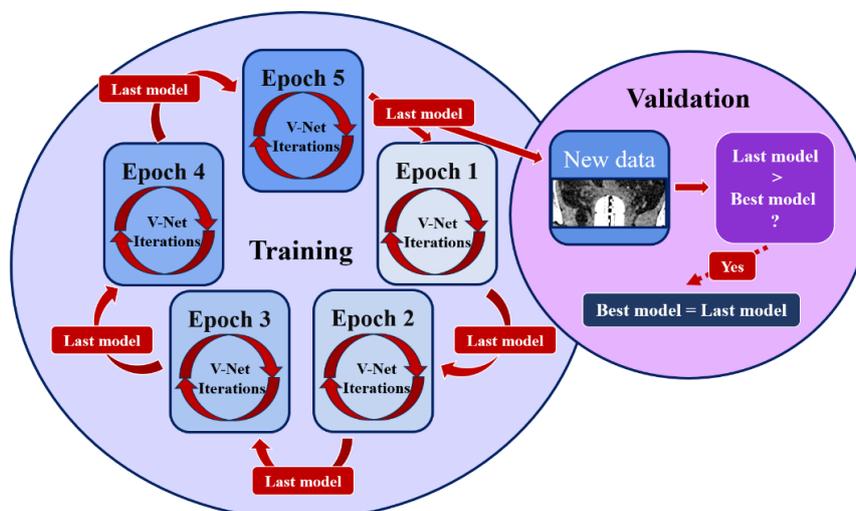


Figure 11. Visual representation of the training process.

On the evaluation stage, the training code was slightly modified to perform only the evaluation of the best metric model in the unseen test set and store the results. About the transforms, they remained all the same with the exception of the *RandSpatialCropd()*, that was erased to make inferences in the entire image. The model part remained the same, but the metrics (DSC MSD and HD95) were added to evaluate the results from different perspectives. About the test scheme, the images were loaded, passed through the V-Net with the best metric weights from the training, and compared with the labels using the different metrics. Then, using *Nibabel* library, the segmentations created by the model were saved in a new folder together with the image and the manual segmentation.

5.3.2. SEPARATED OAR TRAINING

The second phase of the algorithm was to train the model to identify separately each one of the organs to see if it could perform better. To do so, the training code was rearranged to filter the labels to accept only the manual segmentation of one organ of interest. A new variable, named *tissue_class*, was created to select the desired OAR by inserting one of the numbers from the *Table 7* that identified the tissue.

As in the initialization training, only the train and validation sets were loaded. Besides the previously used transforms, two new transforms were introduced to cope with the new needs of the training:

Transform	Function
<i>LabelFilterd()</i>	receives <i>tissue_class</i> as an input and filters the label to return only the needed organ
<i>RandCropByPosNegLabeld()</i>	Based on a positive-negative ratio of the image, detects the centre of the ROI and crops the image to include only the organ.

Table 10. New transforms for separated OAR training.

The rest of the code was left the same as the initialization training, including the model and the train scheme. The test code also remained the same as the initialization test, except for the addition of the mentioned modifications in the transforms and the variables.

Once made the codes, four training and four tests were executed, one for each OAR. The best metric models were stored to load them on the final product.

5.3.3. FINAL PRODUCT

At this point of the project, all the required models had been obtained. First, there was the model that segmented the vagina, which had been developed in [15]. Later, there were the initialization model for the OARs and the four models for the bladder, rectum, bowel and sigma, developed in this project.

The final product would have as an input the DICOM CT slices from the scan and would generate a DICOM-RT struct file with the automatic segmentations. The code was divided into the next steps:

1.Loading of the data: The code was programmed so, when it was run, a window would pop up showing the folder system of the computer to select the folder that contained the CT slices. The six models would also be loaded automatically.

2.Description of the transforms: The transforms used in the final product were the already mentioned `DivisiblePad` and `ScaleIntensityRanged`. Moreover, it was introduced a new transform named `Invertd()`, that inverts all the previously added transforms to return to the original image characteristics so it can be stored.

3.Setting of the model: In this case, the model was actually conformed by 3 different submodels: The CTV model, the initialization model and the OAR models. Each model was loaded with its own best metric weights.

4.CTV prediction and registration to the template: First, the DICOM image would be transformed to NIFTI. Then the vagina would be segmented. Using this prediction, it would be calculated where to derive the bounding box that defines the template of the image.

5.Initialization: By implementing the initialization model, a first approximate segmentation of the OARs would be performed. With this information, each OAR would be cropped to be segmented accurately on the next step.

6.Refinement of the OAR segmentations: Then, the individual OAR models would be implemented in the respective cropped images obtained on the previous point. This way the segmentation would be refined, segmenting only inside the cropping in a more accurate way.

7.Postprocessing: With the aim to maximize the precision of the segmentation, a postprocessing step was added. For the bladder, the rectum and the sigma predicted segmentations, the groups

of foreground voxels were found, and only the biggest group would be kept. This way the false positive organs would be erased, leaving the real segmentation. This process could not be applied to the bowel due to its variability and distribution in separated groups of voxels. It was also applied a hierarchy defining to which organ should a voxel belong in case of overlapping of two different organs.

8. Storing: Finally, the predicted segmentations would be stored after converting the NIFTI file into a RT struct file, which is the right format to insert the segmentations into the planning program of the dosimetry.

5.4. RESULTS

5.4.1. INITIALIZATION

As can be seen on Figure 12, during the initialization training process, the Dice loss of the training was being reduced on each iteration. Meanwhile, the validation loss indicated that the weights were starting to overfit to the training set over the 500 epochs and were not generalizing to the validation set. The best initialization model was met on the early 400 epochs.

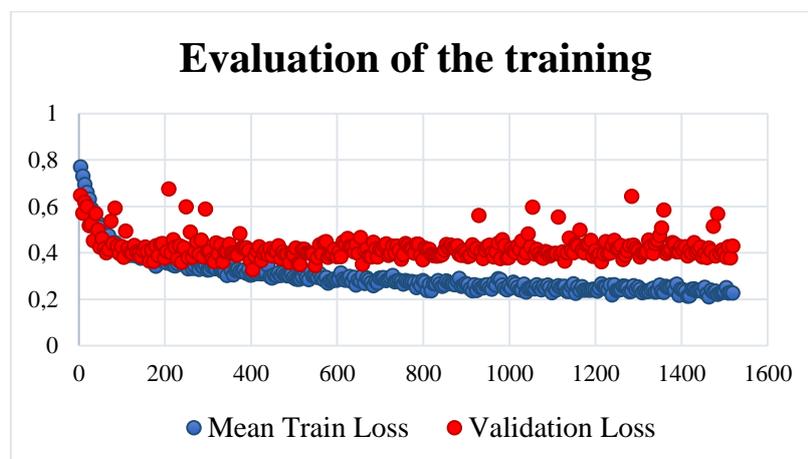


Figure 12. Evolution of Training-Validation Dice Loss

As the results of the test showed, a single model approach to segment the OAR regions would require further improvements. While the rectum and the bladder obtained a median DSC of 0.7 and 0.8 respectively, the bowel and the sigma results were significantly worse, with a DSC of 0.5 and 0.07. These results were corroborated by the MSD values, that showed an overall reasonable performance in the bladder, the rectum and the bowel with less than 2.5 cm of median distance error, but more than 5 cm in the sigma. Finally, HD95 showed that there was a significant amount of false predicted regions, specially in the rectum and the sigma.

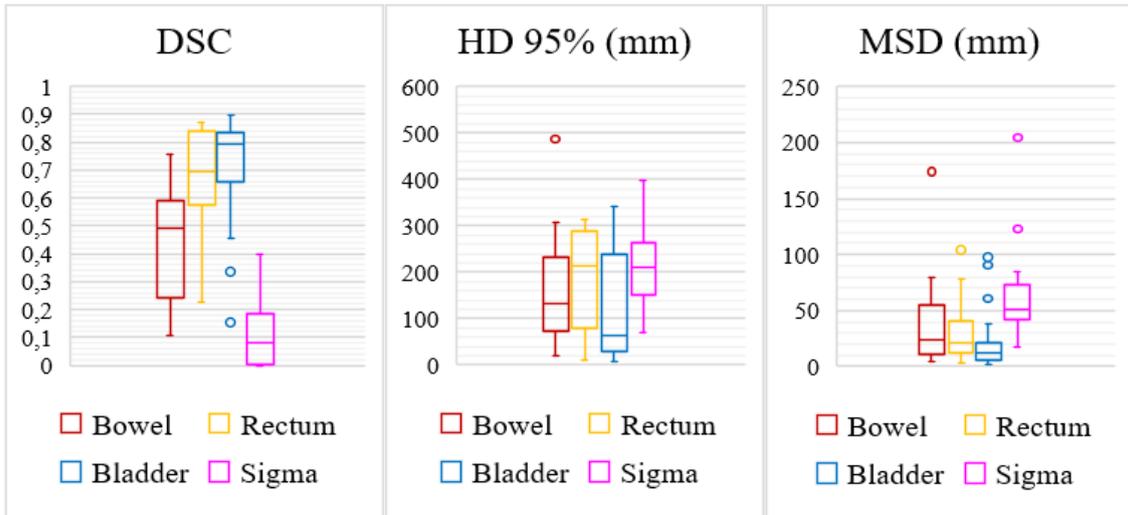


Figure 13. Metrics of the initialization model evaluation in the test set.

On the top part of Figure 15, an example of false prediction of the rectum detected by the HD95 can be appreciated. In this case the rectum is predicted on the bladder region. On the low part, there is an example on how the ground truth (the manual segmentations) isn't always well delineated, compromising the reliability of the metrics. The overall performance of the initialization model can be observed on Figure 14.

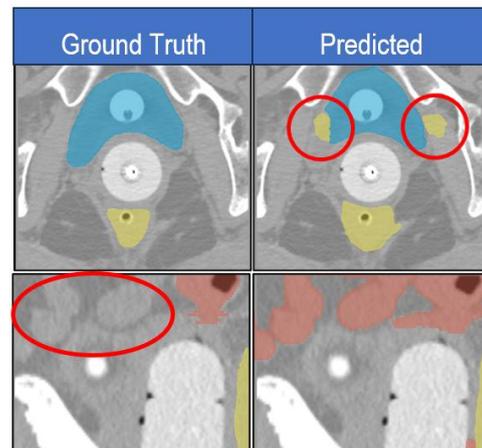


Figure 15. Differences between the manual and the automatic segmentation in the initialization model.

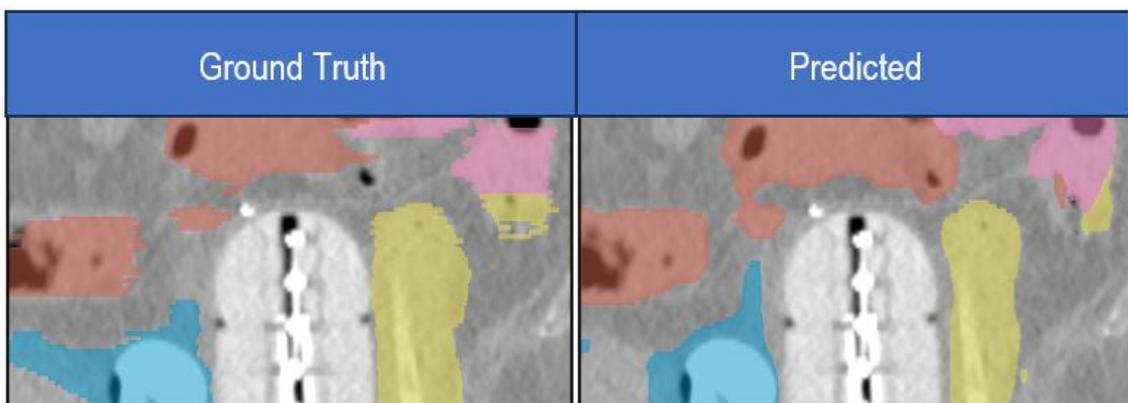


Figure 14. Overall performance of the initialization model.

5.4.2. SEPARATED OAR MODELS

For the individual training of the OARs, the results were enhanced with respect to the initialization. For the DSC, the median values obtained were 0.6 for the bowel (10% of increase), 0.85 for the rectum (15% of increase), 0.88 for the bladder (8% of increase) and 0.3 for the sigma (20% of increase). The HD95 values were reduced a 65% in the bowel, a 93% in the rectum, an 84% in the bladder and a 65% in the sigma with respect to the initialization. This indicated a substantial reduction in the false predicted regions. The MSD also was reduced from a median of 2.6 cm in the OARs to a median of 0.8 cm.

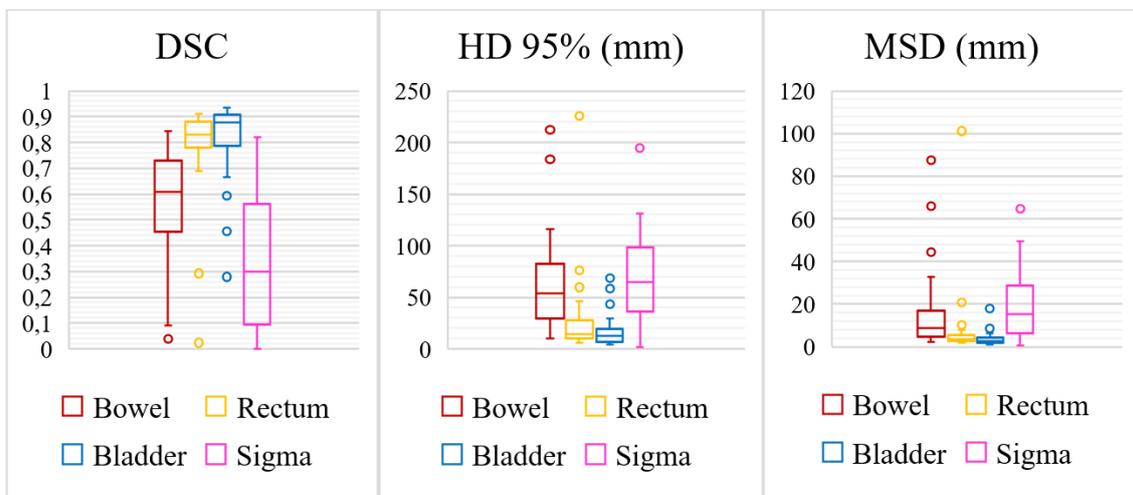


Figure 16. Metrics of the individual OAR models evaluation in the test set.

The Figure 17 shows an example of each one of the model performances. As can be observed qualitatively, the rectum and the bladder are almost perfectly segmented, the bowel offers a good approximation of the real shape and the sigma is partially well segmented but not in all the regions.

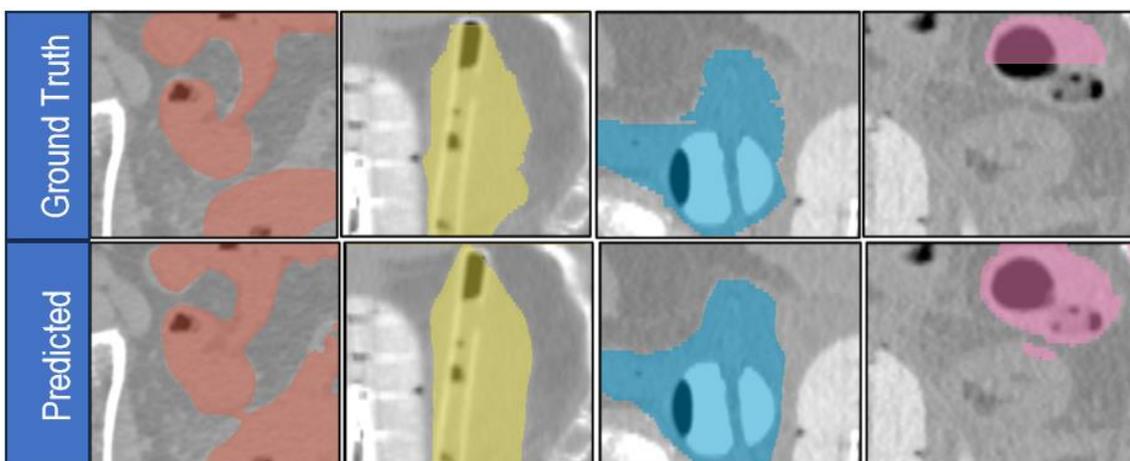


Figure 17. Visual representation of the performance of the individual OAR models. (1. Bowel, 2. Rectum, 3. Bladder, 4. Sigma)

5.4.3. FINAL PRODUCT

The final product consists of an executable file (EXE file) containing the algorithm. This file would be downloaded on the computer where the planning is performed, and with the simple step of opening the file the code would be run. This action would open a window showing the folders of the computer so the CT image could be selected. By eliminating all the intermediate steps to run the code from the terminal, the implementation of the algorithm won't require any programming knowledge.

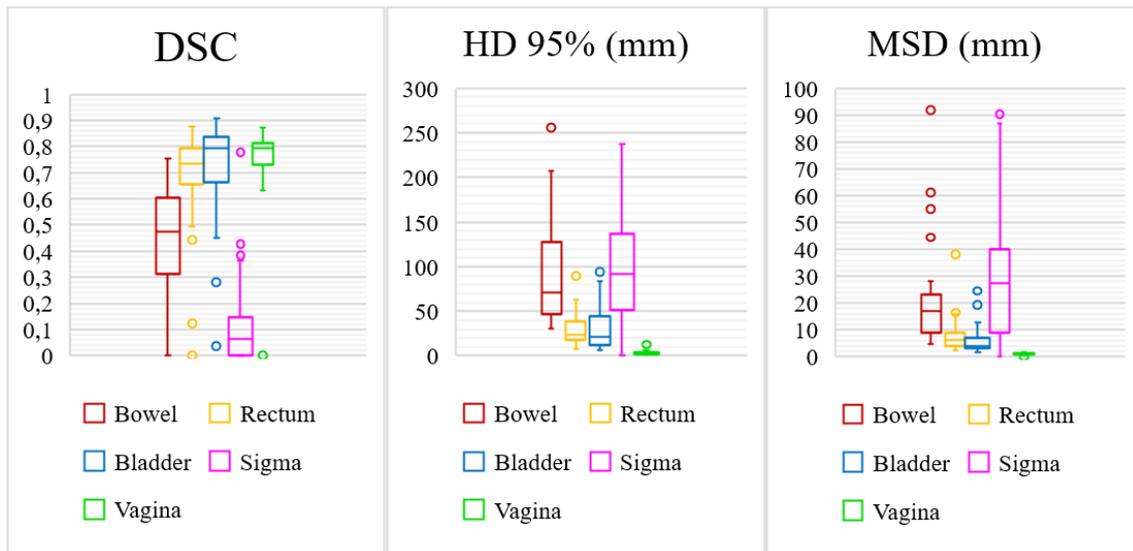


Figure 18. Metrics of the final algorithm evaluation.

Regarding the performance of the final product, the same three metrics were used to evaluate the results. As can be seen on Figure 18, by adding the individualized OAR models into the algorithm, the output improved with respect to the initial algorithm (the initialization). Focusing on the DSC results, while most of the results remained the same, the rectum's DSC increased a 7% in the overlapping. This way, the final median DSC values obtained were 0.47 for the bowel, 0.74 for the rectum, 0.80 for the bladder, 0.07 for the sigma and 0.79 for the vagina wall.

About the error on the contours, the results of HD95 were 6.7 cm for the bowel, 2.3 cm for the rectum, 2 cm for the bladder, 8.8 cm for the sigma and 0.3 cm for the vagina. This shows an overall 69% reduction in the HD95 values with respect to the initialization, meaning that there were substantially less false predicted OARs. On the other hand, the MSD values obtained were 1.3 cm for the bowel, 0.6 cm for the rectum, 0.3 cm for the bladder, 6.5 cm for the sigma and 0.1 cm for the vagina. This way, the contour detail was improved a 48% globally, but the error increased a 30% in the sigma contouring, since there was an undesired reduction of its volumes, as can be seen on Figure 19.

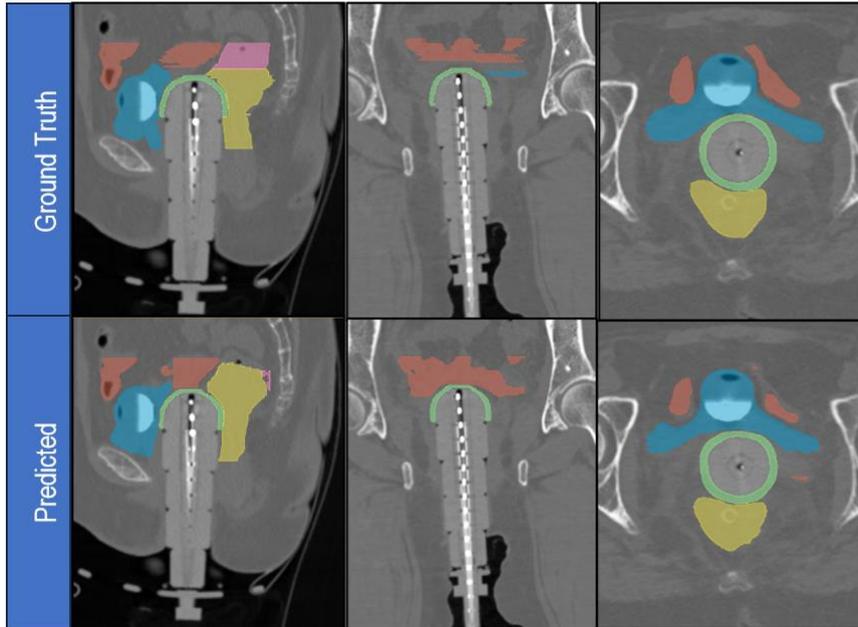


Figure 19. Comparison of the final predicted segmentations with respect to the manual contouring. (Sagittal, coronal and axial planes, respectively in each column)

In a further analysis of the results, the dosimetry of 10 patients was performed with the automatic and manual segmentations to compare the dosimetry output. The minimization of relative difference in the plannings should be the goal to aim for if an algorithm has to be applied in radiotherapy, since it would indicate that there are no significant differences in the result of the treatment.

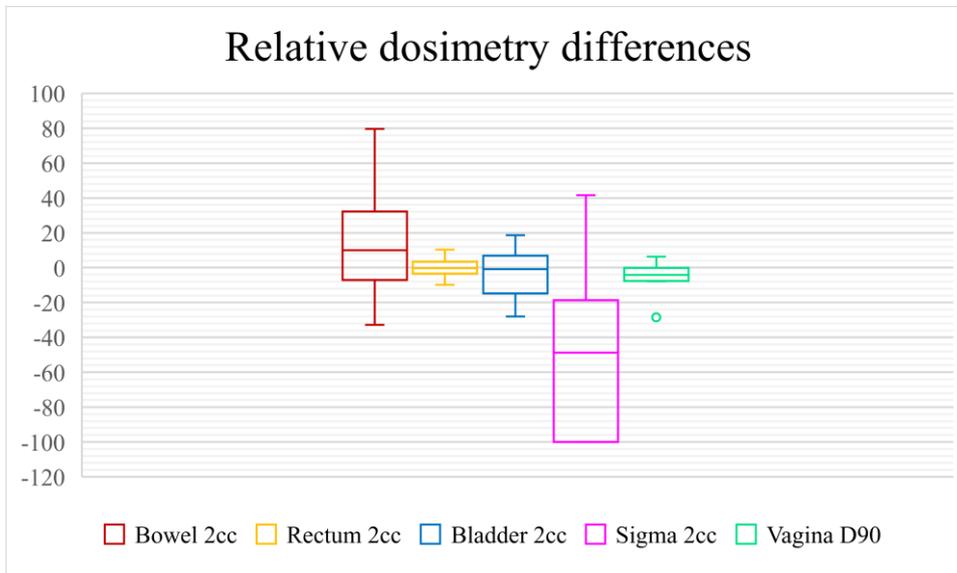


Figure 20. Relative dosimetry differences of the automatic segmentations with respect to the manual ones.

As can be seen in Figure 20, the dosimetry differed a 5% in the CTV (the vagina), a 17% in the bowel, a 3% in the rectum, an 8% in the bladder and a 52% in the sigma. The planning process carries a general uncertainty of an 8% caused by errors in the different steps of the dosimetry [37]. Regarding this fact, the automatic segmentations of the rectum, the bladder and the vagina would

be acceptable in clinics. On the other hand, the bowel and the sigma would require further improvements to be acceptable in clinics, considering that the resultant dosimetry prescribed too much dose on the bowel and far not enough dose in the sigma. Nevertheless, it should be taken into account that this is a preliminary study to indicate the approximated results, but further examinations of the results and the context should be done to obtain completely reliable conclusions.

5.5. DISCUSSION

As has been presented in Section 5.4. (Results), the performance of the model has some disparity in the contour accuracy of the different organs. While the bladder, the rectum and the CTV have shown a satisfactory automatic segmentation, the bowel and the sigma seem to be poorly delineated from the metrics perspective. Nevertheless, these results can be explained by a particular decision in the postprocessing step: To avoid the overlapping of areas between two ROIs, which could lead to higher prescriptions of dose in the planning, it was established a system that selected the bowel rather than the sigma in case the contour coincided in space. This caused a substantial reduction in the sigma volume, which was substituted as bowel volume, since the limit between the bowel and the sigma is quite indistinctive and is prone to cause confusions.

The change should not directly affect the output of the overall dosimetry since the prescription of dose in the sigma tissue is the same as in the bowel tissue, so the treatment would stay the same. Nevertheless, given that the metrics are performed individually on each ROI, the dosimetry results on the sigma were totally biased. Despite not having prior major effects on the final dosimetry, it would be recommendable to follow more sophisticated postprocessing strategies to adjust the segmentations as faithfully as possible to the original manual contours.

Beyond the postprocessing reasoning, the poor results in the segmentation of the sigma and the bowel (just as the errors in the CTV the bladder and the rectum) can also be originated by the nature of the database employed in the training. Despite being the only acceptable option to execute the project, the ground truth used to evaluate the models (the manual segmentations) was not the optimal since the organs weren't completely delineated and there was a lack of detail in some cases. This fact, together with the high variability of the ROIs disposition between patients, caused a limitation in the model performance.

There are two main solutions to this problem. First, the algorithm developed in the project remains at disposal of Hospital Clínic to be reused to train the models with a completely new and reliable

set of manual segmentations, external or internal. The only thing that should be changed in the final product would be the models by the new ones obtained in the training. The second option would be to establish a feedback system in which the product would be applied into the Radiotherapy Department and the automatic segmentations obtained from the new patients would be used as the future ground truth database. In this case, these segmentations should be carefully revised and edited as precise as possible by the specialists so they could be used clinically in the planning and in the new training, improving the output models. By doing one of the two options expounded, the algorithm could be more functional as an automatic tool, implying less editing by the technicians.

Considering completed the study, it can be stated that the main goal of the Project was fulfilled. It was successfully programmed an algorithm capable of segmenting automatically the ROIs in vaginal brachytherapy, reducing the time required to carry out the task to approximately 10 minutes based on preliminary tests. Nevertheless, the models implemented would require further improvements for the algorithm to be applied into clinics as a whole product.

6. EXECUTION CHRONOGRAM

6.1. WORK BREAKDOWN STRUCTURE

In order to plan a project adequately it is necessary to analyse each one of its phases, from the preparation or the execution to its results. The work breakdown structure (WBS) allows a global visualisation of the essential groups of tasks that are needed to carry out the project, so a temporal schedule can be developed.

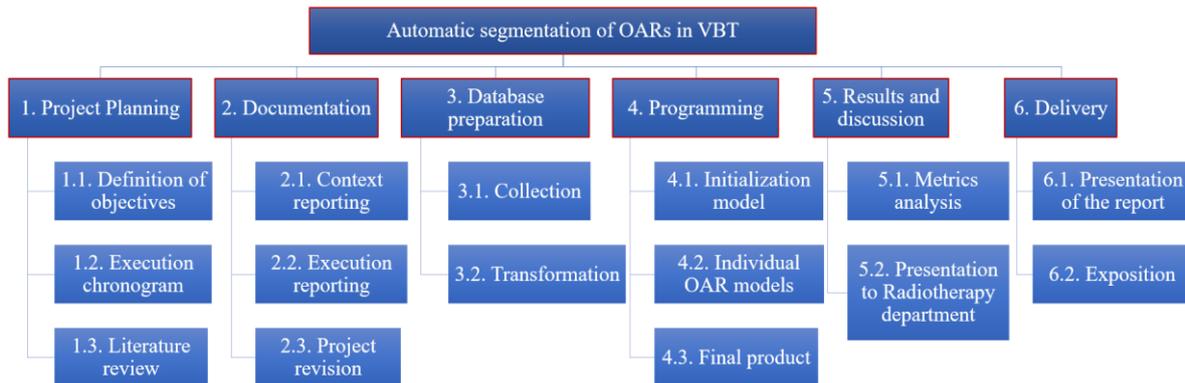


Figure 21. WBS of the project.

The project planning comprised the understanding of the objectives and methodology of the project, the chronological planning structure across the working timeline and the reading of the literature to have a better understanding of the project's scope and context. The documentation could be divided into the part that could be written before the execution of the project and the part reporting its process, adding the revision once the report was finished. The database preparation included the collection and transformation of the data, from anonymizing and changing the format to standardizing it. The programming included the training of the five models and the elaboration of the final algorithm. To analyse the results, the algorithm was tested with three quantitative metrics and by applying it to the radiotherapy department to see the effects on the dosimetry. Finally, the delivery of the project will be divided into the uploading of the report and the exposition.

6.2. PERT-CPM DIAGRAM

The PERT-CPM diagram shows the temporal implication and coordination of the different tasks by indicating their respective timings and interdependencies. This can help to identify the bottleneck path in which any delay could affect the final delivery schedule. With this knowledge it is possible to plan more efficiently the workflow of the project. As can be seen on Figure 22, while tasks as the Literature Review and the Context Reporting had time to spare, all the other tasks were classified as critical, requiring a special focus.

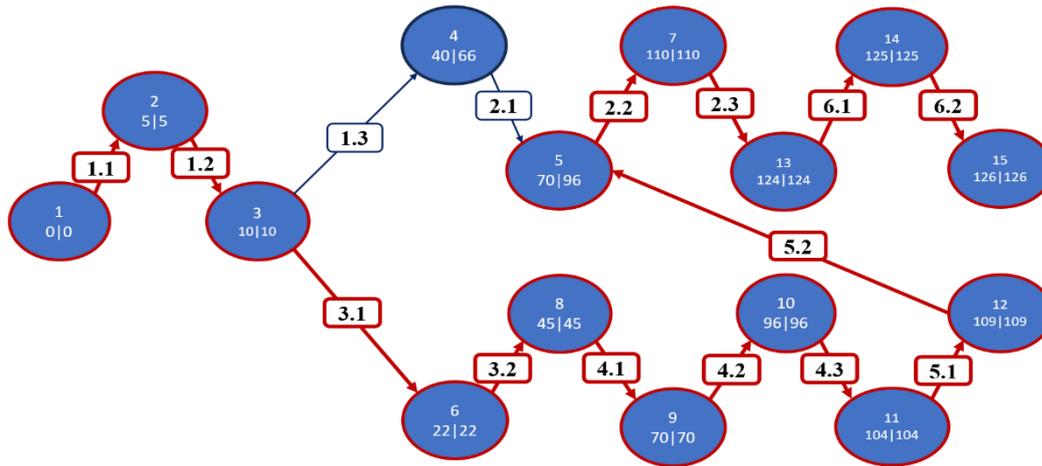


Figure 22. PERT-CPM diagram.

6.3. GANTT DIAGRAM

Finally, the GANTT diagram is set to present the temporal distribution of the tasks on the calendar. This sequence completes the PERT-CPM information by localizing the tasks from an initial date, in this case July 15th of 2023, to an end date, in this case January 26th of 2024. The total duration of the project has been then 6.5 months. As has been mentioned on the previous section, only two activities had spare time. This is representative of the lack of time available to execute this particular project, condensing all the tasks to the maximum to complete it for the delivery date.

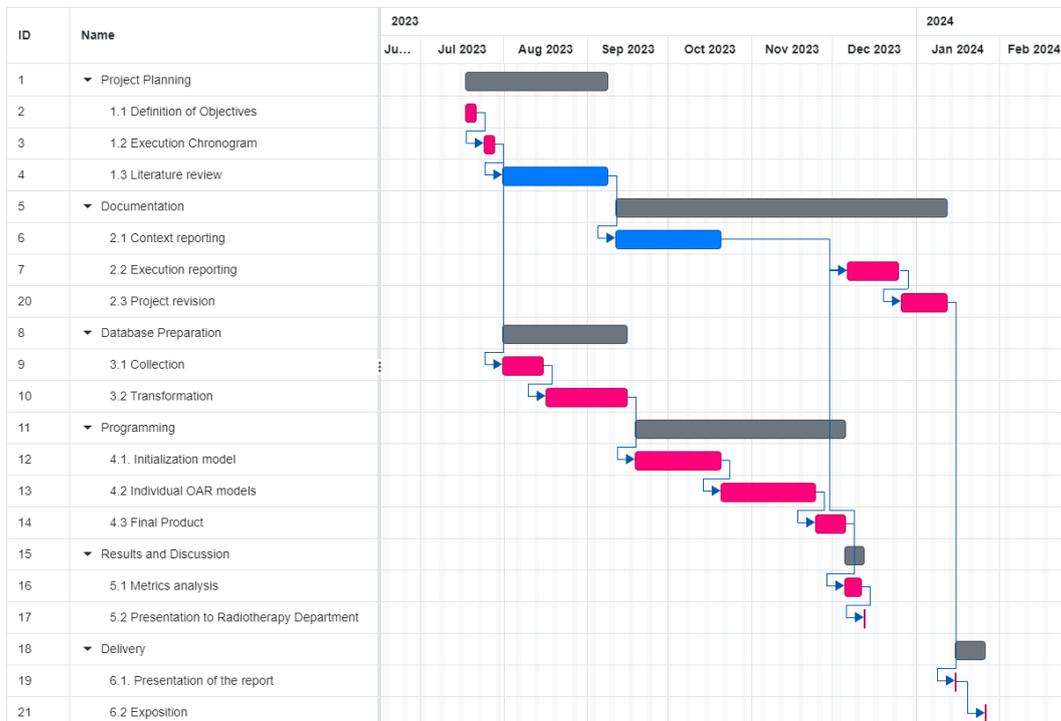


Figure 23. GANTT diagram of the project. [38]

7. TECHNICAL FEASIBILITY

Strengths	Weaknesses
<ul style="list-style-type: none"> ❖ Previous programming knowledge ❖ Satisfactory amount of subjects ❖ Disposition of a powerful computer ❖ Faster method than the current one ❖ Periodic external advice from the UB Biomedical Imaging Group 	<ul style="list-style-type: none"> ❖ Limited time ❖ Unreliable ground truth for the model ❖ High variability of the OAR contour ❖ Required further improvements ❖ Required revision and editing of the results
Opportunities	Threats
<ul style="list-style-type: none"> ❖ Lack of current competence ❖ Addresses endometrial carcinoma cases, which are increasing ❖ Less expensive ❖ Potential future improvement to overcome manual segmentations 	<ul style="list-style-type: none"> ❖ Possibility that big health companies develop the same technology ❖ Legal aspects of using AI in medicine ❖ Required continuous updating

Table 11. SWOT analysis of the project

The development of the automatic segmentation product for endometrial carcinoma treatment has turned out to be an arduous and complex project to make, but the results have been found satisfying. The SWOT analysis of Table 11 indicates the main strengths and weaknesses of the project and the opportunities and threats that would have to deal with in the market.

About the context, the main handicap that had to confront the project was the reduced time to execute it, limiting the study of solutions and improvements that could be applied to perfect the results. Another one of the main weaknesses was the unreliability of the manual segmentations used as the ground truth for the model training. This combined with the fact that the OAR regions have a substantial variability, would set an initial restriction for the model performance. On the other hand, there were some positive parts in the conception of the project. For instance, the number of subjects of the database was more than enough to be reliable. Moreover, the Alfa computer in the biophysics laboratory was available to be used to run the codes, which would save a precious amount of time since this computer has enough power to process all the input data.

During the development of the project, some external advice was received from the Biomedical Imaging Group of the UB, which was useful for the progress of the programming and achievement of the final product. Despite the long periods of time required to train the different models, the

process was followed quite fluently. Analysing the results of the first training (the initialization) it was seen that the algorithm demanded a higher level of complexity than a simple global model. This way, the individual OAR segmentation models were developed. By implementing bounding boxes and postprocessing on the algorithm, the errors of the output were minimized to the maximum extent possible.

About the result, it was found that the algorithm itself worked efficiently according to its function. It automatically segmented the OAR regions much faster than in manual segmentation, and it eliminated the human error. Nevertheless, the ground truth to train the models was not as accurate as it should be, so it would be recommended a secondary training with finer manual segmentations to improve the result. For now, the algorithm output would require editing of some parts of the segmentation in the revision.

About the introduction of this product to the market, there are some opportunities that it could take advantage of. Since there is no current product that covers the same stakeholders, there is a major opportunity to patent the technology and cover the needs of all the market. Moreover, since vaginal brachytherapy use is increasing following the growth rate of the endometrial carcinoma cases, the product would not become obsolete until a better technique was discovered. Given that it does not require much human intervention, it would globally become less expensive and more efficient, redistributing the time of the technicians that used to perform this task into another sector. Finally, benefiting from the external enhancement of deep learning technologies would open the possibility to overcome manual segmentations.

On the other hand, the biggest threat of the project is that big enterprises as Siemens Healthineers develop their own algorithm for the same treatment. Since these companies count on big storages of data, the models of their algorithm would reasonably be more accurate. Another concern that must be considered is the regulation of the AI use, since it would affect the implementation of the product. Finally, it also must be taken into account that this algorithm is not a static product since it needs periodic updates of the models, following the changes in the CT imaging and VBT technique that might modify the way segmentation is performed.

8. ECONOMIC FEASIBILITY

Item	Description	Cost per unit	Total cost
Data acquisition			
CT imaging	Number of patients= 298	140 €/scan [39]	41.720 €
Data processing			
Oncentra Brachy program	Planning program	100.000 €	100.000 €
Python software	Python 3.11	Free	0 €
Materials			
Alfa computer	Workstation Supermicro A+	8.302 € [40]	8.302 €
Human resources			
Biomedical engineering student	Total time= 300 hours	15 €/hour [41]	4.500 €
Director	Total time= 50 hours	25 €/hour [41]	1.250 €
TOTAL COST = 155.772 €		INTRINSIC TOTAL COST= 14.052 €	

Table 12. Economic viability of the project

This project has not entailed a direct economic investment but has taken profit of previously purchased items. The cost that would suppose the performance of CT imaging to 298 patients sums an approximated total of 41.720 €. *Oncentra Brachy*, the program used to exert the VBT planning (where the manual segmentations were done), would add a cost of 100.000 €.

Despite it is interesting to point out the approximate cost that would have entailed the creation of a database, these costs should not be considered to indicate the expense of the project since they were not executed to develop the automatic segmentation. They were only used posteriorly as an internal source of data with no additional cost.

On the other hand, what was strictly required for developing the project was the python software, which is a free tool, and the Alfa computer in the biophysics lab, with a cost of 8.302 €. Adding this to the respective salaries of a junior engineer and a biomedical engineer, the intrinsic final cost of the project would be 14.052 €.

9. REGULATIONS AND LEGAL ASPECTS

Considering that the project has been carried out in Barcelona, it has had to stick to the legislation of Spain and the European Union. This way, the study was executed following the ethical principles of the medical investigation described in the Helsinki Declaration [42]. Moreover, since the study is defined as a final degree project, it is framed in the general regulations of the UB, reported in the Royal Decree 1393/2007 [43].

The law 14/2007 [44] states the regulation of biomedical investigation in Spain. The topics tackled in this law were related to the definition of principles that must follow the researchers and rights of the patients that must be fulfilled. For instance, in the investigation related to humans it must be guaranteed the informed consent, the non-discrimination or the confidentiality and data protection.

The main right that had to be taken into account during the conception of the project was the preservation of confidentiality regarding the use of the records in the Radiotherapy Department. Obeying the constitutional law 3/2018 [45], the General Data Protection Regulation (GDPR) dictates that all the information that could relate the data to an individual has to be properly anonymised [46]. This information can go from locations, ethnicity, gender, biometric data to religious beliefs or political opinions.

In Hospital Clínic [47], there is also an internal regulation for the processing of data, which sticks to the EU Regulation 2016/679. When anonymizing the data, only the main investigator can keep the relationship between the MRN and the code designated to the patient. Any non-anonymized data must be transferred through the <https://compartir.clinic.cat> secured platform, and the access will be restricted only to the participants of the study. The database can only be stored 10 years before its erasing in case the study is followed. After that period, the only information kept will be the trained models created from the database.

The legislation of the insertion of AI in medicine is currently being discussed, but there is no consensus for now. Regarding this project, considering that VBT is not a life-threatening procedure, it is not expected to be highly regulated. Moreover, if there was an error in the dosimetry that endangered the security of the patient, the final responsibility would fall on the designated doctor, since it is always mandatory the revision of the segmentations and its approval from a human specialist before the application.

10. CONCLUSIONS AND FUTURE WORK

The elaboration of this Final Degree Project has culminated the work of three projects involved in the development of an automatic segmentation tool for VBT. The final product obtained is an algorithm that receives the CT images of the patient as an input and returns the segmentations of the CTV and the OARs with the adequate format to be inserted in the planning software. This algorithm can be automatically run without the need for the user to have programming knowledge.

About the secondary goals of the project, the database was successfully widened with recent records of Hospital Clínic's patients after following the right steps of anonymization and transformation during the process. By programming a training structure with V-Net as the DL architecture, it was possible to obtain the models that define the performance of the algorithm. With the preprocessing, the models and the postprocessing steps the algorithm could be completed.

The results obtained showed a satisfactory accuracy for the CTV, the rectum and the bladder. On the other hand, the bowel and the sigma were considered to require further examination. This disbalance was caused by an increase in the bowel's size and a substantial decrease in the sigma size due to the models' limitations and the rudimentary conception of the postprocessing step. Given the impact of these inaccuracies in the dosimetry, the algorithm currently would not be applicable to the clinics.

However, despite the product *per se* isn't still functional on its own, it can be used as a base to spend less time manually segmenting, having to edit only the incorrectly predicted ROIs. If the new segmentations are edited accurately by specialists, they can be used posteriorly as a new database for the same algorithms to perfect the models and obtain a marketable product. Moreover, since the structure of the algorithm is completely functional and it only requires an improvement in the models, it can be used by any healthcare centre or software company that offers VBT services. They would only have to upload their own data into the training algorithm, get the models and upload them on the final product.

A further improvement contemplated to commercialise the product once perfected is the creation of a complete interface instead of using an EXE file, resulting in a more aesthetic packaging.

11. BIBLIOGRAFY

- [1] “Tratamiento del cáncer de endometrio - NCI.” Accessed: Jan. 20, 2024. [Online]. Available: <https://www.cancer.gov/espanol/tipos/uterino/paciente/tratamiento-endometrio-pdq>
- [2] L. H. Ellenson, B. M. Ronnett, R. A. Soslow, R. J. Zaino, and R. J. Kurman, “Endometrial Carcinoma,” *Blaustein’s Pathology of the Female Genital Tract*, pp. 394–452, 2011, doi: 10.1007/978-1-4419-0489-8_9.
- [3] M. Koskas, F. Amant, M. R. Mirza, and C. L. Creutzberg, “Cancer of the corpus uteri: 2021 update,” *International Journal of Gynecology & Obstetrics*, vol. 155, no. S1, pp. 45–60, Oct. 2021, doi: 10.1002/IJGO.13866.
- [4] V. Makker *et al.*, “Endometrial Cancer,” *Nat Rev Dis Primers*, vol. 7, no. 1, p. 88, Dec. 2021, doi: 10.1038/S41572-021-00324-8.
- [5] “Biomedical imaging group - 2022-2023-TFM_SaraOrio.pdf - Tots els documents.” Accessed: Jan. 20, 2024. [Online]. Available: <https://ubarcelona.sharepoint.com/sites/big/Projectes%20estudiants/Forms/AllItems.aspx?id=%2Fsites%2Fbig%2FProjectes%20estudiants%2FTFMs%20entregats%2F2022%2D2023%2DTFM%5FSaraOrio%2Epdf&parent=%2Fsites%2Fbig%2FProjectes%20estudiants%2FTFMs%20entregats>
- [6] C. Janiesch, P. Zschech, and K. Heinrich, “Machine learning and deep learning,” *Electronic Markets*, vol. 31, no. 3, pp. 685–695, Sep. 2021, doi: 10.1007/S12525-021-00475-2/TABLES/2.
- [7] A. Akay and H. Hess, “Deep learning: Current and emerging applications in medicine and technology,” *IEEE J Biomed Health Inform*, vol. 23, no. 3, pp. 906–920, May 2019, doi: 10.1109/JBHI.2019.2894713.
- [8] A. Rosvoll Groendahl *et al.*, “Evaluation of auto-segmentation for brachytherapy of postoperative cervical cancer using deep learning-based workflow,” *Phys Med Biol*, vol. 68, no. 5, p. 055012, Feb. 2023, doi: 10.1088/1361-6560/ACBA76.
- [9] C. Elisabeth Olsson, R. Suresh, J. Niemelä, S. U. Akram, and A. Valdman, “Autosegmentation based on different-sized training datasets of consistently-curated volumes and impact on rectal contours in prostate cancer radiation therapy,” *Phys Imaging Radiat Oncol*, vol. 22, pp. 67–72, Apr. 2022, doi: 10.1016/j.phro.2022.04.007.
- [10] K. Kallis *et al.*, “Automated treatment planning framework for brachytherapy of cervical cancer using 3D dose predictions,” *Phys Med Biol*, vol. 68, no. 8, p. 085011, Apr. 2023, doi: 10.1088/1361-6560/ACC37C.

- [11] M. Lempart, J. Scherman, M. P. Nilsson, and C. Jamtheim Gustafsson, "Deep learning-based classification of organs at risk and delineation guideline in pelvic cancer radiation therapy," *J Appl Clin Med Phys*, vol. 24, no. 9, p. e14022, Sep. 2023, doi: 10.1002/ACM2.14022.
- [12] D. Duprez, C. Trauernicht, H. Simonds, and O. Williams, "Self-configuring nnU-Net for automatic delineation of the organs at risk and target in high-dose rate cervical brachytherapy, a low/middle-income country's experience," *J Appl Clin Med Phys*, vol. 24, no. 8, p. e13988, Aug. 2023, doi: 10.1002/ACM2.13988.
- [13] F. Zabihollahy, A. N. Viswanathan, E. J. Schmidt, and J. Lee, "Fully automated segmentation of clinical target volume in cervical cancer from magnetic resonance imaging with convolutional neural network," *J Appl Clin Med Phys*, vol. 23, no. 9, p. e13725, Sep. 2022, doi: 10.1002/ACM2.13725.
- [14] C. Xiao *et al.*, "RefineNet-based 2D and 3D automatic segmentations for clinical target volume and organs at risks for patients with cervical cancer in postoperative radiotherapy," *J Appl Clin Med Phys*, vol. 23, no. 7, p. e13631, Jul. 2022, doi: 10.1002/ACM2.13631.
- [15] "Biomedical imaging group - 2022-2023-TFG_ArnauAndres.pdf - Tots els documents." Accessed: Jan. 20, 2024. [Online]. Available: <https://ubarcelona.sharepoint.com/sites/big/Projectes%20estudiants/Forms/AllItems.aspx?id=%2Fsites%2Fbig%2FProjectes%20estudiants%2FTFGs%20entregats%2F2022%2D2023%2D2FTFG%5FArnauAndres%2Epdf&parent=%2Fsites%2Fbig%2FProjectes%20estudiants%2FTFGs%20entregats>
- [16] "Rising Endometrial Cancer Rates Spur New Approaches to Prevention | Division of Cancer Prevention." Accessed: Jan. 20, 2024. [Online]. Available: <https://prevention.cancer.gov/news-and-events/blog/rising-endometrial-cancer-rates-spur-new-approaches-prevention>
- [17] G. Kemikler, "History of Brachytherapy," *Oncol*, vol. 34, pp. 1–10, 2019, doi: 10.5505/tjo.2019.1.
- [18] M. J. Rivard, J. L. M. Venselaar, and L. Beaulieu, "The evolution of brachytherapy treatment planning," *Med Phys*, vol. 36, no. 6Part1, pp. 2136–2153, Jun. 2009, doi: 10.1118/1.3125136.
- [19] "Philips - Pinnacle³ Autosegmentación con SPICE Rápido. Consistente. Fiable." Accessed: Jan. 21, 2024. [Online]. Available: <https://www.philips.es/healthcare/product/HCNOCNTN137/autosegmentacin-con-spice-de-pinnacle-rpida-consistente-fiable>

- [20] “Philips - Pinnacle³ Segmentación basada en modelos.” Accessed: Jan. 21, 2024. [Online]. Available: <https://www.philips.es/healthcare/product/HCNOCNTN139/la-segmentacin-basada-en-modelos-de-pinnacle-con-la-funcin-arrastrar-y-soltar-mejora-el-flujo-de-trabajo>
- [21] “AI-segmentation – NMMIttools.” Accessed: Jan. 21, 2024. [Online]. Available: <https://nmmitools.org/category/nmmitools-ai-models/ai-segmentation/>
- [22] “syngo.MR Onco Engine - Siemens Healthineers España.” Accessed: Jan. 21, 2024. [Online]. Available: <https://www.siemens-healthineers.com/es/magnetic-resonance-imaging/options-and-upgrades/clinical-applications/syngo-mr-onco-engine>
- [23] “Whole-Body Dot Engine - Siemens Healthineers España.” Accessed: Jan. 21, 2024. [Online]. Available: <https://www.siemens-healthineers.com/es/magnetic-resonance-imaging/options-and-upgrades/clinical-applications/whole-body-dot-engine>
- [24] “MIM Contour Protégé AI - Auto Contouring | NL-Tec,” <https://www.nl-tec.com.au/>, Accessed: Jan. 21, 2024. [Online]. Available: <https://www.nl-tec.com.au/product/mim-contour-protege-ai/>
- [25] “GitHub - MIRACLE-Center/CTPelvic1K: Resources of the paper ‘Deep Learning to Segment Pelvic Bones: Large-scale CT Datasets and Baseline Models’.” Accessed: Jan. 21, 2024. [Online]. Available: <https://github.com/MIRACLE-Center/CTPelvic1K>
- [26] “The Hourglass Network | U-Net | V-Net | Medium.” Accessed: Jan. 21, 2024. [Online]. Available: <https://medium.com/@callieris.enrico/hourglass-network-6e74cdb9ce2f>
- [27] “V-Net, U-Net’s big brother in Image Segmentation | by François Porcher | Towards Data Science.” Accessed: Jan. 21, 2024. [Online]. Available: <https://towardsdatascience.com/v-net-u-nets-big-brother-in-image-segmentation-906e393968f7>
- [28] A. Hatamizadeh *et al.*, “UNETR: Transformers for 3D Medical Image Segmentation,” *Proceedings - 2022 IEEE/CVF Winter Conference on Applications of Computer Vision, WACV 2022*, pp. 1748–1758, Mar. 2021, doi: 10.1109/WACV51458.2022.00181.
- [29] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation”, Accessed: Jan. 21, 2024. [Online]. Available: <http://promise12.grand-challenge.org/results/>
- [30] “Week 3: Deep Learning model Shenanigans - BASIS Independent Silicon Valley.” Accessed: Jan. 21, 2024. [Online]. Available: <https://siliconvalley.basisindependent.com/2022/03/11/week-3-deep-learning-model-shenanigans/>

- [31] "Introduction to Batch Normalization: Understanding the Basics." Accessed: Jan. 21, 2024. [Online]. Available: <https://www.analyticsvidhya.com/blog/2021/03/introduction-to-batch-normalization/>
- [32] "Review — Gaussian Error Linear Units (GELUs) | by Sik-Ho Tsang | Medium." Accessed: Jan. 21, 2024. [Online]. Available: <https://sh-tsang.medium.com/review-gaussian-error-linear-units-gelus-d4d7347d1e11>
- [33] "A Complete Guide to Data Augmentation | DataCamp." Accessed: Jan. 21, 2024. [Online]. Available: <https://www.datacamp.com/tutorial/complete-guide-data-augmentation>
- [34] "Network architectures — MONAI 1.3.0 Documentation." Accessed: Jan. 21, 2024. [Online]. Available: <https://docs.monai.io/en/stable/networks.html#vnet>
- [35] K. H. Zou *et al.*, "Statistical Validation of Image Segmentation Quality Based on a Spatial Overlap Index: Scientific Reports," *Acad Radiol*, vol. 11, no. 2, p. 178, 2004, doi: 10.1016/S1076-6332(03)00671-8.
- [36] S. J. Reddi, S. Kale, and S. Kumar, "ON THE CONVERGENCE OF ADAM AND BEYOND".
- [37] C. Kirisits *et al.*, "Review of clinical brachytherapy uncertainties: analysis guidelines of GEC-ESTRO and the AAPM," *Radiother Oncol*, vol. 110, no. 1, pp. 199–212, 2014, doi: 10.1016/J.RADONC.2013.11.002.
- [38] "Free Online Gantt Chart Software." Accessed: Jan. 21, 2024. [Online]. Available: <https://www.onlinegantt.com/#/gantt>
- [39] "El precio de un TAC por privado puede ser más barato - iDoctor Asistencia Médica." Accessed: Jan. 21, 2024. [Online]. Available: <https://idoctor.es/blog/precio-de-un-tac-barato-por-privado/>
- [40] "Supermicro A+ SuperWorkstation 5014A-TT (AS-5014A-TT)." Accessed: Jan. 21, 2024. [Online]. Available: <https://www.thinkmate.com/system/a+-superworkstation-5014a-tt>
- [41] "¿Cuánto cobra un biomédico al mes? | UFV." Accessed: Jan. 21, 2024. [Online]. Available: <https://www.ufv.es/cuanto-cobra-un-biomedico-al-mes-preguntas-grados/>
- [42] "ANEXO A Declaración de Helsinki de la Asociación Médica Mundial".
- [43] "NORMES GENERALS REGULADORES DELS TREBALLS DE FI DE GRAU DE LA UNIVERSITAT DE BARCELONA," 2011.
- [44] "Ley 14/2007, de 3 de julio, de Investigación." Accessed: Jan. 21, 2024. [Online]. Available: https://noticias.juridicas.com/base_datos/Admin/l14-2007.html

- [45] “BOE-A-2018-16673 Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.” Accessed: Jan. 21, 2024. [Online]. Available: <https://www.boe.es/buscar/act.php?id=BOE-A-2018-16673>
- [46] “What is GDPR, the EU’s new data protection law? - GDPR.eu.” Accessed: Jan. 21, 2024. [Online]. Available: <https://gdpr.eu/what-is-gdpr/>
- [47] “Model Protocol Projectes Recerca o PS marcat CE,” 2021.
- [48] “Panoptic Segmentation: Definition, Datasets & Tutorial [2023].” Accessed: Jan. 20, 2024. [Online]. Available: <https://www.v7labs.com/blog/panoptic-segmentation-guide>